С. ХОКИНГ, ДЖ. ЭЛЛИС

КРУПНОМАСШТАБНАЯ СТРУКТУРА ПРОСТРАНСТВА-ВРЕМЕНИ

Перевод с английского
Э. А. ТАГИРОВА

Под редакцией
Я. А. СМОРОДИНСКОГО

ИЗДАТЕЛЬСТВО «МИР»
МОСКВА 1977


ИЗДАТЕЛЬСТВО
МИР

THE LARGE SCALE STRUCTURE OF SPACE-TIME

S. W. HAWKING, F. R. S.
Institute of Astronomy
and Gonville and Caius College, Cambridge
and
G. F. R. ELLIS
Department of Applied Mathematics,
University of Capetown

CAMBRIDGE UNIVERSITY PRESS
1973

УДК 52 + 53

Книга посвящена новому подходу к теории относительности и ее астрономическим приложениям, основанному на использовании методов современной дифференциальной геометрии. Применение их оказалось исключительно плодотворным при исследовании свойств пространства-времени в особых условиях, а именно проблем сингулярности в космологии, черных дыр и т. п.

Авторы рассмотрели роль гравитации, методы дифференциальной геометрии и общую относительность, физический смысл кривизны пространства-времени, точные решения и задачу Коши в общей теории относительности, проблему сингулярности и ее приложения к выяснению природы черных дыр и различных этапов расширения Вселенной.

Книга рассчитана на широкие круги физиков, математиков и астрономов, интересующихся проблемами релятивизма, и может служить учебным пособием для студентов.

Редакция космических исследований, астрономии и геофизики

$\text{X}\frac{20605\text{-}114}{041\,(01)\text{-}77}$ 114-77

© Cambridge University Press, 1973
© Перевод на русский язык, «Мир», 1977

Предисловие к русскому изданию

В книге С. Хокинга и Дж. Эллиса общая теория относительности изложена как математическая теория геометрии пространства-времени. Задача авторов — описать на современном математическом языке свойства решений уравнений Эйнштейна. Язык, которым пользуются авторы, пока мало привычен для физиков и может показаться многим слишком абстрактным. Но развитие науки всегда приводит к изменению ее языка. Это происходит тогда, когда новые задачи оказываются слишком запутанными, если их пытаться исследовать, применяя лишь старый привычный аппарат.

Когда в книгах по электродинамике в 30-х гг. появились формулы векторного анализа, то говорили о потере наглядности и физического смысла формул. Столь же непривычен и труден был аппарат квантовой механики. Ясно, что для решения старых задач, таких как движение перигелия Меркурия, старые методы проще и нагляднее. Новый язык понадобился тогда, когда физики от поисков конкретных решений уравнений Эйнштейна перешли к качественному исследованию этих уравнений.

Особенно «тяжелой» оказались проблемы сингулярностей, или «особенностей». Всякое ли решение уравнения Эйнштейна имеет истинную (не устранимую координатным преобразованием) особенность или нет? После многих работ, среди которых надо отметить фундаментальные работы Лифшица, Халатникова, Белинского, исследовавших поведение решений вблизи особенности (старыми методами), необходимость новой точки зрения на всю проблему стала очевидной.

Новую точку зрения — по существу новое направление — открыли работы Пенроуза, а вслед за ними работы Хокинга. Именно это направление — от основ до трудных обобщающих теорем — изложены в предлагаемой книге последовательно и очень четко. Читатель, преодолевший трудности освоения языка современной дифференциальной геометрии, будет вознагражден той красотой, которая откроется в последующих главах, когда речь пойдет уже о конкретных типах пространства-времени, представленных в виде карт и атласов, и увидит, как далеко можно уйти от классических моделей.

Книга построена так, что в ней изложен весь необходимый математический материал. Однако некоторая общая подготовка

читателя, в рамках элементарных курсов топологии, все же необходима. Изучение книги убедит читателя в том, что новые математические средства, вводимые авторами в общую теорию относительности, необычайно сильны и необходимы для дальнейшего развития теории.

В переводе книги мы старались следовать установившейся русской математической терминологии, но иногда приходилось изобретать и новые термины; для объяснения их полезно пользоваться словариком-указателем в конце книги.

Я. Смородинский

Посвящается Д. Шаме

Предисловие

Предмет этой книги — структура пространства-времени в масштабах длин от радиуса элементарной частицы ($\sim 10^{-13}$ см) до радиуса Вселенной ($\sim 10^{28}$ см). По причинам, указанным в гл. 1 и 3, наш подход основан на общей теории относительности Эйнштейна. Эта теория ведет к двум следующим предсказаниям относительно Вселенной: во-первых, конечной участью массивной звезды является коллапс за горизонт событий; при этом образуется «черная дыра», внутри которой заключена сингулярность; во-вторых, в нашем прошлом существует сингулярность, которая представляет собой в некотором смысле начало наблюдаемой Вселенной. Наше изложение в основном направлено на обоснование двух указанных выше результатов. Это связано прежде всего с двумя областями исследований: с теорией поведения семейств времениподобных и изотропных кривых в пространстве-времени и с изучением характера различных отношений причинности в пространстве-времени. Мы подробно обсудим эти проблемы. Кроме того, в книге излагается теория развития во времени решений уравнений Эйнштейна с заданными начальными данными. Параллельно мы исследуем глобальные свойства разнообразных точных решений уравнений поля Эйнштейна, многие из которых обнаруживают неожиданные свойства.

Частично эта книга основывается на обзорном докладе, представленном в связи с награждением одного из нас (С. Х.) премией Адамса. Многие из идей, изложенных здесь, принадлежат Р. Пенроузу и Р. Героку; мы благодарим их за помощь и рекомендуем читателям их обзорные статьи [58, 59, 128, 131]. Большую пользу мы извлекли из дискуссий со многими нашими коллегами, особенно с Б. Картером и Д. Шамой. Им мы также приносим благодарность.

Кембридж
Январь 1973

С. Хокинг

Дж. Эллис

Глава 1

Роль тяготения

Согласно общепринятому в современной физике взгляду, изучение Вселенной можно разделить на две части. Первая — это установление локальных законов, которым удовлетворяют различные физические поля. Эти законы обычно формулируются в виде дифференциальных уравнений. Вторую часть составляют проблема граничных условий для этих уравнений и глобальные свойства их решений. Она в том или ином смысле содержит в себе идею о крае пространства-времени. Эти две части не обязательно независимы. Действительно, уже была высказана мысль, что локальные законы определяются крупномасштабной структурой Вселенной. Эту точку зрения обычно связывают с именем Маха, а в менее отдаленном прошлом ее развивали Дирак [41], Шама [155], Дикке [39], Хойл и Нарликар [82] и др. Мы не будем придерживаться столь возвышенной точки зрения: мы просто начнем с локальных физических законов, установленных экспериментально, и посмотрим, к каким выводам относительно крупномасштабной структуры Вселенной они могут привести.

Конечно, предположение о том, что физические законы, установленные в лаборатории, применимы во всех других точках пространства-времени, где условия могут быть совсем иными — это далеко идущая экстраполяция. Если эти законы перестанут выполняться, мы скажем, что имеется еще какое-то поле, которое повлияло на локальные законы, однако существование его до сих пор не было обнаружено в наших экспериментах из-за того, что оно мало меняется в пределах такой области, как Солнечная система. На самом деле бóльшая часть наших результатов не будет зависеть от конкретного характера рассматриваемых физических законов: они окажутся связанными лишь с определенными общими свойствами, такими, как описание пространства-времени посредством псевдоримановой геометрии и положительная определенность плотности энергии.

Фундаментальные взаимодействия, известные сейчас в физике, можно подразделить на четыре класса: сильные и слабые ядерные взаимодействия, электромагнетизм и тяготение. Среди перечисленных тяготение намного слабее остальных: Gm^2/e^2, отношение силы тяготения между двумя электронами к силе их электростатического взаимодействия, по порядку величины

равно 10^{-40}. Тем не менее тяготение играет доминирующую роль в формировании крупномасштабной структуры Вселенной. Это происходит из-за того, что сильное и слабое взаимодействия обладают весьма коротким радиусом действия ($\sim 10^{-12}$ см и менее), а электромагнетизм хотя и является дальнодействующим взаимодействием, но для тел макроскопических размеров отталкивание одноименных зарядов почти полностью компенсируется притяжением разноименных. Напротив, тяготение, по-видимому, всегда является силой притяжения. Следовательно, гравитационные поля всех частиц тела складываются и создают поле, которое для достаточно больших тел преобладает над всеми другими силами.

Тяготение не только доминирует на больших расстояниях, но и является силой, которая одинаковым образом действует на все частицы. Эту универсальность тяготения впервые выявил Галилей, который обнаружил, что два любых тела падают с одинаковой скоростью. Позднее это было проверено с очень большой точностью опытами Этвеша и Дикке с сотрудниками [39]. Было обнаружено также, что свет отклоняется гравитационными полями. Поскольку принимается, что нет сигналов, распространяющихся быстрее света, это означает, что именно тяготение определяет причинную структуру Вселенной, т. е. определяет, какие события пространства-времени могут быть причинно связаны друг с другом.

Эти свойства тяготения служат источником серьезных трудностей: если бы в некоторой области сконцентрировалось достаточно большое количество вещества, то последнее могло бы отклонить свет, испускаемый из этой области, настолько, что свет был бы втянут обратно в эту область. Эта возможность была осознана в 1798 г. Лапласом, который отметил, что тело такой же плотности, как и Солнце, но с радиусом в 250 раз бóльшим создало бы столь сильное гравитационное поле, что свет не мог бы уйти с его поверхности. Тот факт, что это было предсказано так давно, настолько поразителен, что мы приводим в приложении перевод статьи Лапласа.

Втягивание световых лучей массивным телом можно представить более точно, используя идею Пенроуза о замкнутых ловушечных поверхностях. Рассмотрим сферу $\mathcal{T}$, окружающую некоторое тело. Пусть в какой-то момент от $\mathcal{T}$ исходит импульс света. В некоторый последующий момент t фронты сходящихся и расходящихся от $\mathcal{T}$ волн образуют сферы $\mathcal{T}_1$ и $\mathcal{T}_2$ соответственно. В нормальных условиях площадь $\mathcal{T}_1$ (поскольку она представляет собой сходящийся свет) будет меньше, а площадь $\mathcal{T}_2$ (поскольку она представляет собой расходящийся свет) (см. рис. 1) — больше, чем площадь $\mathcal{T}$. Однако если внутри $\mathcal{T}$ заключено достаточно много вещества, то и площадь $\mathcal{T}_1$, и пло-

щадь $\mathcal{T}_2$ будут *меньше* площади $\mathcal{T}$. В этом случае поверхность $\mathcal{T}$ называют замкнутой ловушечной поверхностью. По мере возрастания времени t площадь $\mathcal{T}_2$ будет становиться все меньше и меньше (при условии, что тяготение остается силой притяжения, т. е. при условии, что плотность энергии не станет отрицательной). Поскольку вещество внутри сферы $\mathcal{T}$ не может иметь скорости, превышающей световую, оно окажется запертым

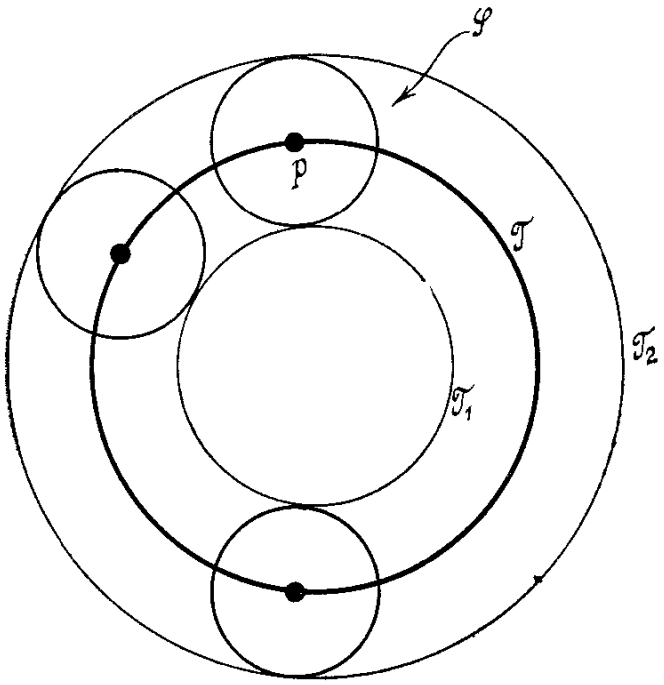


Рис. 1. В некоторый момент времени сфера $\mathcal{T}$ излучает вспышку света. В более поздний момент свет, излученный в точке p, образует сферу $\mathcal{S}$ вокруг p, и огибающие поверхности $\mathcal{T}_1$ и $\mathcal{T}_2$ образуют соответственно фронты сходящихся и расходящихся волн. Если площади *обеих* поверхностей $\mathcal{T}_1$ и $\mathcal{T}_2$ меньше площади $\mathcal{T}$, то $\mathcal{T}$ представляет собой замкнутую ловушечную поверхность.

в области, граница которой стягивается к нулю за конечный промежуток времени. Это наводит на мысль, что наша модель в чем-то ошибочна.

Однако мы покажем, что при выполнении некоторых разумных условий здесь на самом деле возникает сингулярность пространства-времени. Можно думать, что сингулярность — это место, где ныне известные нам физические законы нарушаются. В качестве альтернативы можно считать, что сингулярность представляет часть границы пространства-времени, которая, однако, находится не на бесконечности, а на конечном расстоянии. Тогда сингулярности перестают быть чем-то порочным с точки зрения физики; но и при таком взгляде на вещи нам еще остается решить проблему граничных условий. Другими словами, нам не известно, что именно выходит из сингулярности.

Предполагается, что в двух случаях концентрация материи будет достаточно для возникновения замкнутых ловушечных поверхностей. Первый случай — гравитационный коллапс звезд с массой больше двух масс Солнца; этот случай должен иметь место после выгорания ядерного горючего звезды. При этом звезда, как считают, коллапсирует к сингулярности, которая невидима для внешних наблюдателей. Второй случай — это сама Вселенная как целое. Недавние наблюдения микроволнового фонового излучения свидетельствуют, что Вселенная содержит достаточное количество материи для возникновения замкнутой ловушечной поверхности при обращении времени. Это означает, что в прошлом, в начале нынешней эпохи расширения Вселенной, существовала сингулярность. Эту сингулярность мы можем в принципе наблюдать, и ее можно было бы интерпретировать как начало Вселенной.

В данной книге мы рассматриваем крупномасштабную структуру пространства-времени на основе общей теории относительности Эйнштейна. Выводы из этой теории находятся в согласии со всеми выполненными до сих пор экспериментами. Однако наша трактовка предмета будет достаточно общей и для того, чтобы охватить такие модификации теории Эйнштейна, как теория Бранса — Дикке.

Хотя мы ожидаем, что большинство наших читателей знакомы с общей теорией относительности, мы все же пытались написать книгу так, чтобы от читателя не требовалось иных знаний, кроме математического анализа, алгебры и топологии точечных множеств. Вот почему мы посвящаем гл. 2 дифференциальной геометрии. Наш подход является в разумной мере современным в том отношении, что мы даем наши определения в форме, не зависящей явно от координат. Однако для удобства вычислений временами мы пользуемся индексами; большей частью мы избегали также употребления расслоенных пространств. Читатель, немного знающий дифференциальную геометрию, может при желании пропустить эту главу.

В гл. 3 дана формулировка общей теории относительности в виде трех постулатов математической модели пространства-времени. Эта модель представляет собой многообразие $\mathscr{M}$ вместе с метрикой $\mathbf{g}$ лоренцевой сигнатуры. Физический смысл метрики раскрывается в первых двух постулатах — локальной причинности и локального сохранения энергии-импульса. Эти постулаты общие для специальной и общей теорий относительности, и, следовательно, основанием для них может служить экспериментальное подтверждение первой из этих теорий. Третий постулат — полевые уравнения для метрики $\mathbf{g}$ — экспериментально обоснован не столь хорошо. Однако большинство наших результатов будет зависеть лишь от того следствия уравнений поля, что для

положительной плотности материи тяготение является силой притяжения. К этому свойству тяготения приводят как общая теория относительности, так и некоторые ее модификации, например теория Бранса — Дикке.

В гл. 4 устанавливается физический смысл кривизны, для чего мы изучаем ее влияние на семейства времениподобных и изотропных геодезических. Указанные линии представляют пути малых частиц и световые лучи соответственно. Кривизну можно интерпретировать как приливную силу, которая вызывает относительные ускорения между «соседними» геодезическими. Если

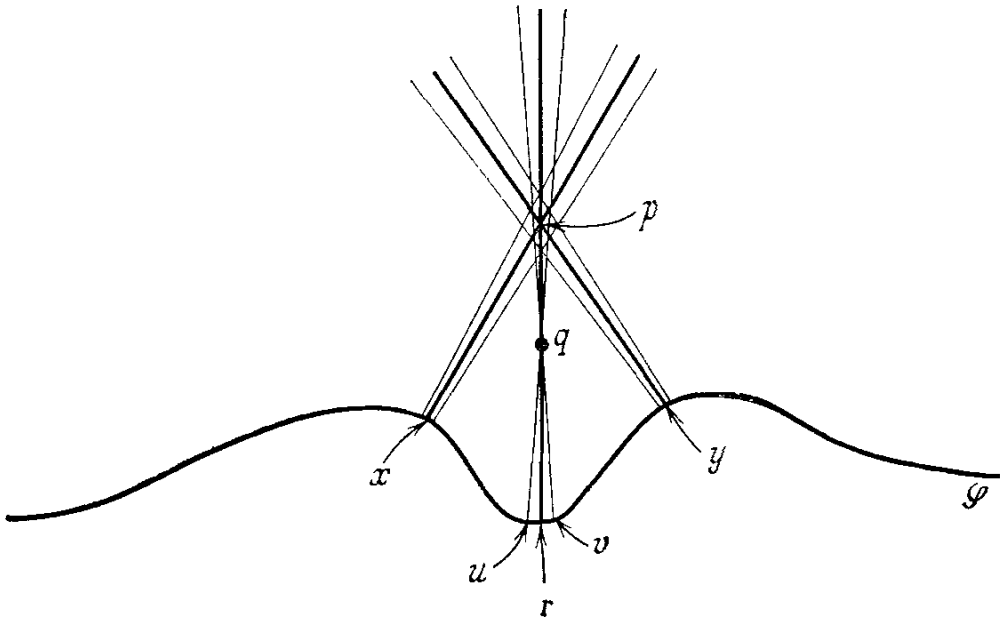


Рис. 2. Линия pr не может быть кратчайшей линией от p до $\mathscr{S}$, поскольку между p и r имеется фокальная точка q. На самом деле кратчайшей линией от p до $\mathscr{S}$ будет px или py.

тензор энергии-импульса удовлетворяет условиям положительной определенности, то результирующее действие приливной силы на невращающиеся семейства геодезических всегда оказывается собирательным. Используя уравнение Райчаудхури (4.26), можно показать, что это приводит к существованию фокальных или сопряженных точек, в которых соседние геодезические пересекаются.

Чтобы выявить роль этих фокальных точек, рассмотрим одномерную поверхность $\mathscr{S}$ в двумерном евклидовом пространстве (рис. 2). Пусть p — точка, не лежащая на $\mathscr{S}$. Тогда найдется кривая, соединяющая $\mathscr{S}$ и p, которая будет короче (или не длиннее) любой другой кривой от $\mathscr{S}$ до p. Ясно, что эта кривая будет геодезической, т. е. прямой линией и будет пересекать $\mathscr{S}$ ортогонально. На самом деле, в ситуации, изображенной на рис. 2, имеется три геодезические, ортогональные к $\mathscr{S}$ и проходящие через p. Геодезическая, проходящая через точку r, очевидно, не

есть кратчайшая линия от $\mathscr{S}$ до p. В этом можно убедиться [103], например, заметив, что соседние геодезические, ортогональные к $\mathscr{S}$ в точках u и v, пересекают геодезическую, проведенную из r, в фокальной точке q, лежащей между $\mathscr{S}$ и p. Тогда, присоединяя к отрезку uq отрезок qp, мы получим кривую от $\mathscr{S}$ до p, длина которой такая же, как у прямой rp. Но поскольку uqp не есть прямая линия, можно было бы срезать угол у точки q и получить кривую короче, чем rp. Отсюда видно, что rp не есть кратчайшая кривая от $\mathscr{S}$ до p. Значит, кратчайшей кривой будут или xp, или yp.

Эти соображения можно перенести на случай четырехмерного пространственно-временного многообразия $\mathscr{M}$ с лоренцевой метрикой $\mathbf{g}$. Вместо прямых линий тогда нужно рассматривать геодезические, а вместо кратчайшей кривой искать времениподобную кривую наибольшей длины между точкой p и пространственноподобной поверхностью $\mathscr{S}$ (ввиду лоренцевой сигнатуры метрики здесь не будет кратчайшей времениподобной кривой, но, возможно, найдется такая кривая наибольшей длины). Эта кривая должна быть геодезической, пересекающей $\mathscr{S}$ ортогонально, и между $\mathscr{S}$ и p не может быть какой-либо фокальной точки геодезических, ортогональных к $\mathscr{S}$. Аналогичные результаты можно получить для изотропных геодезических. Эти результаты используются в гл. 8 для доказательства существования сингулярностей при определенных условиях.

В гл. 5 мы описываем ряд точных решений уравнений Эйнштейна. Эти решения не являются реалистическими в том смысле, что все они обладают точными симметриями. Однако они служат полезными примерами для последующих глав и иллюстрируют различные возможные поведения решений. В частности, космологические модели, обладающие высокой симметрией, почти все имеют пространственно-временные сингулярности. Долгое время думали, что эти сингулярности, возможно, обусловлены именно высокой степенью симметрии и они будут отсутствовать в более реалистических моделях. Показать, что это на самом деле не так — одна из основных наших целей.

В гл. 6 мы исследуем причинную структуру пространства-времени. В частной теории относительности те события, которые могут причинно влиять на некоторое данное событие, и те, на которые это событие само может причинно влиять, образуют внутренние части световых конусов прошлого и будущего соответственно (рис. 3). В общей теории относительности метрика $\mathbf{g}$, которая задает световые конусы, вообще говоря, меняется от точки к точке, а топология пространственно-временного многообразия $\mathscr{M}$ не обязательно совпадает с топологией евклидова пространства R^4. Это приводит к большему разнообразию возможностей. К примеру, можно было отождествить соответствую-

щие точки на поверхностях $\mathscr{S}_1$ и $\mathscr{S}_2$ на рис. 3, чтобы получить пространство-время с топологией $R^3 \times S^1$. В нем есть замкнутые времениподобные кривые. Однако существование таких кривых

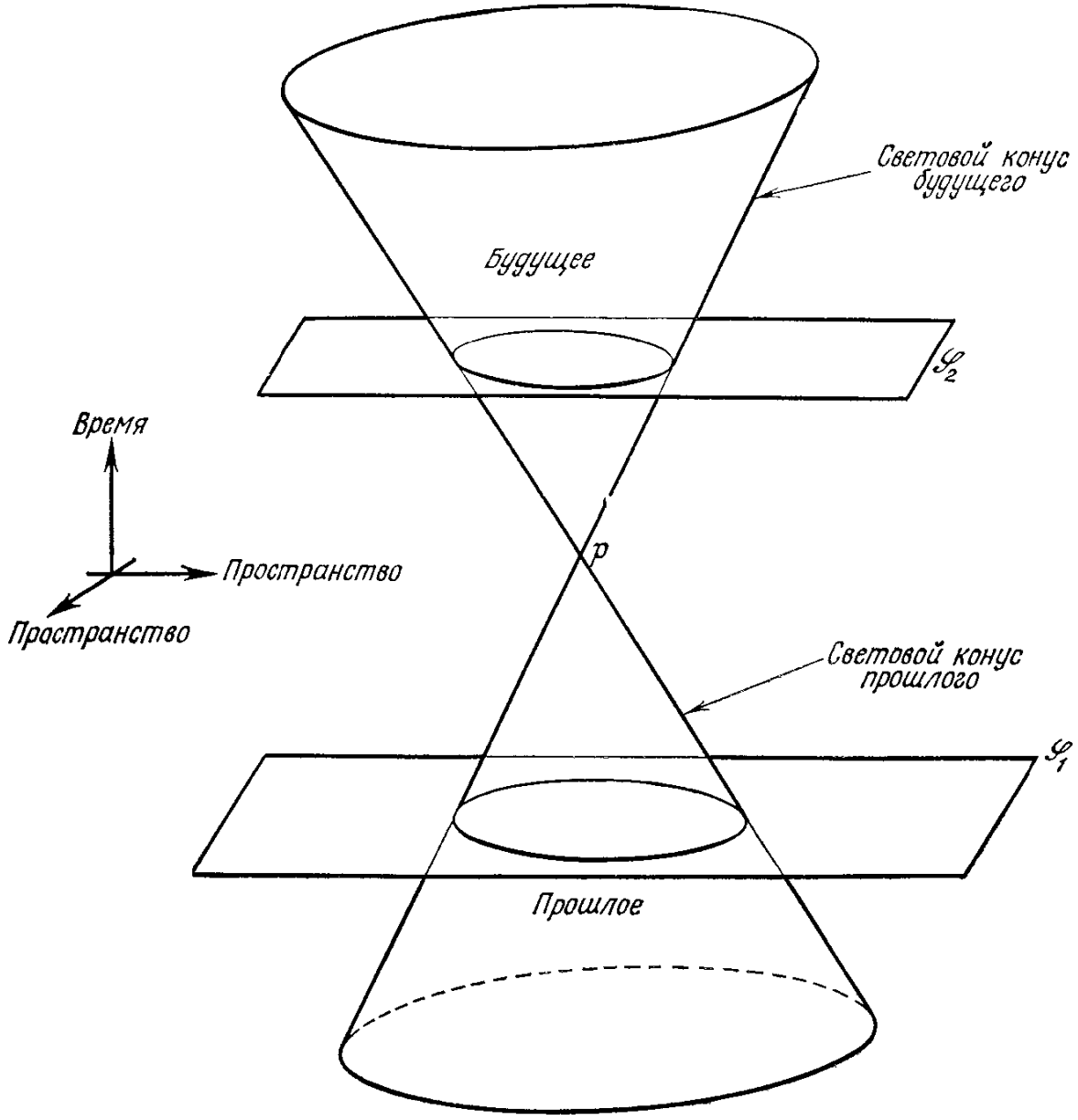


Рис. 3. В специальной теории относительности световой конус события p является множеством всех лучей света, проходящих через p. Прошлое событие p — это внутренность светового конуса прошлого, а будущее событие p — внутренность светового конуса будущего.

привело бы к нарушению причинности: можно было бы путешествовать в собственное прошлое. Большей частью мы ограничимся рассмотрением пространственно-временных многообразий, в которых такого рода нарушения причинности исключены. Тогда в пространстве-времени для любой данной пространственноподобной поверхности $\mathscr{S}$ существует максимальная область

пространства-времени — назовем ее областью Коши (Cauchy development *)), — в которой можно делать предсказания по известным начальным данным на $\mathscr{S}$. Область Коши обладает тем свойством («глобальная гиперболичность»), что при наличии в ней двух точек, которые можно соединить времениподобной кривой, между ними существует такая кривая наибольшей длины; она будет геодезической.

Причинная структура пространства-времени может быть использована для определения границы или края пространства-времени. Эта граница включает в себя как бесконечность, так и те части края пространства-времени, которые находятся на конечном расстоянии, т. е. сингулярные точки.

В гл. 7. мы рассматриваем задачу Коши для общей теории относительности. Будет показано, что начальные данные на пространственноподобной поверхности определяют решение однозначно в области Коши для этой поверхности и что в некотором смысле это решение зависит непрерывно от начальных данных. Эта глава включается для полноты изложения, а также потому, что в ней используется ряд результатов предыдущей главы, но чтение ее не обязательно для понимания последующих глав.

В гл. 8 обсуждается определение пространственно-временных сингулярностей. Это представляет известную трудность, поскольку нет возможности рассматривать сингулярные точки как часть пространственно-временного многообразия $\mathscr{M}$.

Затем мы доказываем четыре теоремы, которые устанавливают наличие сингулярностей при определенных условиях. Эти условия распадаются на три категории. Во-первых, требуется чтобы тяготение было силой притяжения; это может быть выражено в виде некоторого неравенства для тензора энергии-импульса. Во-вторых требуется, чтобы количества материи, содержащегося в некоторой области, было бы достаточно для того, чтобы ничто не могло покинуть эту область. Это произойдет, если имеется замкнутая ловушечная поверхность или если вся Вселенная сама пространственно замкнута. Третье требование состоит в том, чтобы не было никаких нарушений причинности. Однако в одной из доказываемых теорем это требование не является необходимым. Основная идея доказательства состоит в том, чтобы с помощью результатов гл. 6 показать, что между определенными парами точек должны существовать времениподобные кривые наибольшей длины. Тогда можно доказать, что если бы сингулярности отсутствовали, то были бы фокальные

*) В нашей литературе термин «Cauchy development» иногда переводят как «область зависимости». — *Прим. перев.*

точки, а, следовательно, не было бы кривой наибольшей длины между этими парами точек.

После этого мы даем описание предложенной Шмидтом процедуры построения границы пространства-времени, отображающей его сингулярные точки. Эта граница может отличаться от той части причинной границы (ее определение дано в гл. 6), которая соответствует сингулярностям.

В гл. 9 показано, что второе условие теоремы 2 гл. 8 должно выполняться вблизи звезд с массой 1,5 массы Солнца на конечных стадиях их эволюции. Сингулярности, которые возникают при этом, вероятно, находятся за горизонтом событий и ненаблюдаемы извне. Там, где когда-то была звезда, внешний наблюдатель видит «черную дыру». Мы рассмотрим свойства этих черных дыр и покажем, что вероятным итогом процесса образования их является установление в пространстве-времени метрики из семейства решений Керра. В предположении, что так на самом деле и происходит, можно наложить определенные ограничения сверху на количество энергии, которое можно извлечь из черной дыры. В гл. 10 показано, что второе условие теорем 2 и 3 гл. 8 выполнено при обращении знака времени для Вселенной в целом. В этом случае сингулярности были в прошлом по отношению к нам и представляют собой начало развития всей наблюдаемой Вселенной или ее части.

Существенная часть вводного материала содержится в разделах 3.1, 3.2 и 3.4. Читателю, желающему понять лишь теоремы, предсказывающие существование сингулярностей во Вселенной, достаточно затем прочитать только гл. 4 и разд. 6.2—6.7, 8.1, 8.2. Применение этих теорем к коллапсирующим звездам дано в разд. 9.1 (где используются также результаты приложения Б); в разд. 10.1 эти теоремы на основе некоторого истолкования моделей Робертсона — Уокера применяются к Вселенной в целом. Обсуждение характера сингулярностей содержится в разд. 8.1, 8.3—8.5 и 10.2; важную роль при этом играет пример пространства Тауба — НУТ (разд. 5.8). Определенный интерес представляет также модель Вселенной типа I Бианки (разд. 5.4).

Читателю, желающему ознакомиться лишь с вопросом о черных дырах, необходимо прочитать гл. 4 и разд. 6.2—6.6, 6.9, 9.1—9.3. Здесь мы основываемся на определенной интерпретации решения Шварцшильда (разд. 5.5) и решения Керра (разд. 5.6).

Наконец, читатель, который в первую очередь интересуется свойствами эволюции во времени уравнений Эйнштейна, может ограничиться чтением разд. 6.2—6.6 и гл. 7. Для него окажутся интересными примеры, приведенные в разд. 5.1, 5.2 и 5.5.

Мы старались сделать указатель полезным путеводителем по всем введенным определениям и связям между ними.

Глава 2

Дифференциальная геометрия

Структура пространства-времени, о которой пойдет речь в следующей главе и которая является основной для изложения во всей остальной части книги, — это структура многообразия, наделенного лоренцевой метрикой и определяемой ею аффинной связностью.

В настоящей главе, в разд. 2.1, мы введем понятие многообразия, а в разд. 2.2 — векторы и тензоры, являющиеся естественными геометрическими объектами, определяемыми на многообразии. В разд. 2.3 будут рассмотрены отображения многообразий, что приведет нас к понятиям подмногообразий и индуцированных отображений тензоров. Производная индуцированного отображения, определяемого векторным полем, дает производную Ли; она вводится в разд. 2.4. Там же мы даем определение внешнего дифференцирования — другой дифференциальной операции, зависящей только от структуры многообразия. Эта операция встречается в обобщенной форме теоремы Стокса.

В разд. 2.5 вводится дополнительная структура — связность; это позволяет ввести ковариантную производную и тензор кривизны. В разд. 2.6 рассмотрено соотношение между метрикой и связностью и дано выражение для тензора кривизны через тензор Вейля и тензор Риччи, которые связаны между собой тождествами Бианки.

В оставшейся части главы мы обсудим ряд других вопросов дифференциальной геометрии. В разд. 2.7 рассмотрены метрика и связность, индуцированные на гиперповерхности, и выведены уравнения Гаусса — Кодацци. В разд. 2.8 вводится задаваемый метрикой элемент объема, который затем используется для доказательства теоремы Гаусса. Наконец, в разд. 2.8 мы вкратце остановимся на расслоенных пространствах, уделяя особое внимание касательным и реперным расслоениям. Это позволит переформулировать многие из введенных ранее понятий на красивом геометрическом языке. Сведения из разд. 2.7 и 2.9 в дальнейшем используются только в одном или двух местах и несущественны для понимания основного содержания книги.

2.1. Многообразия

По существу многообразие есть пространство, локально подобное евклидову пространству в том смысле, что оно может быть покрыто кусками координатных сетей. Такая структура позволяет ввести дифференцирование, но не позволяет различить внутренним образом разные координатные системы. Следовательно, структурой многообразия определяются такие понятия, которые не зависят от выбора системы координат. Мы дадим точную формулировку понятия многообразия, введя предварительно некоторые определения.

Обозначим через R^n *евклидово пространство n измерений*, т. е. множество всевозможных наборов n чисел $(x^1, x^2, \ldots, x^n)$, $(-\infty < x^i < \infty)$ с обычной топологией (открытые и замкнутые множества определены, как обычно). Пусть $\frac{1}{2}R^n$ означает «нижнюю половину» R^n, т. е. область R^n, в которой $x^1 \leqslant 0$. Отображение ϕ открытого множества $\mathcal{O} \subset R^n$ (соответственно $\frac{1}{2}R^n$) на открытое множество $\mathcal{O}' \subset R^n$ (соответственно $\frac{1}{2}R^n$) называется отображением класса C^r, или, короче, C^r-отображением, если координаты $(x'^1, x'^2, \ldots, x'^m)$ точки $\phi(p)$ в $\mathcal{O}'$, являющейся образом точки p в $\mathcal{O}$, есть r-кратно непрерывно дифференцируемые функции (т. е. их r-е производные существуют и непрерывны) координат $(x^1, x^2, \ldots, x^n)$ точки p. Если отображение является отображением класса C^r для любого $r \geqslant 0$, то говорят, что это отображение класса C^∞, или C^∞-отображение. Под C^0-отображением мы подразумеваем непрерывное отображение.

Говорят, что функция f на открытом множестве $\mathcal{O} \subset R^n$ локально удовлетворяет условию Липшица, если для каждого открытого множества $\mathcal{U} \subset \mathcal{O}$ с компактным замыканием существует постоянная K, такая, что для любой пары точек $p, q \in \mathcal{U}$. $|f(p) - f(q)| \leqslant K|p - q|$, где $|p|$ означает

$$\{(x^1(p))^2 + (x^2(p))^2 + \ldots + (x^n(p))^2\}^{1/2}.$$

Отображение ϕ будем называть удовлетворяющим локально условию Липшица и обозначать это как C^{1-}, если координаты точки $\phi(p)$ как функции координат точки p локально удовлетворяют условию Липшица. Аналогично, будем говорить, что ϕ есть C^r-отображение, если оно класса C^r и производные $(r-1)$-го порядка координат $\phi(p)$ являются функциями координат p, локально удовлетворяющими условию Липшица. В дальнейшем мы обычно рассматриваем функции класса C^r, но аналогичные определения и результаты справедливы и для функций класса C^{r-}.

Пусть $\mathscr{P}$ есть произвольное множество в R^n (соответственно в $\frac{1}{2}R^n$); тогда отображение ϕ из $\mathscr{P}$ в некоторое множество $\mathscr{P}' \subset R^m$ (соответственно $\frac{1}{2}R^m$) называется C^r-отображением, если ϕ является сужением на $\mathscr{P}$ и $\mathscr{P}'$ C^r-отображения некоторого открытого множества $\mathscr{O}$, содержащего $\mathscr{P}$, на некоторое открытое множество $\mathscr{O}'$, содержащее $\mathscr{P}'$.

n-Мерное многообразие $\mathscr{M}$ класса C^r, или C^r-многообразие, есть множество $\mathscr{M}$ вместе с *атласом* $\{\mathscr{U}_\alpha, \phi_\alpha\}$ класса C^r, т. е. с набором карт $\{\mathscr{U}_\alpha, \phi_\alpha\}$, где $\mathscr{U}_\alpha$ — подмножества $\mathscr{M}$, а ϕ_α — взаимно-однозначные отображения соответствующих $\mathscr{U}_\alpha$ на такие открытые множества в R^n, что

1) $\mathscr{U}_\alpha$ образуют покрытие $\mathscr{M}$, т. е. $\mathscr{M} = \bigcup\limits_\alpha \mathscr{U}_\alpha$

2) если $\mathscr{U}_\alpha \cap \mathscr{U}_\beta$ не пусто, то отображение

$$\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(\mathscr{U}_\alpha \cap \mathscr{U}_\beta) \to \phi_\alpha(\mathscr{U}_\alpha \cap \mathscr{U}_\beta)$$

есть C^r-отображение некоторого открытого подмножества R^n на открытое подмножество R^n (рис. 4).

Каждое $\mathscr{U}_\alpha$ есть локально-координатная окрестность в том смысле, что в ней определены локальные координаты x^a, $a = 1, \ldots, n$, заданные отображением ϕ_α [т. е. если $p \in \mathscr{U}_\alpha$, то координатами точки p являются координаты точки $\phi_\alpha(p) \in R^n$].

Условие (2) требует, чтобы в пересечении двух координатных окрестностей координаты одной окрестности были функциями класса C^r координат другой окрестности, и наоборот.

Какой-либо другой атлас называется *совместным* с данным C^r-атласом, если их объединение есть атлас класса C^r для всего многообразия $\mathscr{M}$. Атлас, состоящий из всех атласов, совместных с данным, называется *полным атласом* многообразия. Таким образом, полный атлас есть множество всех возможных систем координат, покрывающих $\mathscr{M}$. Топология в $\mathscr{M}$ определяется требованием, чтобы все открытые множества $\mathscr{M}$ состояли из объединений множества вида $\mathscr{U}_\alpha$, принадлежащих полному атласу. В такой топологии каждое отображение ϕ_α становится гомеоморфизмом.

Определение дифференцируемого *многообразия класса* C^r *с краем* получается заменой «R^n» на «$\frac{1}{2}R^n$». Тогда край многообразия $\mathscr{M}$, обозначаемый $\partial\mathscr{M}$, определяется как множество всех точек многообразия $\mathscr{M}$, образы которых при отображении ϕ_α лежат на границе области $\frac{1}{2}R^n$ в R^n. $\partial\mathscr{M}$ представляет собой $(n-1)$-мерное многообразие без края.

Эти определения могут показаться излишне сложными, однако простые примеры показывают, что описание пространства

единственной координатной окрестностью, вообще говоря, невозможно. Очевидно, *двумерная евклидова плоскость*, R^2, является многообразием. Декартовы координаты $(x, y; -\infty < x < \infty, -\infty < y < \infty)$ покрывают всю плоскость как одну координатную окрестность, в которой ϕ является тождественным отображением. Полярные координаты r, θ покрывают координатную окрестность $(r > 0, 0 < \theta < 2\pi)$; нужны по меньшей мере две такие координатные окрестности, чтобы покрыть R^2. *Двумерный*

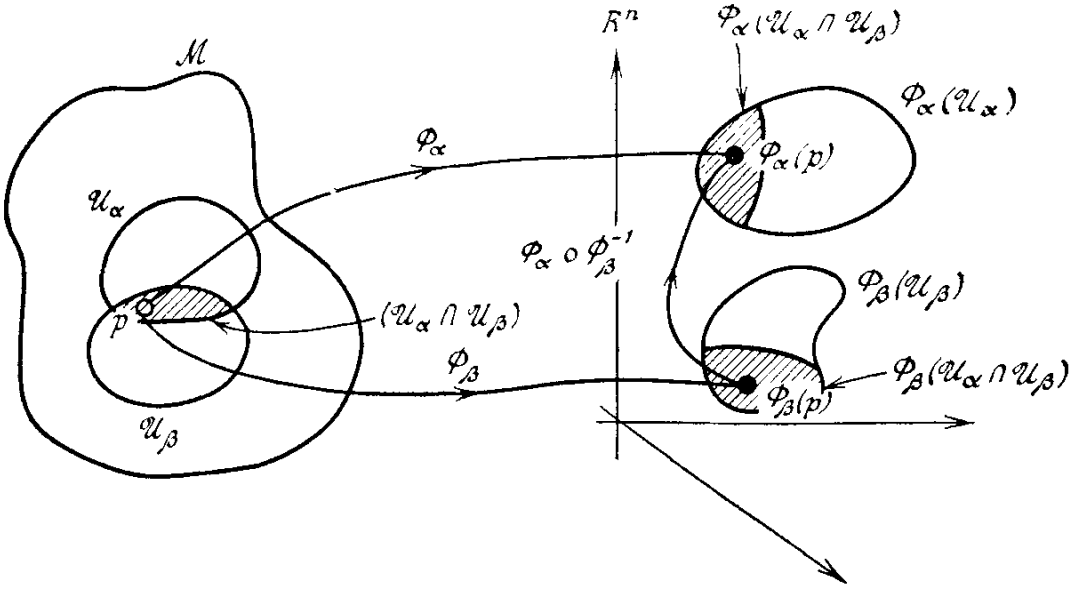


Рис. 4. Там, где кооординатные окрестности $\mathcal{U}_\alpha$ и $\mathcal{U}_\beta$ пересекаются, координаты связаны C^r-отображением $\phi_\alpha \circ \phi_\beta{}^{-1}$.

цилиндр C^2 представляет собой многообразие, получающееся из R^2 отождествлением точек (x, y) и $(x + 2\pi, y)$. Тогда (x, y) являются координатами в окрестности $(0 < x < 2\pi; -\infty < y < \infty)$, и, чтобы покрыть C^2, нужны две такие координатные окрестности. *Лист Мёбиуса* является многообразием, которое получается подобным же образом: отождествляются точки (x, y) и $(x+2\pi, -y)$. *Двумерную сферу* (2-сферу) S^2 единичного радиуса можно определить как поверхность в R^3, задаваемую уравнением $(x^1)^2 + (x^2)^2 + (x^3)^2 = 1$. Тогда

$$x^2, \quad x^3; \quad -1 < x^2 < 1, \quad -1 < x^3 < 1,$$

будут координатами в каждой из областей $x^1 > 0$, $x^1 < 0$, и понадобится шесть таких координатных окрестностей, чтобы покрыть всю сферу. В действительности S^2 вообще невозможно покрыть только одной координатной окрестностью. Аналогично, n-мерную сферу (n-сферу) S^n можно определить как множество точек

$$(x^1)^2 + (x^2)^2 + \ldots + (x^{n+1})^2 = 1$$

в R^{n+1}.

Многообразие называется *ориентируемым*, если в полном атласе есть такой атлас $\{\mathcal{U}_\alpha, \phi_\alpha\}$, что в каждом непустом пересечении $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ якобиан $|\partial x^i/\partial x'^j|$, где $x^1, \ldots, x^n$ и $x'^1, \ldots, x'^n$ — соответственно координаты в $\mathcal{U}_\alpha$ и $\mathcal{U}_\beta$, положителен. Примером неориентируемого многообразия является лист Мёбиуса.

Пока мы дали самое общее определение многообразия. В большинстве случаев полезно наложить еще два условия, а именно, что $\mathcal{M}$ хаусдорфово и паракомпактно. Это гарантирует нам обычные локальные свойства.

Топологическое пространство называется *хаусдорфовым*, если оно удовлетворяет аксиоме разделения Хаусдорфа: для любых двух различных точек $p, q \in \mathcal{M}$ существуют непересекающиеся открытые множества $\mathcal{U}, \mathcal{V} \subset \mathcal{M}$, такие, что $p \in \mathcal{U}$, $q \in \mathcal{V}$.

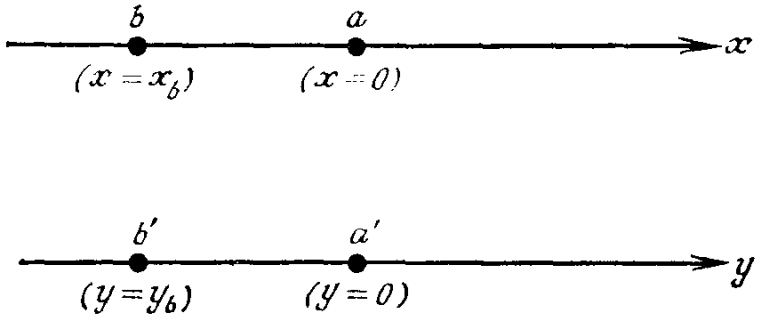


Рис. 5. Пример нехаусдорфова многообразия. Эти две линии отождествлены при $x = y < 0$. Однако точки a $(x = 0)$ и a' $(y = 0)$ не отождествлены.

Нельзя считать, что всякое многообразие — непременно хаусдорфово. Рассмотрим, например, ситуацию, изображенную на рис. 5. На двух прямых отождествим те и только те точки b и b', для которых $x_b = y_{b'} < 0$. Тогда любая точка содержится в некоторой (координатной) окрестности, гомеоморфной открытому подмножеству в R^1. Однако не существует непересекающихся открытых окрестностей $\mathcal{U}$, $\mathcal{V}$, удовлетворяющих условию $a \in \mathcal{U}$, $a' \in \mathcal{V}$, где a — точка $x = 0$ и a' — точка $y = 0$.

Атлас $\{\mathcal{U}_\alpha, \phi_\alpha\}$ называют *локально-конечным*, если для любой точки $p \in \mathcal{M}$ найдется открытая окрестность, которая пересекается лишь с конечным числом множеств $\mathcal{U}_\alpha$. $\mathcal{M}$ называют *паракомпактным*, если для любого атласа $\{\mathcal{U}_\alpha, \phi_\alpha\}$ существует локально-конечный атлас $\{\mathcal{V}_\beta, \psi_\beta\}$, каждое $\mathcal{V}_\alpha$ которого содержится в некотором $\mathcal{U}_\alpha$. Связное хаусдорфово многообразие паракомпактно тогда и только тогда, когда оно обладает счетным базисом, т. е. существует такая счетная система открытых множеств, что любое открытое множество можно представить как объединение множеств из этой системы ([91], p. 271).

Если не оговаривается иное, *все рассматриваемые далее многообразия предполагаются паракомпактными связными хаусдор-*

фовыми многообразиями класса C^∞ без края. Позднее мы увидим, что при наделении $\mathcal{M}$ дополнительной структурой (о существовании аффинной связности см. разд. 2.4) требование паракомпактности выполняется автоматически в силу прочих ограничений.

Функция f на C^k-многообразии $\mathcal{M}$ есть отображение из $\mathcal{M}$ в R^1. Говорят, что она класса C^r $(r \leqslant k)$ в точке $p \in \mathcal{M}$, если ее представление $f \circ \phi_\alpha^{-1}$ в некоторой локально-координатной окрестности $\mathcal{U}_\alpha$ есть функция класса C^r (C^r-функция) локальных координат точки p; f называется *C^r-функцией* на множестве $\mathcal{V} \subset \mathcal{M}$, если f есть C^r-функция в каждой точке $p \in \mathcal{V}$.

В дальнейшем мы воспользуемся следующим свойством паракомпактных многообразий (см. [91], p. 272). Пусть $\{\mathcal{U}_\alpha, \phi_\alpha\}$ — любой локально-конечный атлас на паракомпактном C^k-многообразии; всегда можно найти систему функций g_α класса C^k, таких, что

1) $0 \leqslant g_\alpha \leqslant 1$ на $\mathcal{M}$ при любом α;

2) носитель g_α, т. е. замыкание множества $\{p \in \mathcal{M}: g_\alpha(p) \neq 0\}$, содержится в соответствующем $\mathcal{U}_\alpha$;

3) $\sum_\alpha g_\alpha(p) = 1$ для всех $p \in \mathcal{M}$.

Такую систему функций будем называть *разложением единицы.* Этот результат справедлив, в частности, для функций класса C^∞, но он, очевидно, не имеет места для аналитических функций (аналитическую функцию можно представить в виде сходящегося степенного ряда в некоторой окрестности каждой точки $p \in \mathcal{M}$, причем если она равна нулю в некоторой открытой окрестности, то она равна нулю всюду).

Наконец, *прямое* произведение $\mathcal{A} \times \mathcal{B}$ многообразий $\mathcal{A}$, $\mathcal{B}$ есть многообразие со структурой, которая естественным образом определяется структурой $\mathcal{A}$ и $\mathcal{B}$: для произвольных точек $p \in \mathcal{A}$, $q \in \mathcal{B}$ существуют координатные окрестности $\mathcal{U}$, $\mathcal{V}$, содержащие соответственно p и q, такие, что точка $(p, q) \in \mathcal{A} \times \mathcal{B}$ содержится в координатной окрестности $\mathcal{U} \times \mathcal{V} \subset \mathcal{A} \times \mathcal{B}$ и имеет в ней координаты (x^i, y^j), где x^i — координаты p в $\mathcal{U}$ и y^j — координаты q в $\mathcal{V}$.

2.2. Векторы и тензоры

Тензорные поля представляют собой множество геометрических объектов на многообразии, которые определяются естественным образом структурой многообразия. Тензорное поле эквивалентно заданию тензора в каждой точке многообразия. Поэтому в первую очередь мы определим тензоры в точке многообразия, а начнем с основного понятия вектора в точке.

Кривая $\lambda(t)$ класса C^k, или, короче, C^k-кривая на $\mathscr{M}$, есть отображение класса C^k некоторого интервала вещественной прямой R^1 в $\mathscr{M}$. *Вектор* (контравариантный вектор) $(\partial/\partial t)_\lambda|_{t_0}$, касательный к C^1-кривой $\lambda(t)$ в точке $\lambda(t_0)$, есть оператор, который ставит в соответствие каждой C^1-функции f в $\lambda(t_0)$ число $(\partial f/\partial t)_\lambda|_{t_0}$. Иначе говоря, $(\partial f/\partial t)_\lambda$ есть производная f в направлении $\lambda(t)$ по параметру t:

$$\left(\frac{\partial f}{\partial t}\right)_\lambda\Big|_t = \lim_{s\to 0}\frac{1}{s}\{f(\lambda(t+s)) - f(\lambda(t))\}. \tag{2.1}$$

Сам параметр t вдоль кривой, очевидно, удовлетворяет соотношению $(\partial/\partial t)_\lambda t = 1$.

Пусть $(x^1, \ldots, x^n)$ — локальные координаты в некоторой окрестности точки p; тогда

$$\left(\frac{\partial f}{\partial t}\right)_\lambda\Big|_{t_0} = \sum_{j=1}^{n}\frac{\mathrm{d}x^j(\lambda(t))}{\mathrm{d}t}\Big|_{t=t_0}\cdot\frac{\partial f}{\partial x^j}\Big|_{\lambda(t_0)} = \frac{\mathrm{d}x^j}{\mathrm{d}t}\,\frac{\partial f}{\partial x^j}\Big|_{\lambda(t_0)}.$$

(Здесь и повсюду в книге мы пользуемся *правилом суммирования*, согласно которому по повторяющемуся индексу подразумевается суммирование по всем значениям этого индекса.) Таким образом, любой касательный вектор в точке p можно представить в виде линейной комбинации производных по координатам. Обратно, пусть дана линейная комбинация $V^j(\partial/\partial x^j)|_p$, где V^j — какие-то числа; рассмотрим кривую $\lambda(t)$, определяемую уравнением $x^j(\lambda(t)) = x^j(p) + tV^j$ для t из некоторого интервала $[-\varepsilon, \varepsilon]$; тогда касательным вектором в этой кривой точке p является $V^j(\partial/\partial x^j)_p$. Следовательно, касательные векторы в p образуют векторное пространство над R^1, натянутое на координатные производные $(\partial/\partial x^j)|_p$; структура векторного пространства задана соотношением

$$(\alpha\mathbf{X} + \beta\mathbf{Y})f = \alpha(\mathbf{X}f) + \beta(\mathbf{Y}f),$$

справедливым для всех векторов $\mathbf{X}$, $\mathbf{Y}$, чисел α, β и функций f. Векторы $(\partial/\partial x^j)_p$ линейно независимы. (Допустим, что они зависимы, т. е. существуют числа V^j, не все равные нулю, такие, что $V^j(\partial/\partial x^j)|_p = 0$; тогда, применив это соотношение к каждой координате, мы получим равенство

$$V^l\frac{\partial x^k}{\partial x^l} = V^k = 0$$

т. е. противоречие.) Таким образом, пространство всех касательных к $\mathscr{M}$ в точке p векторов, обозначаемое $T_p(\mathscr{M})$ или просто T_p, есть n-мерное векторное пространство. Это пространство, представляющее собой множество всех направлений в p, называется

касательным к $\mathcal{M}$ векторным пространством в p. Можно представлять себе вектор $\mathbf{V} \in T_p$ в виде стрелки в p, направленной вдоль кривой $\lambda(t)$, т. е. вдоль касательного вектора $\mathbf{V}$ в точке p; при этом «длина» $\mathbf{V}$ зависит от выбора параметра t вдоль кривой и определяется соотношением $V(t) = 1$. [Жирным шрифтом отмечено, что $\mathbf{V}$ — оператор; его компоненты V^j и $V(f)$ — результат действия $\mathbf{V}$ на функцию f, — являются числами и печатаются курсивом.]

Пусть $\{\mathbf{E}_a\}$ $(a = 1, \ldots, n)$ — какая-либо совокупность n линейно независимых векторов в точке p. Тогда любой вектор $\mathbf{V} \in T_p$ можно записать в виде $\mathbf{V} = V^a \mathbf{E}_a$, где числа $\{V^a\}$ суть компоненты $\mathbf{V}$ относительно базиса $\{\mathbf{E}_a\}$ векторов в p. (В частности, в качестве $\{\mathbf{E}_\alpha\}$ можно выбрать координатный базис $(\partial/\partial x^i)_p$; тогда компонентами $V^i = V(x^i) = (dx^i/dt)\,|_p$ будут производные координат x^i в направлении $\mathbf{V}$.)

Линейная форма *) (ковариантный вектор) $\boldsymbol{\omega}$ в p есть вещественная линейная функция на пространстве T_p. Число, которое линейная форма $\boldsymbol{\omega}$ ставит в соответствие вектору $\mathbf{X}$ в точке p, будем записывать в виде $\langle \boldsymbol{\omega}, \mathbf{X} \rangle$; тогда из линейности следует, что при всех $\alpha, \beta \in R^1$ и $\mathbf{X}, \mathbf{Y} \in T_p$ выполняется равенство

$$\langle \boldsymbol{\omega}, \alpha\mathbf{X} + \beta\mathbf{Y} \rangle = \alpha \langle \boldsymbol{\omega}, \mathbf{X} \rangle + \beta \langle \boldsymbol{\omega}, \mathbf{Y} \rangle.$$

Подпространство в T_p, определяемое при заданной линейной форме $\boldsymbol{\omega}$ уравнением $\langle \boldsymbol{\omega}, \mathbf{X} \rangle = \text{const}$, линейно. Поэтому линейную форму в точке p можно представлять себе как пару плоскостей в T_p, таких, что вектор $\mathbf{X}$ лежит в одной плоскости, если $\langle \boldsymbol{\omega}, \mathbf{X} \rangle = 0$, и касается концом другой, если $\langle \boldsymbol{\omega}, \mathbf{X} \rangle = 1$.

Пусть дан базис $\{\mathbf{E}_a\}$ векторов в p; тогда можно построить такую единственную систему $\{\mathbf{E}^a\}$ из n линейных форм, что линейная форма $\mathbf{E}^i$ отображает любой вектор $\mathbf{X}$ в число X^i (i-ю компоненту X относительно базиса $\{\mathbf{E}_a\}$). При этом, в частности, $\langle \mathbf{E}^a, \mathbf{E}_b \rangle = \delta^a{}_b$. Определив линейные комбинации линейных форм равенством

$$\langle \alpha\boldsymbol{\omega} + \beta\boldsymbol{\eta}, \mathbf{X} \rangle = \alpha \langle \boldsymbol{\omega}, \mathbf{X} \rangle + \beta \langle \boldsymbol{\eta}, \mathbf{X} \rangle$$

для любых линейных форм $\boldsymbol{\omega}$, $\boldsymbol{\eta}$ и произвольных α, $\beta \in R^1$, $\mathbf{X} \in T_p$, можно рассматривать $\{\mathbf{E}^a\}$ как базис для линейных форм, ибо любую линейную форму можно записать в виде $\boldsymbol{\omega} = \omega_i \mathbf{E}$, где $\omega_i = \langle \boldsymbol{\omega}, \mathbf{E}_i \rangle$. Таким образом, множество всех линейных форм в точке p образует n-мерное векторное *пространство* T^*_p, *дуальное* касательному пространству. Базис линейных

*) В оригинале «one-form» (1-форма). Мы пользуемся здесь более привычным термином «линейная форма», но после введения на стр. 30 термина «*p-форма*», мы тоже будем называть линейную форму 1-формой. — *Прим. перев.*

форм $\{\mathbf{E}^a\}$ есть *базис*, *дуальный* базису векторов $\{\mathbf{E}_a\}$. Для любых $\boldsymbol{\omega} \in T^*_p$, $\mathbf{X} \in T_p$ число $\langle \boldsymbol{\omega}, \mathbf{X} \rangle$ можно выразить через компоненты ω_i и X^i в дуальных базисах $\{\mathbf{E}^a\}$, $\{\mathbf{E}_a\}$:

$$\langle \boldsymbol{\omega}, \mathbf{X} \rangle = \langle \omega_i \mathbf{E}^i, X^j \mathbf{E}_j \rangle = \omega_i X^i.$$

Любая функция f на $\mathscr{M}$ определяет в p линейную форму $\mathrm{d}f$ по правилу: для произвольного вектора $\mathbf{X}$

$$\langle \mathrm{d}f, \mathbf{X} \rangle = Xf.$$

$\mathrm{d}f$ называется *дифференциалом* f. Пусть $x^1, \ldots, x^n$ — локальные координаты; совокупность дифференциалов $(\mathrm{d}x^1, \mathrm{d}x^2, \ldots, \mathrm{d}x^n)$ в точке p образует базис линейных форм, дуальный базису векторов $(\partial/\partial x^1, \partial/\partial x^2, \ldots, \partial/\partial x^n)$ в p, поскольку

$$\langle \mathrm{d}x^i, \partial/\partial x^j \rangle = \frac{\partial x^i}{\partial x^j} = \delta^i{}_j.$$

Дифференциал $\mathrm{d}f$ произвольной функции f записывается в этом базисе в виде

$$\mathrm{d}f = \frac{\partial f}{\partial x^i}\, \mathrm{d}x^i.$$

Если $\mathrm{d}f \neq 0$, поверхности $f = \mathrm{const}$ являются $(n-1)$-мерными многообразиями. Множество всех векторов $\mathbf{X}$, для которых $\langle \mathrm{d}f, \mathbf{X} \rangle = 0$, образует подпространство T_p, состоящее из всех векторов, которые касательны в точке p к кривым, лежащим на поверхности $f = \mathrm{const}$. Следовательно, $\mathrm{d}f$ можно рассматривать как нормаль к поверхности $f = \mathrm{const}$ в p. Если $\alpha \neq 0$, то $\alpha \mathrm{d}t$ также будет нормалью к этой поверхности.

Из пространства T_p векторов в p и пространства T^*_p линейных форм в p мы можем образовать прямое произведение

$$\Pi^s_r = \underbrace{T^*_p \times T^*_p \times \ldots \times T^*_p}_{r \text{ сомножителей}} \times \underbrace{T_p \times T_p \times \ldots \times T_p}_{s \text{ сомножителей}},$$

т. е. упорядоченное множество векторов и линейных форм $(\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^r, \mathbf{Y}_1, \ldots, \mathbf{Y}_s)$, где $\mathbf{Y}$ — произвольный вектор, $\boldsymbol{\eta}$ — произвольная линейная форма.

Тензор типа (r, s) в точке p есть функция на Π^s_r, линейная по каждому из аргументов. Пусть $\mathbf{T}$ — тензор типа (r, s) в p; он ставит в соответствие элементу $(\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^r, \mathbf{Y}_1, \ldots, \mathbf{Y}_s)$ пространства Π^s_r число, которое запишем в виде

$$T(\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^r, \mathbf{Y}_1, \ldots, \mathbf{Y}_s).$$

Тогда линейность означает, что

$$T(\eta^1, \ldots, \eta^r, \alpha\mathbf{X}+\beta\mathbf{Y}, \mathbf{Y}_2, \ldots, \mathbf{Y}_s) = \\ = \alpha \cdot T(\eta^1, \ldots, \eta^r, \mathbf{X}, \mathbf{Y}_2, \ldots, \mathbf{Y}_s) + \\ + \beta \cdot T(\eta^1, \ldots, \eta^r, \mathbf{Y}, \mathbf{Y}_2, \ldots, \mathbf{Y}_s)$$

при любых $\alpha, \beta \in R^1$ и $\mathbf{X}, \mathbf{Y} \in T_p$.

Пространство всех таких тензоров называется *тензорным произведением*

$$T_s^r(p) = \underbrace{T_p \otimes \ldots \otimes T_p}_{r \text{ сомножителей}} \otimes \underbrace{T^*_p \otimes \ldots \otimes T^*_p}_{s \text{ сомножителей}}.$$

В частности, $T_0^1(p) = T_p$ и $T_1^0(p) = T^*_p$.

Сложение тензоров типа (r, s) определяется следующим образом: $(\mathbf{T}+\mathbf{T}')$ есть тензор типа (r, s) в точке p, такой, что для произвольных $\mathbf{Y}_i \in T_p$, $\eta^j \in T^*_p$

$$(T+T')(\eta^1, \ldots, \eta^r, \mathbf{Y}_1, \ldots, \mathbf{Y}_s) = \\ = T(\eta^1, \ldots, \eta^r, \mathbf{Y}_1, \ldots, \mathbf{Y}_s) + T'(\eta^1, \ldots, \eta^r, \mathbf{Y}_1, \ldots, \mathbf{Y}_s).$$

Аналогично определяется *умножение тензора на скаляр* α: $\alpha\mathbf{T}$ есть тензор, такой, что для любых $\mathbf{Y}_i \in T_p$, $\eta^j \in T^*_p$

$$\alpha\mathbf{T}(\eta^1, \ldots, \eta^2, \mathbf{Y}_1, \ldots, \mathbf{Y}_s) = \alpha \cdot \mathbf{T}(\eta^1, \ldots, \eta^r, \mathbf{Y}_1, \ldots, \mathbf{Y}_s).$$

В сочетании с этими правилами сложения и умножения тензорное произведение $T_s^r(p)$ является векторным пространством размерности n^{r+s} над R^1.

Пусть $\mathbf{X}_i \in T_p$ $(i = 1, \ldots, r)$ и $\omega^j \in T^*_p$ $(j = 1, \ldots, s)$. Тогда мы будем обозначать через $\mathbf{X}_1 \otimes \ldots \otimes \mathbf{X}_r \otimes \omega^1 \otimes \ldots \otimes \omega^s$ такой элемент из $T_s^r(p)$, который ставит в соответствие элементу $(\eta^1, \ldots, \eta^r, \mathbf{Y}_1, \ldots, \mathbf{Y}_s)$ из Π_r^s число

$$\langle \eta^1, \mathbf{X}_1 \rangle \langle \eta^2, \mathbf{X}_2 \rangle \ldots \langle \eta^r, \mathbf{X}_r \rangle \langle \omega^1, \mathbf{Y}_1 \rangle \ldots \langle \omega^s, \mathbf{Y}_s \rangle.$$

Аналогично, если $\mathbf{R} \in T_s^r(p)$ и $\mathbf{S} \in T_q^p(p)$, будем обозначать через $\mathbf{R} \otimes \mathbf{S}$ такой элемент пространства $T_{q+s}^{p+r}(p)$, который ставит в соответствие элементу $(\eta^1, \ldots, \eta^r, \mathbf{Y}_1, \ldots, \mathbf{Y}_{s+q})$ из Π_{p+r}^{s+q} число

$$R(\eta^1, \ldots, \eta^s, \mathbf{Y}_1, \ldots, \mathbf{Y}_r) S(\eta^{s+1}, \ldots, \eta^{s+q}, \mathbf{Y}_{r+1}, \ldots, \mathbf{Y}_{r+p}).$$

Тензорные пространства в p вместе с операцией умножения $\otimes$ образуют алгебру над R^1.

Если $\{\mathbf{E}_a\}$, $\{\mathbf{E}^a\}$ — дуальные базисы соответственно в T_p и T^*_p, то

$$\{\mathbf{E}_{a_1} \otimes \ldots \otimes \mathbf{E}_{a_r} \otimes \mathbf{E}^{b_1} \otimes \ldots \otimes \mathbf{E}^{b_s}\}, \quad a_i, b_j = 1, \ldots, n,$$

будет базисом в $T^r_s(p)$. Произвольный тензор $\mathbf{T} \in T^r_s(p)$ можно записать в этом базисе как

$$\mathbf{T} = T^{a_1 \dots a_r}{}_{b_1 \dots b_s} \mathbf{E}_{a_1} \otimes \dots \otimes \mathbf{E}_{a_r} \otimes \mathbf{E}^{b_1} \otimes \dots \otimes \mathbf{E}^{b_s},$$

где $T^{a_1 \dots a_r}{}_{b_1 \dots b_s}$ — *компоненты* $\mathbf{T}$ относительно дуальных базисов $\{\mathbf{E}_a\}$, $\{\mathbf{E}^a\}$:

$$T^{a_1 \dots a_r}{}_{b_1 \dots b_s} = T\left(\mathbf{E}^{a_1}, \dots, \mathbf{E}^{a_r}, \mathbf{E}_{b_1}, \dots, \mathbf{E}_{b_s}\right).$$

Соотношения тензорной алгебры в p можно написать для компонент тензоров:

$$(T + T')^{a_1 \dots a_r}{}_{b_1 \dots b_s} = T^{a_1 \dots a_r}{}_{b_1 \dots b_s} + T'^{a_1 \dots a_r}{}_{b_1 \dots b_s};$$

$$(\alpha T)^{a_1 \dots a_r}{}_{b_1 \dots b_s} = \alpha \cdot T^{a_1 \dots a_r}{}_{b_1 \dots b_s};$$

$$(T \otimes T')^{a_1 \dots a_{r+p}}{}_{b_1 \dots b_{s+q}} = T^{a_1 \dots a_r}{}_{b_1 \dots b_s} T'^{a_{r+1} \dots a_{r+p}}{}_{b_{s+1} \dots b_{s+q}}.$$

Обычно мы будем представлять тензорные соотношения именно в такой записи ввиду ее удобства.

Если $\{\mathbf{E}_{a'}\}$ и $\{\mathbf{E}^{a'}\}$ — другая пара дуальных базисов в T_p и T^*_p, их элементы можно записать в базисах $\{\mathbf{E}_a\}$, $\{\mathbf{E}^a\}$ в виде

$$\mathbf{E}_{a'} = \Phi_{a'}{}^a \mathbf{E}_a, \tag{2.2}$$

$$\mathbf{E}^{a'} = \Phi^{a'}{}_a \mathbf{E}^a, \tag{2.3}$$

где $\Phi_{a'}{}^a$ и $\Phi^{a'}{}_a$ — несингулярные матрицы $n \times n$. Поскольку $\{\mathbf{E}_{a'}\}$, $\{\mathbf{E}^{a'}\}$ — дуальные базисы,

$$\delta^{b'}{}_{a'} = \langle \mathbf{E}^{b'}, \mathbf{E}_{a'} \rangle = \langle \Phi^{b'}{}_b \mathbf{E}^b, \Phi_{a'}{}^a \mathbf{E}_a \rangle = \Phi_{a'}{}^a \Phi^{b'}{}_b \delta_a{}^b = \Phi_{a'}{}^a \Phi^{b'}{}_a,$$

т. е. $\Phi_{a'}{}^a$, $\Phi^{a'}{}_a$ — взаимно-обратные матрицы и $\delta^a{}_b = \Phi^a{}_{b'} \Phi^{b'}{}_b$.

Компонентами $T^{a'_1 \dots a'_r}{}_{b'_1 \dots b'_s}$ тензора $\mathbf{T}$ относительно дуальных базисов $\{\mathbf{E}_{a'}\}$, $\{\mathbf{E}^{a'}\}$ являются

$$T^{a'_1 \dots a'_r}{}_{b'_r \dots b'_s} = T\left(\mathbf{E}^{a'_1}, \dots, \mathbf{E}^{a'_r}, \mathbf{E}_{b'_1}, \dots, \mathbf{E}_{b'_s}\right).$$

Они связаны с компонентами $T^{a_1 \dots a_r}{}_{b_1 \dots b_s}$ относительно базисов $\{\mathbf{E}_a\}$, $\{\mathbf{E}^a\}$ соотношением

$$T^{a'_1 \dots a'_r}{}_{b'_1 \dots b'_s} = T^{a_1 \dots a_r}{}_{b_1 \dots b_s} \Phi^{a'_1}{}_{a_1} \dots \Phi^{a'_r}{}_{a_r} \Phi_{b'_1}{}^{b_1} \dots \Phi_{b'_s}{}^{b_s}. \tag{2.4}$$

Свертка тензора $\mathbf{T}$ типа (r, s) с компонентами $T^{ab \dots d}{}_{ef \dots g}$ относительно базисов $\{\mathbf{E}_a\}$, $\{\mathbf{E}^a\}$ по первому контравариантному

и первому ковариантному индексам есть по определению тензор $C_1^1(\mathbf{T})$ типа $(r-1, s-1)$, компоненты которого относительно тех же базисов равны $T^{ab\ldots d}{}_{af\ldots g}$, т. е.

$$C_1^1(\mathbf{T}) = T^{ab\ldots d}{}_{af\ldots g}\mathbf{E}_b \otimes \ldots \otimes \mathbf{E}_d \otimes \mathbf{E}^f \otimes \ldots \otimes \mathbf{E}^g.$$

Если $\{\mathbf{E}_{a'}\}$, $\{\mathbf{E}^{a'}\}$ — другая пара базисов, то определяемая ею свертка равна

$$C'^1_1(\mathbf{T}) = T^{a'b'\ldots d'}{}_{a'f'\ldots g'}\mathbf{E}_{b'} \otimes \ldots \otimes \mathbf{E}_{d'} \otimes \mathbf{E}^{f'} \otimes \ldots \otimes \mathbf{E}^{g'} =$$
$$= \Phi^{a'}{}_a \Phi^a{}_{h'} T^{h'b'\ldots d'}{}_{a'f'\ldots g'} \Phi_{b'}{}^b \ldots \Phi_{d'}{}^d \Phi^{f'}{}_f \ldots \Phi^{g'}{}_g \cdot \mathbf{E}_b \otimes \ldots$$
$$\ldots \otimes \mathbf{E}_d \otimes \mathbf{E}^f \otimes \ldots \otimes \mathbf{E}^g =$$
$$= T^{ab\ldots d}{}_{af\ldots g}\mathbf{E}_b \otimes \ldots \otimes \mathbf{E}_d \otimes \mathbf{E}^f \otimes \ldots \otimes \mathbf{E}^g = C_1^1(\mathbf{T}),$$

так что свертка тензора C_1^1 не зависит от базиса, использованного при ее определении. Подобным же образом можно сворачивать $\mathbf{T}$ по любой паре контравариантных и ковариантных индексов. (Если бы мы свернули по двум контравариантным или по двум ковариантным индексам, то полученный тензор зависел бы от базиса.)

Симметричная часть тензора $\mathbf{T}$ типа (2,0) есть тензор

$$S(\mathbf{T})(\eta_1, \eta_2) = \frac{1}{2!}\{T(\eta_1, \eta_2) + T(\eta_2, \eta_1)\}$$

при произвольных $\eta_1, \eta_2 \in T^*_p$. Будем обозначать компоненты $S(\mathbf{T})^{ab}$ этого тензора через $T^{(ab)}$; тогда

$$T^{(ab)} = \frac{1}{2!}\{T^{ab} + T^{ba}\}.$$

Аналогично, компоненты антисимметричной части T будем обозначать через

$$T^{[ab]} = \frac{1}{2!}\{T^{ab} - T^{ba}\}.$$

Вообще для обозначения компонент симметричной или антисимметричной части тензора по данной совокупности ковариантных или контравариантных индексов мы будем заключать эти индексы в круглые или квадратные скобки. Так,

$$T_{(a_1\ldots a_r)}{}^{b\ldots f} = \frac{1}{r!}\{\text{сумма по всем перестановкам индексов } a_1, \ldots, a_r \text{ в } T_{a_1\ldots a_r}{}^{b\ldots f}\}$$

и

$$T_{[a_1\ldots a_r]}{}^{b\ldots} = \frac{1}{r!}\{\text{альтернированная сумма по всем перестановкам индексов } a_1, \ldots, a_r \text{ в } T_{a_1\ldots a_r}{}^{b\ldots f}\}.$$

Например,

$$K^a{}_{[bcd]} = \frac{1}{6}\{K^a{}_{bcd} + K^a{}_{dbc} + K^a{}_{cdb} - K^a{}_{bdc} - K^a{}_{cbd} - K^a{}_{dcb}\}.$$

Тензор *симметричен* по данной совокупности контравариантных или ковариантных индексов, если он равен своей симметричной части по этим индексам, и *антисимметричен*, если он равен своей антисимметричной части. Так, например, тензор $\mathbf{T}$ типа (0,2) симметричен, если $T_{ab} = \frac{1}{2}(T_{ab} + T_{ba})$; это свойство иначе можно выразить равенством $T_{[ab]} = 0$.

Особый интерес представляют тензоры типа $(0, q)$, антисимметричные по всем q индексам (при этом необходимо $q \leqslant n$); такой тензор называется *q-формой*. Если $\mathbf{A}$ и $\mathbf{B}$ соответственно p- и q-формы, можно из них построить $(p+q)$-форму $\mathbf{A} \wedge \mathbf{B}$, где $\wedge$ означает антисимметризованное тензорное произведение $\otimes$; иначе говоря, $\mathbf{A} \wedge \mathbf{B}$ есть тензор типа $(0, p+q)$ с компонентами

$$(A \wedge B)_{a \ldots bc \ldots f} = A_{[a \ldots b} B_{c \ldots f]}.$$

Из этого следует, что $\mathbf{A} \wedge \mathbf{B} = (-1)^{pq}(\mathbf{B} \wedge \mathbf{A})$. Пространство внешних форм (т. е. пространство всех p-форм при всех p, *включая линейные формы в качестве 1-форм и скаляры* в качестве 0-форм) вместе с операцией умножения образуют грассманову алгебру (внешних) форм. Пусть $\{\mathbf{E}^a\}$ — базис линейных форм; тогда формы $\mathbf{E}^{a_1} \wedge \ldots \wedge \mathbf{E}^{a_p}$ $(a_i = 1, \ldots, n)$ образуют базис p-форм, поскольку любая p-форма $\mathbf{A}$ может быть представлена в виде $\mathbf{A} = A_{a \ldots b}\mathbf{E}^a \wedge \ldots \wedge \mathbf{E}^b$, где $A_{a \ldots b} = A_{[a \ldots b]}$.

До сих пор мы рассматривали тензоры, определенные в некоторой фиксированной точке многообразия. Система локальных координат $\{x^i\}$ на открытом множестве $\mathcal{U}$ в $\mathcal{M}$ задает базис векторов $\{(\partial/\partial x^i)|_p\}$ и базис линейных форм (1-форм $\{(dx^i)|_p\}$), а следовательно, и базис тензоров типа (r, s) в каждой точке $\mathcal{U}$. Такой базис тензоров называется координатным базисом. По определению мы имеем *тензорное поле* $\mathbf{T}$ *типа* (r, s) *класса* C^k (*тензорное C^k-поле*) на множестве $\mathcal{V} \subset \mathcal{M}$, если в каждой точке p множества $\mathcal{V}$ задан элемент пространства $T^r_s(p)$, такой, что компоненты $\mathbf{T}$ относительно любого координатного базиса, определенного на $\mathcal{V}$, являются функциями класса C^k.

Вообще говоря, в использовании координатного базиса тензоров нет необходимости: вовсе не обязательно, чтобы на $\mathcal{V}$ для заданного базиса векторов $\{\mathbf{E}_a\}$ и дуального ему базиса линейных форм $\{\mathbf{E}^a\}$ нашлось такое открытое подмножество, в котором существовали бы локальные координаты $\{x^a\}$, удовлетворяющие условию $\mathbf{E}_a = \partial/\partial x^a$ и $\mathbf{E}^a = dx^a$. Однако использование координатного базиса приводит к некоторым упрощениям, в част-

ности, для любой функции f выполняется соотношение $E_a(E_b f) = E_b(E_a f)$, эквивалентное равенству $\partial^2 f/\partial x^a \partial x^b = \partial^2 f/\partial x^b \partial x^a$. При переходе от координатного базиса $\mathbf{E}_a = \partial/\partial x^a$ к координатному базису $\mathbf{E}_{a'} = \partial/\partial x^{a'}$ в силу (2.2) и (2.3) имеем

$$\Phi_{a'}{}^{a} = \frac{\partial x^a}{\partial x^{a'}}, \qquad \Phi^{a'}{}_{a} = \frac{\partial x^{a'}}{\partial x^a}.$$

Ясно, что базис общего вида $\{\mathbf{E}_a\}$ можно получить из координатного базиса $\{\partial/\partial x^i\}$, задав функции $E_a{}^i$, компоненты $\mathbf{E}_a$ в базисе $\{\partial/\partial x^i\}$; тогда (2.2) и (2.3) запишутся соответственно в виде $\mathbf{E}_a = E^i{}_a \partial/\partial x^i$ и $\mathbf{E}^a = E^a{}_i\, dx^i$, причем матрица $E^a{}_i$ дуальна матрице $E_a{}^i$.

2.3. Отображение многообразий

В этом параграфе, пользуясь общим понятием отображения C^k-многообразия, мы дадим определение понятий вложения, погружения и индуцированного отображения тензоров. Первые два из них будут полезны позднее при рассмотрении подмногообразий, а третье играет важную роль при исследовании поведения семейств кривых и при установлении свойств симметрии многообразий.

Отображение ϕ из n-мерного C^k-многообразия $\mathcal{M}$ в n'-мерное $C^{r'}$-многообразие называется отображением класса C^r или *C^r-отображением* ($r \leqslant k$, $r \leqslant k'$), если для каких-либо локальных координатных систем в $\mathcal{M}$ и $\mathcal{M}'$ координаты точки-образа $\phi(p) \in \mathcal{M}'$ суть C^r-функции координат точки $p \in \mathcal{M}$. Поскольку такое отображение, вообще говоря, не является взаимно однозначным (например, оно не может быть таковым, если $n > n'$), то в общем случае оно не имеет обратного. Даже, если некоторое C^r-отображение имеет обратное, то последнее, вообще говоря, не будет класса C^r (пусть, например, ϕ есть отображение $R^1 \to R^1$, устанавливающее соответствие $x \to x^3$; тогда ϕ^{-1} недифференцируемо в точке $x = 0$).

Пусть f — функция на $\mathcal{M}'$. Отображение ϕ задает на $\mathcal{M}$ функцию ϕ^*, значением которой в точке p является значение f в точке $\phi(p)$, т. е.

$$\phi^* f(p) = f(\phi(p)). \tag{2.5}$$

Таким образом, когда ϕ отображает точки из $\mathcal{M}$ в $\mathcal{M}'$, ϕ^* линейно отображает функции из $\mathcal{M}'$ в $\mathcal{M}$.

Если $\lambda(t)$ — кривая, проходящая через точку $p \in \mathcal{M}$, то кривая $\phi(\lambda(t))$ в $\mathcal{M}'$, на которую отображается $\lambda(t)$, проходит через точку $\phi(p)$. При $r \geqslant 1$ касательный вектор к этой кривой в точке $\phi(p)$ будем обозначать как $\phi_*(\partial/\partial t)_\lambda|_{\phi(p)}$; его можно рассматривать как образ вектора $(\partial/\partial t)_\lambda|_p$ при отображении ϕ.

Очевидно, ϕ_* есть линейное отображение из $T_p(\mathscr{M})$ в $T_{\phi(p)}(\mathscr{M}')$. Из (2.5) и определения вектора как производной по направлению следует, что отображение векторов ϕ_* можно охарактеризовать соотношением

$$X(\phi^* f)|_p = \phi_* X(f)|_{\phi(p)} \tag{2.6}$$

для каждой C^r-функции ($r \geqslant 1$) в точке $\phi(p)$ и вектора $\mathbf{X}$ в p.

Используя отображение векторов ϕ_* из $\mathscr{M}$ в $\mathscr{M}'$, можно при $r \geqslant 1$ определить линейное отображение 1-форм ϕ^* из $T^*_{\phi(p)}(\mathscr{M}')$ в $T^*_p(\mathscr{M})$, требуя, чтобы свертка вектора и 1-формы при этих отображениях сохранялась. Тогда 1-форма $\mathbf{A} \in T^*_{\phi(p)}$ отображается в 1-форму $\phi^*\mathbf{A} \in T^*_p$, причем для произвольного вектора $\mathbf{X} \in T_p$ имеем

$$\langle \phi^*\mathbf{A}, \mathbf{X} \rangle|_p = \langle \mathbf{A}, \phi_*\mathbf{X} \rangle|_{\phi(p)}.$$

Следствием этого является равенство

$$\phi^*(\mathrm{d}f) = \mathrm{d}(\phi^* f). \tag{2.7}$$

Отображениям ϕ_* и ϕ^* можно придать более широкий смысл отображений контравариантных тензоров из $\mathscr{M}$ в $\mathscr{M}'$ и ковариантных тензоров из $\mathscr{M}'$ в $\mathscr{M}$ соответственно, если задать следующие правила:

$$\phi_* : \mathbf{T} \in T^0_r(p) \to \phi_*\mathbf{T} \in T^r_0(\phi(p)),$$

$$\phi^* : \mathbf{T} \in T^s_0(\phi(p)) \to \phi^*\mathbf{T} \in T^0_s(p),$$

причем для любого $\eta^i \in T^*_{\phi(p)}$

$$T(\phi^*\eta^1, \ldots, \phi^*\eta^r)|_p = \phi_* T(\eta^1, \ldots, \eta^r)|_{\phi(p)},$$

и для любого

$$\phi^* T(\mathbf{X}_1, \ldots, \mathbf{X}_s)|_p = T(\phi_*\mathbf{X}_1, \ldots, \phi_*\mathbf{X}_s)|_{\phi(p)}.$$

При $r \geqslant 1$ говорят, что C^r-отображение ϕ из $\mathscr{M}$ в $\mathscr{M}'$ имеет *ранг* s в точке p, если размерность $\phi_*(T_p(\mathscr{M}))$ равна s. Оно называется *инъективным* в p, если $s = n$ (и, следовательно, $n \leqslant n'$) в p; тогда в T_p нет вектора, который обращается отображением ϕ_* в нуль. ϕ_* называется *сюръективным*, если $s = n'$ (и, следовательно, $n \geqslant n'$).

C^r-отображение ϕ ($r \geqslant 0$) называют *погружением*, если оно и обратное ему отображение локально также являются C^r-отображениями, т. е. если у каждой точки $p \in \mathscr{M}$ существует такая ее окрестность $\mathscr{U}$, что обратное отображение ϕ^{-1}, ограниченное на $\phi(\mathscr{U})$, тоже принадлежит классу C^r. Это означает, что $n \leqslant n'$. В силу теоремы о неявной функции ([161], p. 41), при $r \geqslant 1$ ϕ тогда и только тогда будет погружением, когда оно инъективно в каждой точке $p \in \mathscr{M}$; в этом случае ϕ_* есть изоморфизм T_p на

его образ $\phi_*(T_p) \subset T_{\phi(p)}$. В этом случае говорят, что образ $\phi(\mathcal{M})$ является n-мерным *погруженным подмногообразием* в $\mathcal{M}'$. Это подмногообразие может иметь самопересечения, т. е. ϕ может не быть взаимно-однозначным, если ограничиться малой окрестностью в $\mathcal{M}'$.

Погружение является *вложением*, если оно представляет собой гомеоморфизм на свой образ по индуцированной топологии. Таким образом, вложение есть взаимно-однозначное погружение;

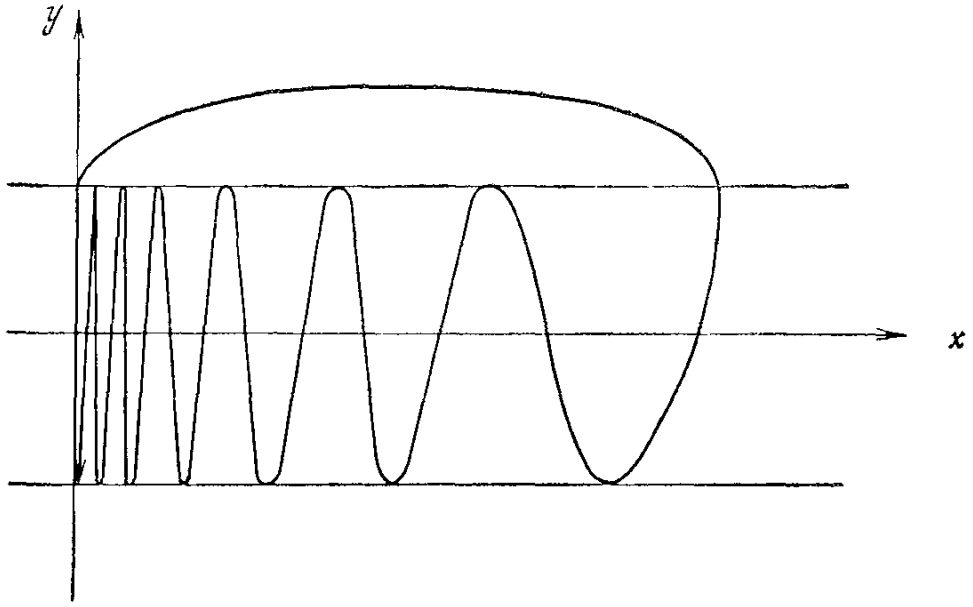


Рис. 6. Взаимно однозначное погружение R^1 в R^2, не являющееся вложением. Оно получено гладким присоединением части кривой $y = \sin(1/x)$ к кривой $\{(y, 0);\ -\infty < y < 1\}$.

однако не всякое взаимно-однозначное погружение есть вложение, см., например, рис. 6. Отображение ϕ называется собственным, если образ любого компактного множества $\mathcal{K} \subset \mathcal{M}'$ при обратном отображении $\phi^{-1}(\mathcal{K})$ компактен. Можно показать, что собственное взаимно-однозначное погружение есть вложение. Образ $\phi(\mathcal{M})$ многообразия $\mathcal{M}$ при отображении ϕ называется *n-мерным вложенным подмногообразием* $\mathcal{M}'$.

Отображение из $\mathcal{M}$ в $\mathcal{M}'$ называют C^r-*диффеоморфизмом*, если оно является взаимно-однозначным C^r-отображением и обратное отображение ϕ^{-1} есть C^r-отображение из $\mathcal{M}'$ в $\mathcal{M}$. В этом случае $n = n'$ и при $r \geqslant 1$ ϕ одновременно и инъективно, и сюръективно. Обратно, из теоремы о неявной функции следует, что если ϕ_* и инъективно, и сюръективно в точке p, то p имеет открытую окрестность $\mathcal{U}$, такую, что $\phi: \mathcal{U} \to \phi(\mathcal{U})$ есть диффеоморфизм. Таким образом, ϕ есть локальный диффеоморфизм в окрестности p, если ϕ_* является изоморфизмом из T_p в $T_{\phi(p)}$.

Если отображение ϕ является C^r-диффеоморфизмом ($r \geqslant 1$), то ϕ_* отображает $T_p(\mathscr{M})$ в $T_{\phi(p)}(\mathscr{M}')$ и $(\phi^{-1})^*$ отображает $T^*_p(\mathscr{M})$ в $T^*_{\phi(p)}(\mathscr{M}')$. Следовательно, мы можем определить отображение ϕ_* пространства $T^r_s(p)$ в $T^r_s(\phi(p))$ при любых (r, s) равенством

$$T(\eta^1, \ldots, \eta^s, \mathbf{X}_1, \ldots, X_r)|_p =$$
$$= \phi_* T((\phi^{-1})^* \eta^1, \ldots, (\phi^{-1})^* \eta^s, \phi_* \mathbf{X}_1, \ldots, \phi_* \mathbf{X}_r)|_{\phi(p)},$$

где $\mathbf{X}_i \in T_p$ и $\eta^i \in T^*_p$ произвольны. Это отображение тензоров типа (r, s) на $\mathscr{M}$ в тензоры типа (r, s) на $\mathscr{M}'$ сохраняет симметрии и соотношения тензорной алгебры; например, свертка $\phi_* \mathbf{T}$ равна отображению ϕ_* свертки $\mathbf{T}$.

2.4. Внешнее дифференцирование и производная Ли

Теперь нам предстоит рассмотреть три дифференциальных оператора на многообразии, два из которых полностью определяются структурой многообразия, а определение третьего (разд. 2.5) связано с введением дополнительной структуры.

Оператор *внешнего дифференцирования* d линейно отображает поле r-формы в поле $(r+1)$-формы. Его действие на поле 0-формы (т. е. на функцию) f дает поле 1-формы $\mathrm{d}f$ по правилу (ср. с разд. 2.2):

$$\langle \mathrm{d}f, \mathbf{X} \rangle = Xf \tag{2.8}$$

для всех векторных полей $\mathbf{X}$; действие d на поле r-формы

$$\mathbf{A} = A_{ab \ldots d}\, \mathrm{d}x^a \wedge \mathrm{d}x^b \wedge \ldots \wedge \mathrm{d}x^d$$

дает поле $(r+1)$-формы $\mathrm{d}\mathbf{A}$:

$$\mathrm{d}\mathbf{A} = \mathrm{d}A_{ab \ldots d} \wedge dx^a \wedge \mathrm{d}x^b \wedge \ldots \wedge dx^d. \tag{2.9}$$

Чтобы убедиться, что это поле $(r+1)$-формы не зависит от выбора координат $\{x^a\}$, использованных при его определении, рассмотрим другую систему координат $\{x^{a'}\}$. Тогда

$$\mathbf{A} = A_{a'b' \ldots d'}\, \mathrm{d}x^{a'} \wedge dx^{b'} \wedge \ldots \mathrm{d}x^{d'},$$

где

$$A_{a'b' \ldots d'} = \frac{\partial x^a}{\partial x^{a'}} \frac{\partial x^b}{\partial x^{b'}} \ldots \frac{\partial x^d}{dx^{d'}} A_{ab \ldots d}.$$

Следовательно, $(r+1)$-формой $\mathrm{d}\mathbf{A}$, определяемой этими координатами, является

$$\mathrm{d}\mathbf{A} = \mathrm{d}A_{a'b'\ldots d'}\,dx^{a'} \wedge \mathrm{d}x^{b'} \wedge \ldots \wedge dx^{d'} =$$

$$= \mathrm{d}\left(\frac{\partial x^a}{\partial x^{a'}}\frac{\partial x^b}{\partial x^{b'}}\ldots\frac{\partial x^d}{\partial x^{d'}}A_{ab\ldots d}\right)\wedge \mathrm{d}x^{a'}\wedge \mathrm{d}x^{b'}\wedge\ldots\mathrm{d}x^{d'} =$$

$$= \frac{\partial x^a}{\partial x^{a'}}\frac{\partial x^b}{\partial x^{b'}}\ldots\frac{\partial x^d}{\partial x^{d'}}\mathrm{d}A_{ab\ldots d}\wedge \mathrm{d}x^{a'}\wedge \mathrm{d}x^{b'}\wedge\ldots\wedge \mathrm{d}x^{d'} +$$

$$+\frac{\partial^2 x^a}{\partial x^{a'}\partial x^{e'}}\frac{\partial x^b}{\partial x^{b'}}\ldots\frac{\partial x^d}{\partial x^{d'}}A_{ab\ldots d}\,\mathrm{d}x^{e'}\wedge \mathrm{d}x^{a'}\wedge dx^{b'}\wedge\ldots\wedge dx^{d'}+\ldots+\ldots$$

$$= \mathrm{d}A_{ab\ldots d}\wedge \mathrm{d}x^a\wedge \mathrm{d}x^b\wedge\ldots\wedge \mathrm{d}x^d$$

ввиду симметрии $\partial^2 x^a/\partial x^{a'}\partial x^{e'}$ и антисимметрии $\mathrm{d}x^{e'}\wedge \mathrm{d}x^{a'}$ по a' и e'. Отметим, что такое определение подходит только для внешних форм: оно не будет независимым от выбора координат, если внешнее умножение $\wedge$ заменить тензорным произведением. Используя равенство $\mathrm{d}(fg) = g\mathrm{d}f + f\mathrm{d}g$, справедливое для произвольных функций f, g, для любых r-формы $\mathbf{A}$ и 1-формы $\mathbf{B}$ получим, что $\mathrm{d}(\mathbf{A}\wedge\mathbf{B}) = \mathrm{d}\mathbf{A}\wedge\mathbf{B} + (-1)^r\mathbf{A}\wedge\mathrm{d}\mathbf{B}$. Согласно (2.8), локально-координатное выражение $\mathrm{d}f$ имеет вид $\mathrm{d}f = (\partial f/\partial x^i)\,\mathrm{d}x^i$; отсюда следует, что $\mathrm{d}(\mathrm{d}f) = (\partial^2 f/\partial x^i\partial x^j)\,\mathrm{d}x^i\wedge \mathrm{d}x^j = 0$, поскольку первый множитель симметричен, а второй антисимметричен. Аналогично, из (2.9) следует, что

$$\mathrm{d}(\mathrm{d}\mathbf{A}) = 0$$

для любого поля r-формы $\mathbf{A}$.

Оператор d коммутирует с отображениями многообразия в следующем смысле: пусть $\phi:\mathscr{M}\to\mathscr{M}'$ есть C^r-отображение $(r\geqslant 2)$ и $\mathbf{A}$ — некоторое C^k-поле внешней формы $(k\geqslant 2)$ на $\mathscr{M}'$; тогда, согласно (2.7),

$$\mathrm{d}(\phi^*\mathbf{A}) = \phi^*(\mathrm{d}\mathbf{A})$$

(что эквивалентно правилу дифференцирования сложных функций).

Оператор d естественным образом появляется в общей формулировке теоремы Стокса на многообразии. Сначала мы определим интегрирование n-формы. Пусть $\mathscr{M}$ — компактное ориентируемое n-мерное многообразие с границей $\partial\mathscr{M}$ и пусть $\{f_\alpha\}$ — разложение единицы для конечного ориентированного атласа $\{\mathscr{U}_\alpha, \phi_\alpha\}$. Тогда интеграл по $\mathscr{M}$ поля n-формы $\mathbf{A}$ на $\mathscr{M}$ определяется как

$$\int_{\mathscr{M}}\mathbf{A} = \sum_\alpha \int_{\phi_\alpha(\mathscr{U}_\alpha)} f_\alpha A_{12\ldots n}\,\mathrm{d}x^1\,\mathrm{d}x^2\ldots\mathrm{d}x^n, \qquad (2.10)$$

где $A_{12\ldots n}$ — компоненты $\mathbf{A}$ относительно локальных координат в координатной окрестности $\mathcal{U}_\alpha$, а интегралы в правой части — это обычные кратные интегралы по открытым множествам $\phi_\alpha(\mathcal{U}_\alpha)$ в R^n. Таким образом интегрирование n-формы по $\mathcal{M}$ сводится к отображению формы в R^n (путем введения локальных координат) и обычному кратному интегрированию в R^n; существование разложения единицы гарантирует глобальную справедливость этой операции.

Интеграл (2.10) является хорошо определенной величиной, поскольку, выбирая другой атлас $\{\mathcal{V}_\beta, \psi_\beta\}$ и разложение единицы $\{g_\beta\}$ для этого атласа, мы получили бы интеграл

$$\sum_\beta \int_{\psi_\beta(\mathcal{V}_\beta)} g_\beta A_{1'2'\ldots n'}\, \mathrm{d}x^{1'}\, \mathrm{d}x^{2'} \ldots \mathrm{d}x^{n'},$$

где $x^{i'}$ — соответствующие локальные координаты. Из сравнения этих двух величин в пересечении $\mathcal{U}_\alpha \cap \mathcal{V}_\beta$ координатных окрестностей, принадлежащих двум атласам, видно, что первую из них можно представить в виде

$$\sum_\alpha \sum_\beta \int_{\phi_\alpha(\mathcal{U}_\alpha \cap \mathcal{V}_\beta)} f_\alpha g_\beta A_{12\ldots n}\, \mathrm{d}x^1\, dx^2 \ldots \mathrm{d}x^n,$$

а вторую в виде

$$\sum_\alpha \sum_\beta \int_{\psi_\beta(\mathcal{U}_\alpha \cap \mathcal{V}_\beta)} f_\alpha g_\beta A_{1'2'\ldots n'}\, \mathrm{d}x^{1'}\, \mathrm{d}x^{2'} \ldots \mathrm{d}x^{n'}.$$

В силу законов преобразования n-формы $\mathbf{A}$ и кратных интегралов в R^n эти выражения равны в окрестности любой точки, и, следовательно, $\int_{\mathcal{M}} \mathbf{A}$ не зависит от выбора атласа и разложения единицы.

Подобным же образом можно показать, что этот интеграл инвариантен относительно диффеоморфизмов: если ϕ есть C^r-диффеоморфизм ($r \geqslant 1$) из $\mathcal{M}$ в $\mathcal{M}'$, то

$$\int_{\mathcal{M}'} \phi_* \mathbf{A} = \int_{\mathcal{M}} \mathbf{A}.$$

Обобщенную теорему Стокса можно сформулировать теперь с помощью оператора d следующим образом: пусть $\mathbf{B}$ — поле $(n-1)$-формы на $\mathcal{M}$, тогда

$$\int_{\partial\mathcal{M}} \mathbf{B} = n \int_{\mathcal{M}} \mathrm{d}\mathbf{B},$$

в чем можно убедиться (см., например, [161]), используя данные выше определения; по существу мы получили общую формулировку этой фундаментальной теоремы анализа. Чтобы выполнить интегрирование в левой части, нужно задать ориентацию границы $\partial\mathcal{M}$ многообразия $\mathcal{M}$. Это делается следующим образом. Пусть $\mathcal{U}_\alpha$ — такая координатная окрестность из ориентированного атласа многообразия $\mathcal{M}$, что $\mathcal{U}_\alpha$ пересекает $\partial\mathcal{M}$; тогда, согласно определению $\partial\mathcal{M}$, $\phi_\alpha(\mathcal{U}_\alpha \cap \partial\mathcal{M})$ лежит в плоскости $x^1 = 0$ в R^n и $\phi_\alpha(\mathcal{U}_\alpha \cap \mathcal{M})$ лежит в нижней половине $x^1 \leqslant 0$. Координаты $(x^2, x^3, \ldots, x^n)$ являются тогда ориентированными координатами в окрестности $\mathcal{U}_\alpha \cap \partial\mathcal{M}$ границы $\partial\mathcal{M}$. Можно убедиться, что таким путем получается ориентированный атлас на $\partial\mathcal{M}$.

Структурой многообразия естественным образом определяется другой вид дифференцирования — *производная Ли*. Рассмотрим произвольное векторное C^r-поле ($r \geqslant 1$) $\mathbf{X}$ на $\mathcal{M}$.

Согласно фундаментальной теореме для систем обыкновенных дифференциальных уравнений [18], через каждую точку $p \in \mathcal{M}$ проходит единственная максимальная кривая $\lambda(t)$, такая, что $\lambda(0) = p$ и ее касательный вектор в точке $\lambda(t)$ есть вектор $\mathbf{X}|_{\lambda(t)}$. Пусть $\{x^i\}$ — локальные координаты, так что кривая $\lambda(t)$ имеет координаты $x^i(t)$, а вектор $\mathbf{X}$ — компоненты X^i. Тогда локально эта кривая есть решение системы дифференциальных уравнений

$$\mathrm{d}x^i/\mathrm{d}t = X^i(x^1(t), \ldots, x^n(t)).$$

Эта кривая называется *интегральной кривой* векторного поля $\mathbf{X}$ с начальной точкой p. Для каждой точки $q \in \mathcal{M}$ найдется ее открытая окрестность $\mathcal{U}$ и $\varepsilon > 0$, такие, что $\mathbf{X}$ задает семейство диффеоморфизмов $\phi_t: \mathcal{U} \to \mathcal{M}$, если $|t| < \varepsilon$, которые получаются переносом каждой точки $p \in \mathcal{U}$ на параметрическое расстояние t вдоль интегральных кривых $\mathbf{X}$. В самом деле, отображения $\phi(t)$ образуют однопараметрическую локальную группу диффеоморфизмов, поскольку $\phi_{t+s} = \phi_t \circ \phi_s = \phi_s \circ \phi_t$ для $|t|, |s|, |t+s| < \varepsilon$, так что $\phi_{-t} = (\phi_t)^{-1}$, а ϕ_0 есть единица группы. Этот диффеоморфизм отображает в $\phi_{t*}\mathbf{T}|_{\phi_t(p)}$ в точке p любое тензорное поле $\mathbf{T}$ типа (r, s).

По определению *производная Ли* $L_{\mathbf{X}}\mathbf{T}$ тензорного поля $\mathbf{T}$ относительно $\mathbf{X}$ есть взятое с обратным знаком значение в точке $t = 0$ производной семейства тензоров $\phi_{t*}\mathbf{T}|_p$ по t, т. е.

$$L_{\mathbf{X}}\mathbf{T}|_p = \lim_{t\to 0} \frac{1}{t}\{\mathbf{T}|_p - \phi_{t*}\mathbf{T}|_p\}.$$

Из свойств ϕ_* следует

1) $L_{\mathbf{X}}$ сохраняет тип тензора, т. е. если $\mathbf{T}$ — тензорное поле типа (r, s), то $L_{\mathbf{X}}\mathbf{T}$ — тензорное поле того же типа;

2) $L_{\mathbf{X}}$ — линейное отображение и сохраняет свертку.

Можно доказать, что как и в обычном анализе справедливо правило Лейбница;

3) для произвольных тензоров $\mathbf{S}, \mathbf{T}$ $L_{\mathbf{X}}(\mathbf{S}\otimes\mathbf{T}) = L_{\mathbf{X}}\mathbf{S}\otimes\mathbf{T} + \mathbf{S}\otimes L_{\mathbf{X}}\mathbf{T}$.

Из предыдущих определений непосредственно следует

4) $L_{\mathbf{X}} f = Xf$, где f — любая функция. При отображении ϕ_t точка $q = \phi_{-t}(p)$ отображается в p. Поэтому ϕ_{t*} есть отображение из T_q в T_p. Следовательно, в силу (2.6)

$$(\phi_{t*}Y)\, f|_p = Y(\phi_t^* f)|_q.$$

В локальных координатах $\{x^i\}$, введенных в окрестности точки p, координаты компоненты $\phi_{t*}Y$ в p равны

$$(\phi_{t*}Y)^i|_p = \phi_{t*}Y|_p\, x^i = Y^j|_q \frac{\partial}{\partial x^j(q)}(x^i(p)) = \frac{\partial x^i(\phi_t(q))}{\partial x^j(q)} Y^j|_q.$$

Поскольку

$$\frac{\mathrm{d}x^i(\phi_t(q))}{\mathrm{d}t} = \mathbf{X}^i|_{\phi_t(q)},$$

то

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial x^i(\phi_t(q))}{\partial x^j(q)}\right)\Big|_{t=0} = \frac{\partial X^i}{\partial x^j}\Big|_p,$$

и, следовательно,

$$(L_{\mathbf{X}}Y)^i = -\frac{\mathrm{d}}{\mathrm{d}t}(\phi_{t*}Y)^i|_{t=0} = \frac{\partial Y_i}{\partial x^j}X^j - \frac{\partial X^i}{\partial x^j}Y^j. \qquad (2.11)$$

Для любых C^2-функций f последнее равенство можно переписать в виде

$$(L_{\mathbf{X}}Y)f = X(Yf) - Y(Xf).$$

Иногда мы будем обозначать $L_{\mathbf{X}}\mathbf{Y}$ через $[\mathbf{X}, \mathbf{Y}]$, т. е.

$$L_{\mathbf{X}}\mathbf{Y} = -L_{\mathbf{Y}}\mathbf{X} = [\mathbf{X}, \mathbf{Y}] = [\mathbf{Y}, \mathbf{X}].$$

Когда производная Ли двух векторных полей $\mathbf{X}$, $\mathbf{Y}$ равна нулю, говорят, что эти векторные поля коммутируют. В этом случае, если из точки p двигаться вдоль интегральной кривой поля $\mathbf{X}$ на расстояние t (по параметру кривой), а потом вдоль интегральной кривой $\mathbf{Y}$ на расстояние s, то конечная точка будет той же, как при движении сначала на расстояние s вдоль интегральной кривой $\mathbf{Y}$, а затем на расстояние t вдоль интегральной кривой $\mathbf{X}$ (см. рис. 7). Таким образом, множество всех точек, которых можно достичь, двигаясь из данной точки p вдоль интегральных кривых $\mathbf{X}$ и $\mathbf{Y}$, образует погруженное двумерное подмногообразие, содержащее p.

Компоненты производной Ли 1-формы ω можно получить сверткой равенства

$$L_{\mathbf{X}}(\omega \otimes \mathbf{Y}) = L_{\mathbf{X}}\omega \otimes \mathbf{Y} + \omega \otimes L_{\mathbf{X}}\mathbf{Y}$$

[свойство (3) производной Ли]; в результате имеем

$$L_{\mathbf{X}}\langle \omega, \mathbf{Y}\rangle = \langle L_{\mathbf{X}}\omega, \mathbf{Y}\rangle + \langle \omega, L_{\mathbf{X}}\mathbf{Y}\rangle$$

[по свойству (2) производной Ли], где $\mathbf{X}$, $\mathbf{Y}$ — произвольные векторные C^1-поля. Выбирая в качестве $\mathbf{Y}$ базисный вектор $\mathbf{E}_i = \partial/\partial x^i$, получим, что координатные компоненты равны

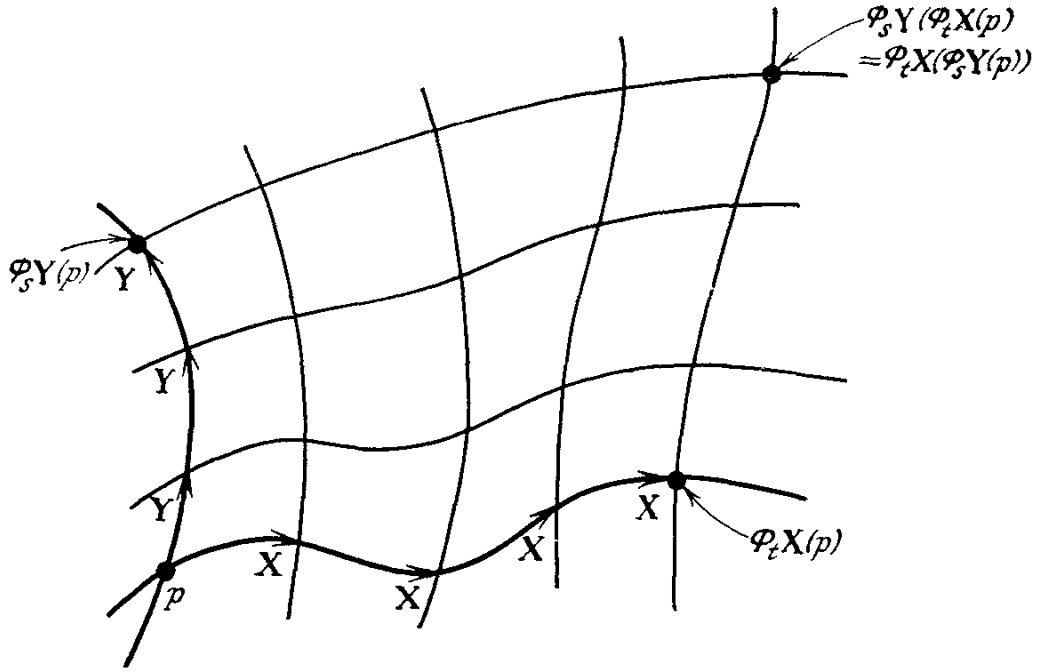


Рис. 7. Преобразования, порождаемые коммутирующими векторными полями X, Y, переводят точку p соответственно в точки $\phi_{t\mathbf{X}}(p)$, $\phi_{s\mathbf{Y}}(p)$.

$$(L_{\mathbf{X}}\omega)_i = \left(\frac{\partial \omega_i}{\partial x_j}\right) X^j + \omega_j \left(\frac{\partial X^j}{\partial x^i}\right),$$

поскольку из (2.11) следует, что

$$\left(L_{\mathbf{X}}\left(\frac{\partial}{\partial x^i}\right)\right)^j = -\frac{\partial X^j}{\partial x^i}.$$

Аналогично, применяя правило Лейбница к выражению

$$L_{\mathbf{X}}(\mathbf{T} \otimes \mathbf{E}^a \otimes \ldots \otimes \mathbf{E}^d \otimes \mathbf{E}_e \otimes \ldots \otimes \mathbf{E}_g)$$

и затем свертывая по всем индексам, можно найти координатные компоненты производной Ли любого тензорного C^r-поля

$\mathbf{T}$ типа (r, s); они равны

$$(L_{\mathbf{X}}T)^{ab\ldots d}{}_{ef\ldots g} = \left(\frac{\partial T^{ab\ldots d}{}_{ef\ldots g}}{\partial x^i}\right) X^i - \\ - T^{ib\ldots d}{}_{ef\ldots g}\frac{\partial X^a}{\partial x^i} - \ldots - T^{ab\ldots i}{}_{ef\ldots g}\frac{\partial X^d}{\partial x^i} + \\ + T^{ab\ldots d}{}_{if\ldots g}\frac{\partial X^i}{\partial x^e} + \ldots + T^{ab\ldots d}{}_{ef\ldots i}\frac{\partial X^i}{\partial x^g}. \qquad (2.12)$$

В силу (2.7) любая производная Ли коммутирует с d, т. е. для любого поля p-формы $\boldsymbol{\omega}$

$$\mathrm{d}(L_{\mathbf{X}}\boldsymbol{\omega}) = L_{\mathbf{X}}(\mathrm{d}\boldsymbol{\omega}).$$

Как из этих формул, так и из их геометрической интерпретации вытекает, что производная Ли $L_{\mathbf{X}}\mathbf{T}_p$ тензорного поля $\mathbf{T}$ типа (r, s) зависит от направления векторного поля $\mathbf{X}$ не только в точке p, но и в соседних точках.

Таким образом, введенные нами два дифференциальных оператора, определяемые структурой многообразия, имеют слишком ограниченный характер для того, чтобы служить обобщением понятия частной производной, которое необходимо для задания уравнения полей физических величин на многообразиях. В самом деле, d действует только на внешние формы, а производная Ли отличается от обычной частной производной тем, что последняя (как производная по направлению) зависит только от направления в рассматриваемой точке. Нужное нам обобщение производной — ковариантную производную — можно получить при введении на многообразии дополнительной структуры. Это будет сделано в следующем параграфе.

2.5. Ковариантная производная и тензор кривизны

Дополнительная структура, которую мы введем, — это (аффинная) *связность* на $\mathscr{M}$. *Связность* ∇ в точке $p \in \mathscr{M}$ есть правило, по которому любому векторному полю $\mathbf{X}$ ставится в точке p в соответствие дифференциальный оператор $\nabla_{\mathbf{X}}$, который отображает произвольное векторное C^r-поле $(r \geqslant 1)$ в векторное поле $\nabla_{\mathbf{X}}\mathbf{Y}$ и обладает следующими свойствами:

1) $\nabla_{\mathbf{X}}\mathbf{Y}$ есть тензор по аргументу $\mathbf{X}$, т. е. для произвольных функций f, g и векторных C^1-полей $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$

$$\nabla_{f\mathbf{X}+g\mathbf{Y}}\mathbf{Z} = f\nabla_{\mathbf{X}}\mathbf{Z} + g\nabla_{\mathbf{Y}}\mathbf{Z}$$

(это эквивалентно условию зависимости $\nabla_{\mathbf{X}}$ только от направления $\mathbf{X}$ в p);

2) $\nabla_{\mathbf{X}}\mathbf{Y}$ линейно по $\mathbf{Y}$, т. е.

$$\nabla_{\mathbf{X}}(\alpha\mathbf{Y} + \beta\mathbf{Z}) = \alpha\nabla_{\mathbf{X}}\mathbf{Y} + \beta\nabla_{\mathbf{X}}\mathbf{Z}$$

для любых векторных C^1-полей $\mathbf{Y}$, $\mathbf{Z}$ и $\alpha, \beta \in R^1$.

3) для любой C^1-функции f и векторного C^1-поля $\mathbf{Y}$

$$\nabla_{\mathbf{X}}(f\mathbf{Y}) = Xf\mathbf{Y} + f\nabla_{\mathbf{X}}\mathbf{Y}.$$

При выполнении всех этих условий $\nabla_{\mathbf{X}} Y$ есть *ковариантная производная* (относительно связности ∇) *векторного поля* $\mathbf{Y}$ *в направлении* $\mathbf{X}$ в точке p. Согласно условию (1), *ковариантную производную векторного поля* $\mathbf{Y}$ можно определить как такое тензорное поле типа (1, 1), которое при свертке с $\mathbf{X}$ дает вектор $\nabla_{\mathbf{X}}\mathbf{Y}$. Тогда имеем

$$3) \Leftrightarrow \nabla(f\mathbf{Y}) = \mathrm{d}f \otimes \mathbf{Y} + f\nabla\mathbf{Y}.$$

Связность ∇ *класса* C^r (C^r-связность) на C^k-многообразии $\mathscr{M}$ ($k \geqslant r+2$) есть правило, по которому в каждой точке задается связность ∇, такая, что, если $\mathbf{Y}$ — векторное C^{r+1}-поле на $\mathscr{M}$, то $\nabla\mathbf{Y}$ — тензорное C^r-поле.

Пусть в некоторой окрестности $\mathscr{U}$ даны произвольный базис C^{r+1}-векторов $\{\mathbf{E}_a\}$ и дуальный базис 1-форм $\{\mathbf{E}^a\}$; будем записывать компоненты $\nabla\mathbf{Y}$ как $Y^a{}_{;b}$, тогда

$$\nabla\mathbf{Y} = Y^a{}_{;b}\mathbf{E}^b \otimes \mathbf{E}_a.$$

На $\mathscr{U}$ связность задают n^3 C^r-функций $\Gamma^a{}_{bc}$, определяемые равенствами

$$\Gamma^a{}_{bc} = \langle \mathbf{E}^a, \quad \nabla_{\mathbf{E}_b}\mathbf{E}_c\rangle \Leftrightarrow \nabla\mathbf{E}_c = \Gamma^a{}_{bc}\mathbf{E}^b \otimes \mathbf{E}_a.$$

Для любого векторного C^1-поля $\mathbf{Y}$

$$\nabla\mathbf{Y} = \nabla(Y^c\mathbf{E}_c) = dY^c \otimes \mathbf{E}_c + Y^c\Gamma^a{}_{bc}\mathbf{E}^b \otimes \mathbf{E}_a.$$

Таким образом, компонентами $\nabla\mathbf{Y}$ относительно координатных базисов $\{\partial/\partial x^a\}$ и $\{\mathrm{d}x^b\}$ являются

$$Y^a{}_{;b} = \frac{\partial Y^a}{\partial x^b} + \Gamma^a{}_{bc}Y^c.$$

Трансформационные свойства функций $\Gamma^a{}_{bc}$ задаются свойствами связности (1) — (3):

$$\Gamma^{a'}{}_{b'c'} = \langle \mathbf{E}^{a'}, \quad \nabla_{\mathbf{E}_{b'}}\mathbf{E}_{c'}\rangle = \langle \Phi^{a'}{}_a\mathbf{E}^a, \quad \nabla_{\Phi_{b'}{}^b\mathbf{E}_b}(\Phi_{c'}{}^c\mathbf{E}_c)\rangle =$$
$$= \Phi^{a'}{}_a\Phi_{b'}{}^b\left(E_b(\Phi_{c'}{}^a) + \Phi_{c'}{}^c\Gamma^a{}_{bc}\right),$$

поскольку $\mathbf{E}_{a'} = \Phi_{a'}{}^a\mathbf{E}_a$, $\mathbf{E}^{a'} = \Phi^{a'}{}_a\mathbf{E}^a$. Это равенство можно переписать в виде

$$\Gamma^{a'}{}_{b'c'} = \Phi^{a'}{}_a\left(E_{b'}(\Phi_{c'}{}^a) + \Phi_{b'}{}^b\Phi_{c'}{}^c\Gamma^a{}_{bc}\right).$$

В частности, если базисами являются координатные базисы, определяемые координатами $\{x^a\}$, $\{x^{a'}\}$, закон преобразования

имеет вид

$$\Gamma^{a'}{}_{b'c'} = \frac{\partial x^{a'}}{\partial x^a}\left(\frac{\partial^2 x^a}{\partial x^{b'}\,\partial x^{c'}} + \frac{\partial x^b}{\partial x^{b'}}\frac{\partial x^c}{\partial x^{c'}}\,\Gamma^a{}_{bc}\right).$$

Из-за члена $E_{b'}(\Phi_{c'}{}^a)$ функции $\Gamma^a{}_{bc}$ не преобразуются как компоненты тензора. Однако если ковариантные производные $\nabla\mathbf{Y}$ и $\hat{\nabla}\mathbf{Y}$ получены с использованием двух различных связностей, то

$$\hat{\nabla}\mathbf{Y} - \nabla\mathbf{Y} = (\Gamma^a{}_{bc} - \hat{\Gamma}^a{}_{bc})\,Y^c\mathbf{E}^b \otimes \mathbf{E}_a$$

будет тензором. Таким образом, разности $(\Gamma^a{}_{bc} - \hat{\Gamma}^a{}_{bc})$ являются компонентами тензора.

Более широкое понятие ковариантной производной произвольного тензора C^r-поля ($r \geqslant 1$) можно ввести, задав следующие правила (ср. со свойствами производной Ли):

1) если $\mathbf{T}$ — тензорное C^r-поле типа (q, s), то $\nabla\mathbf{T}$ — тензорное C^{r-1}-поле типа $(q, s+1)$;

2) ∇ — линейный оператор, коммутирующий со сверткой;

3) для произвольных тензорных полей $\mathbf{S}$, $\mathbf{T}$ выполняется правило Лейбница

$$\nabla(\mathbf{S}\otimes\mathbf{T}) = \nabla\mathbf{S}\otimes\mathbf{T} + \mathbf{S}\otimes\nabla\mathbf{T};$$

4) $\nabla f = \mathrm{d}f$ для любой функции f.

Запишем компоненты $\nabla\mathbf{T}$ в виде $(\nabla_{\mathbf{E}_h}T)^{a\ldots d}{}_{e\ldots g} = T^{a\ldots d}{}_{e\ldots g;h}$. Согласно правилам (2) и (3), имеем

$$\nabla_{\mathbf{E}_b}\mathbf{E}^c = -\Gamma^c{}_{ba}\mathbf{E}^a,$$

где $\{\mathbf{E}^a\}$ — базис, дуальный $\{\mathbf{E}_a\}$. Методом, аналогичным выводу равенства (2.12), можно показать, что координатные компоненты $\nabla\mathbf{T}$ равны

$$T^{ab\ldots d}{}_{ef\ldots g;h} = \frac{\partial T^{ab\ldots d}{}_{ef\ldots g}}{\partial x^h} + \Gamma^a{}_{hj}T^{jb\ldots d}{}_{ef\ldots g} + \\ + \Gamma^b{}_{hj}T^{aj\ldots d}{}_{ef\ldots g} + \ldots + \Gamma^d{}_{hj}T^{ab\ldots j}{}_{ef\ldots g} - \\ - \Gamma^j{}_{he}T^{ab\ldots d}{}_{jf\ldots g} - \Gamma^j{}_{hf}T^{ab\ldots d}{}_{ej\ldots g} - \ldots - \Gamma^j{}_{hg}T^{ab\ldots d}{}_{ef\ldots j}. \qquad (2.13)$$

Интересным примером является единичный тензор $\mathbf{E}_b\otimes\mathbf{E}^a$, имеющий компонентами $\delta^a{}_b$; его ковариантная производная равна нулю, и, следовательно, нулю равны также и ковариантные производные обобщенных единичных тензоров с компонентами $\delta^{(a_1}{}_{b_1}\delta^{a_2}{}_{b_2}\ldots\delta^{a_s)}{}_{b_s}$, $\delta^{[a_1}{}_{b_1}\delta^{a_2}{}_{b_2}\ldots\delta^{a_p]}{}_{b_p}$ $(p \leqslant n)$.

Пусть $\mathbf{T}$ — тензорное C^r-поле ($r \geqslant 1$), заданное вдоль C^r-кривой $\lambda(t)$; тогда можно ввести $\mathrm{D}\mathbf{T}/\partial t$, *ковариантную производную* $\mathbf{T}$ *вдоль* $\lambda(t)$, как $\nabla_{\partial/\partial t}\bar{\mathbf{T}}$, где $\bar{\mathbf{T}}$ — любое тензорное C^r-поле,

расширяющее $\mathbf{T}$ на некоторую открытую окрестность кривой $\lambda(t)$. $\mathrm{D}\mathbf{T}/\partial t$ является тензорным C^{r-1}-полем, которое определено вдоль кривой $\lambda(t)$ и не зависит от расширения $\bar{\mathbf{T}}$. Пусть $\mathbf{X}$ — касательный вектор к $\lambda(t)$, тогда в записи через компоненты имеем $\mathrm{D}T^{a\ldots d}{}_{e\ldots g}/\partial t = T^{a\ldots d}{}_{e,\ldots g;\,h}X^h$. В частности, можно выбрать такие локальные координаты, в которых координаты кривой $\lambda(t)$ равны $x^a(t)$, $X^a = \mathrm{d}x^a/\mathrm{d}t$, и тогда для векторного поля

$$\mathrm{D}Y^a/\partial t = \partial Y^a/\partial t + \Gamma^a{}_{bc}Y^c\,\mathrm{d}x^b/\mathrm{d}t. \tag{2.14}$$

Говорят, что тензор $\mathbf{T}$ *параллельно переносится* вдоль кривой λ, если $\mathrm{D}\mathbf{T}/\partial t = 0$. Пусть дана кривая $\lambda(t)$ с концами в точках p и q, тогда из теории обыкновенных дифференциальных уравнений при условии, что связность ∇ — по меньшей мере класса C^{1-}, следует единственность тензора, получаемого в точке q параллельным переносом вдоль кривой $\lambda(t)$ тензора, заданного в точке p. Таким образом, параллельный перенос вдоль λ есть линейное отображение из $T^r_s(p)$ в $T^r_s(q)$, сохраняющее все тензорные произведения и тензорные свертки. В частности, из этого следует, что параллельный перенос базиса векторов вдоль данной кривой из p в q задает изоморфизм из T_p в T_q. (Если у кривой есть самопересечения, то p и q могут быть *одной и той же* точкой.)

Важное понятие возникает при рассмотрении ковариантной производной вдоль λ самого касательного вектора. Кривая $\lambda(t)$ называется *геодезической кривой*, если вектор

$$\nabla_{\mathbf{X}}\mathbf{X} = \frac{\mathrm{D}}{\partial t}\left(\frac{\partial}{\partial t}\right)_\lambda$$

параллелен вектору $(\partial/\partial t)_\lambda$, т. е. если найдется такая функция f (возможно, равная нулю), что $X^a{}_{;\,b}X^b = fX^a$. В этом случае можно ввести новый параметр $v(t)$ вдоль кривой, так что будет выполняться равенство

$$\frac{\mathrm{D}}{\partial v}\left(\frac{\partial}{\partial v}\right)_\lambda = 0.$$

Такой параметр называется *аффинным параметром*. Связанный с ним касательный вектор $\mathbf{V} = (\partial/\partial v)_\lambda$ параллелен $\mathbf{X}$, но масштаб определяется условием $V(v) = 1$; $\mathbf{V}$ подчиняется уравнениям

$$V^a{}_{;\,b}V^b = 0 \Leftrightarrow \frac{\mathrm{d}^2x^a}{\mathrm{d}v^2} + \Gamma^a{}_{bc}\frac{\mathrm{d}x^b}{\mathrm{d}v}\frac{\mathrm{d}x^c}{\mathrm{d}v} = 0, \tag{2.15}$$

причем второе (локально-координатное) выражение получается применением (2.14) к вектору $\mathbf{V}$. Аффинный параметр геодезической кривой определен с точностью до аддитивной и мультипликативной постоянных, т. е. с точностью до преобразования

$v' = av + b$, где a, b — постоянные; свобода в выборе b соответствует свободе выбора новой начальной точки $\lambda(0)$, свобода в выборе a — возможности перенормировки масштаба вектора $\mathbf{V}$ умножением на постоянную: $\mathbf{V}' = (1/a)\mathbf{V}$. Будем называть геодезическую кривую просто *геодезической*, если эта кривая параметризуется каким-либо аффинным параметром.

Пусть задана C^r-связность ($r \geqslant 0$); тогда из известных теорем для обыкновенных дифференциальных уравнений следует, что для любой точки $p \in \mathscr{M}$ и для любого вектора $\mathbf{X}_p$ в p существует максимальная геодезическая $\lambda_{\mathbf{X}}(v)$ в $\mathscr{M}$ с начальной точкой p и начальным направлением $\mathbf{X}_p$, т. е. $\lambda_{\mathbf{X}}(0) = p$ и $(\partial/\partial v)_\lambda|_{v=0} = \mathbf{X}_p$. Если $r \geqslant 1$, то эта геодезическая единственна и ее зависимость от p и $\mathbf{X}_p$ непрерывна. Это означает, что при $r \geqslant 1$ можно определить такое C^r-отображение $\exp: T_p \to \mathscr{M}$, что для каждого $\mathbf{X} \in T_p$, $\exp(\mathbf{X})$ есть точка в $\mathscr{M}$, находящаяся на единичном расстоянии по параметру вдоль геодезической $\lambda_{\mathbf{X}}$, начинающейся в p. Это отображение не обязательно определено для всех $\mathbf{X} \in T_p$, поскольку геодезическая $\lambda_{\mathbf{X}}(v)$ может не быть определена для всех v. Если же v может принимать все значения, то говорят, что геодезическая $\lambda_{\mathbf{X}}(v)$ — *полная*. Многообразие $\mathscr{M}$ называется *геодезически полным*, если все геодезические на $\mathscr{M}$ полны, т. е. если отображение exp определено на всем пространстве T_p каждой точки $p \in \mathscr{M}$.

Независимо от того, полно или нет многообразие $\mathscr{M}$, отображение $\exp_p$ имеет ранг n в p. Поэтому в силу теоремы о неявной функции [161] существует открытая окрестность $\mathscr{N}_0$ начала в T_p и открытая окрестность $\mathscr{N}_p$ точки $p \in \mathscr{M}$, такие, что exp есть C^r-диффеоморфизм $\mathscr{N}_0$ на $\mathscr{N}_p$. Такая окрестность $\mathscr{N}_p$ называется *нормальной окрестностью* точки p. Далее, $\mathscr{N}_p$ можно считать выпуклой, т. е. такой, что всякая точка $q \in \mathscr{N}_p$ может быть соединена с любой другой точкой $r \in \mathscr{N}_p$ геодезической, начинающейся в q и целиком лежащей в $\mathscr{N}_p$. Внутри выпуклой нормальной окрестности $\mathscr{N}$ координаты $(x^1, \ldots, x^n)$ можно ввести следующим образом: выбираем какую-либо точку $q \in \mathscr{N}$, базис $\{\mathbf{E}_a\}$ в T_q и координаты точки $r \in \mathscr{N}$ задаем соотношением $r = \exp(x^a \mathbf{E}_a)$ (т. е. точке r приписываем координаты в базисе $\{\mathbf{E}_a\}$ точки $\exp^{-1}(r) \in T_q$). Тогда $(\partial/\partial x^i)|_q = \mathbf{E}_i$ и, согласно (2.15), $\Gamma^i_{(jk)}|_q = 0$. Такие координаты называются *нормальными координатами* с началом в q. Герок [55] использовал существование нормальных окрестностей для доказательства наличия счетного базиса у связного хаусдорфова C^3-многообразия с C^1-связностью. Таким образом паракомпактность C^3-многообразия можно связать с существованием на нем C^1-связности. «Нормальное» локальное поведение геодезических в этих окрестностях контрастирует с поведением геодезических в большом в общем случае,

когда, с одной стороны, две произвольные точки, вообще говоря, не могут быть соединены какой-либо геодезической, а, с другой, некоторые геодезические, исходящие из одной точки, могут собраться как фокусе в некоторой другой точке. Позднее мы столкнемся с примерами и того и другого поведения.

Если задана C^r-связность ∇, то равенство

$$\mathbf{T}(\mathbf{X}, \mathbf{Y}) = \nabla_{\mathbf{X}}\mathbf{Y} - \nabla_{\mathbf{Y}}\mathbf{X} - [\mathbf{X}, \mathbf{Y}]$$

определяет некоторое тензорное C^{r-1}-поле типа (1, 2), где $\mathbf{X}$, $\mathbf{Y}$ — произвольные векторные C^r-поля. Такой тензор называется *тензором кручения*. В координатном базисе его компоненты равны

$$T^i{}_{jk} = \Gamma^i{}_{jk} - \Gamma^i{}_{kj}.$$

Мы будем иметь дело лишь со связностями *без кручения*; иначе говоря, мы предполагаем, что $\mathbf{T} = 0$. В этом случае координатные компоненты связности удовлетворяют условию $\Gamma^i{}_{jk} = \Gamma^i{}_{kj}$, и поэтому такую связность часто называют симметричной. Связность тогда и только тогда является связностью без кручения, когда $f_{;ij} = f_{;ji}$ для всех функций f. Из уравнения геодезической (2.15) следует, что связность без кручения полностью определяется заданием геодезических на $\mathscr{M}$.

В отсутствие кручения ковариантные производные произвольных векторных C^1-полей связаны с их производными Ли соотношением

$$[\mathbf{X}, \mathbf{Y}] = \nabla_{\mathbf{X}}\mathbf{Y} - \nabla_{\mathbf{Y}}\mathbf{X} \Leftrightarrow (L_{\mathbf{X}}\mathbf{Y})^a = Y^a{}_{;b}X^b - X^a{}_{;b}Y^b, \qquad (2.16)$$

а для любого тензорного C^1-поля T типа (r, s)

$$\begin{aligned}(L_{\mathbf{X}}T)^{ab\ldots d}{}_{ef\ldots g} = {} & T^{ab\ldots d}{}_{ef\ldots g;h}X^h - T^{jb\ldots d}{}_{ef\ldots g}X^a{}_{;j} - \\ & - T^{aj\ldots d}{}_{ef\ldots g}X^b{}_{;j} - \ldots - T^{ab\ldots j}{}_{ef\ldots g}X^d{}_{;j} + \\ & + T^{ab\ldots d}{}_{jf\ldots g}X^j{}_{;e} + T^{ab\ldots d}{}_{ej\ldots g}X^j{}_{;f} + \ldots + T^{ab\ldots d}{}_{ef\ldots j}X^j{}_{;g}.\end{aligned} \qquad (2.17)$$

Легко проверить также, что внешняя производная связана с ковариантной следующим образом:

$$\mathbf{dA} = A_{a\ldots c;d}\,dx^d \wedge dx^a \wedge \ldots \wedge dx^c \Leftrightarrow (dA)_{a\ldots cd} = (-1)^p A_{[a\ldots c;d]},$$

где $\mathbf{A}$ — произвольная p-форма. Таким образом, уравнения, содержащие внешнюю производную или производную Ли, всегда можно выразить через ковариантную производную. Однако в силу своих определений производная Ли и внешняя производная не зависят от связности.

Если вектор $\mathbf{X}_p$ из данной точки p перенести параллельно вдоль некоторой замкнутой кривой γ снова в ту же точку p, то получится вектор $\mathbf{X}'_p$, вообще говоря, отличный от $\mathbf{X}_p$; если выбрать другую кривую γ', то полученный при этом новый вектор в p будет, вообще говоря, отличен от $\mathbf{X}_p$ и от $\mathbf{X}'_p$. Эта неинтегрируемость параллельного переноса соответствует тому, что ковариантные производные в общем случае не коммутируют. Меру этой некоммутативности дает тензор (кривизны) Римана. Пусть даны векторные C^{r+1}-поля $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, а векторное C^{r-1}-поле $\mathbf{R}(\mathbf{X}, \mathbf{Y})\mathbf{Z}$ определено C^r-связностью ∇ следующим образом:

$$\mathbf{R}(\mathbf{X}, \mathbf{Y})\mathbf{Z} = \nabla_{\mathbf{X}}(\nabla_{\mathbf{Y}}\mathbf{Z}) - \nabla_{\mathbf{Y}}(\nabla_{\mathbf{X}}\mathbf{Z}) - \nabla_{[\mathbf{X}, \mathbf{Y}]}\mathbf{Z}. \tag{2.18}$$

Тогда поле $\mathbf{R}(\mathbf{X}, \mathbf{Y})\mathbf{Z}$ линейно по $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, и можно убедиться, что значение $\mathbf{R}(\mathbf{X}, \mathbf{Y})\mathbf{Z}$ в p зависит только от значений $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ в p, т. е. $\mathbf{R}(\mathbf{X}, \mathbf{Y})\mathbf{Z}$ есть тензорное C^{r-1}-поле типа (3, 1). Чтобы записать (2.18) в компонентах, определим вторую ковариантную производную $\nabla\nabla\mathbf{Z}$ вектора $\mathbf{Z}$ как ковариантную производную $\nabla(\nabla\mathbf{Z})$ тензора $\nabla\mathbf{Z}$; ее компоненты равны

$$Z^a{}_{;bc} = (Z^a{}_{;b})_{;c}.$$

Тогда (2.18) можно записать в виде

$$R^a{}_{bcd}X^cY^dZ^b = (Z^a{}_{;d}Y^d)_{;c}\,X^c - (Z^a{}_{;d}X^d)_{;c}\,Y^c - \\ - Z^a{}_{;d}\,(Y^d{}_{;c}X^c - X^d{}_{;c}Y^c) = (Z^a{}_{;dc} - Z^a{}_{;cd})\,X^cY^d,$$

где компоненты тензора Римана $R^a{}_{bcd}$ относительно дуальных базисов $\{\mathbf{E}_a\}$ и $\{\mathbf{E}^a\}$ определяются как $R^a{}_{bcd} = \langle \mathbf{E}^a, \mathbf{R}(\mathbf{E}_c, \mathbf{E}_d)\mathbf{E}_b\rangle$. Поскольку $\mathbf{X}$, $\mathbf{Y}$ — произвольные векторы, равенство

$$Z^a{}_{;dc} - Z^a{}_{;cd} = R^a{}_{bcd}Z^b \tag{2.19}$$

выражает некоммутативность вторых ковариантных производных $\mathbf{Z}$ через тензор Римана.

Поскольку для любых векторных C^2-полей $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ и любого C^2-поля 1-формы η выполняется равенство

$$\nabla_{\mathbf{X}}(\eta \otimes \nabla_{\mathbf{Y}}\mathbf{Z}) = \nabla_{\mathbf{X}}\eta \otimes \nabla_{\mathbf{Y}}\mathbf{Z} + \eta \otimes \nabla_{\mathbf{X}}\nabla_{\mathbf{Y}}\mathbf{Z} \Rightarrow \langle \eta, \nabla_{\mathbf{X}}\nabla_{\mathbf{Y}}\mathbf{Z}\rangle = \\ = X(\langle \eta, \nabla_{\mathbf{Y}}\mathbf{Z}\rangle) - \langle \nabla_{\mathbf{X}}\eta, \nabla_{\mathbf{Y}}\mathbf{Z}\rangle,$$

то из (2.18) следует, что

$$\langle \mathbf{E}^a, \mathbf{R}(\mathbf{E}_c, \mathbf{E}_d)\mathbf{E}_b\rangle = E_c\left(\langle \mathbf{E}^a, \nabla_{\mathbf{E}_d}\mathbf{E}_b\rangle\right) - E_d\left(\langle \mathbf{E}^a, \nabla_{\mathbf{E}_c}\mathbf{E}_b\rangle\right) - \\ - \langle \nabla_{\mathbf{E}_c}\mathbf{E}^a, \nabla_{\mathbf{E}_d}\mathbf{E}_b\rangle + \langle \nabla_{\mathbf{E}_d}\mathbf{E}^a, \nabla_{\mathbf{E}_c}\mathbf{E}_b\rangle - \langle \mathbf{E}^a, \nabla_{[\mathbf{E}_c, \mathbf{E}_d]}\mathbf{E}_b\rangle.$$

Выбрав в качестве $\{\mathbf{E}_a\}$, $\{\mathbf{E}^a\}$ координатные базисы, можно выразить координатные компоненты тензора Римана через коор-

динатные компоненты связности:

$$R^a{}_{bcd} = \frac{\partial \Gamma^a{}_{db}}{\partial x^c} - \frac{\partial \Gamma^a{}_{cb}}{\partial x^d} + \Gamma^a{}_{cf}\Gamma^f{}_{db} - \Gamma^a{}_{df}\Gamma^f{}_{cb}. \qquad (2.20)$$

Пользуясь этими определениями, можно убедиться, что кроме симметрии

$$R^a{}_{bcd} = -R^a{}_{bdc} \Leftrightarrow R^a{}_{b(cd)} = 0 \qquad (2.21а)$$

тензор кривизны обладает также симметрией вида

$$R^a{}_{[bcd]} = 0 \Leftrightarrow R^a{}_{bcd} + R^a{}_{dbc} + R^a{}_{cdb} = 0. \qquad (2.21б)$$

Подобным же образом можно убедиться, что первые ковариантные производные тензора Римана удовлетворяют *тождествам Бианки*:

$$R^a{}_{b[cd;e]} = 0 \Leftrightarrow R^a{}_{bcd;e} + R^a{}_{bec;d} + R^a{}_{bde;c} = 0. \qquad (2.22)$$

Теперь мы видим, что параллельный перенос произвольного вектора вдоль произвольной замкнутой кривой локально интегрируем (т. е. $\mathbf{X}'_p$ необходимо совпадает с $\mathbf{X}_p$ в каждой $p \in \mathscr{M}$), только если $R^a{}_{bcd} = 0$ во всех точках $\mathscr{M}$; в этом случае мы говорим, что связность *плоская*. Свертывая тензор кривизны, можно определить *тензор Риччи* как тензор типа (0, 2) с компонентами

$$R_{bd} = R^a{}_{bad}.$$

2.6. Метрика

Метрический тензор $\mathbf{g}$ в точке $p \in \mathscr{M}$ есть симметричный тензор типа (0, 2) в p, так что C^r-метрика на $\mathscr{M}$ есть C^r-поле симметричного тензора $\mathbf{g}$. Метрика $\mathbf{g}$ приписывает каждому вектору $\mathbf{X} \in T_p$ его «длину» $|g(\mathbf{X}, \mathbf{X})|^{1/2}$ и задает «косинус угла»

$$\frac{g(\mathbf{X}, \mathbf{Y})}{(|g(\mathbf{X}, \mathbf{X})\, g(\mathbf{Y}, \mathbf{Y})|)^{1/2}}$$

между любыми двумя векторами $\mathbf{X}$, $\mathbf{Y} \in T_p$, для которых $g(\mathbf{X}, \mathbf{X}) \cdot g(\mathbf{Y}, \mathbf{Y}) \neq 0$. Будем говорить, что векторы $\mathbf{X}$, $\mathbf{Y}$ *ортогональны*, если $g(\mathbf{X}, \mathbf{Y}) = 0$.

Компонентами $\mathbf{g}$ относительно базиса $\{\mathbf{E}_a\}$ являются

$$g_{ab} = g(\mathbf{E}_a, \mathbf{E}_b) = g(\mathbf{E}_b, \mathbf{E}_a),$$

т. е. компоненты $\mathbf{g}$ равны просто скалярным произведениям базисных векторов $\mathbf{E}_a$. Если пользоваться координатным базисом, то

$$\mathbf{g} = g_{ab}\, dx^a \otimes dx^b. \qquad (2.23)$$

Длины касательного пространства, определяемые метрикой, связаны с длинами на многообразии следующим образом. *Длина пути* между точками $p = \gamma(a)$ и $q = \gamma(b)$ вдоль кривой $\gamma(t)$

класса C^0, кусочно класса C^1, с касательным вектором $\partial/\partial t$, таким, что $g(\partial/\partial t, \partial/\partial t)$ имеет один знак во всех точках $\gamma(t)$, равна

$$L = \int_a^b |g(\partial/\partial t, \partial/\partial t)|^{1/2}\, dt. \tag{2.24}$$

Выражения (2.23), (2.24) можно символически представить в форме

$$ds^2 = g_{ij}\, dx^i\, dx^j,$$

используемой в классических руководствах для выражения длины «бесконечно малой» дуги, задаваемой смещением координат $x^i \to x^i + dx^i$.

Метрика называется *невырожденной*, если не существует такого ненулевого вектора $\mathbf{X} \in T_p$, для которого $g(\mathbf{X}, \mathbf{Y}) = 0$ при всех $\mathbf{Y} \in T_p$. В терминах компонент метрика не вырождена, если матрица (g_{ab}) из компонент $\mathbf{g}$ неособенна. Далее мы всегда будем предполагать, что метрический тензор не вырожден. Тогда соотношением

$$g^{ab} g_{bc} = \delta^a{}_c$$

мы можем однозначно определить симметричный тензор типа $(2, 0)$ с компонентами g^{ab} относительно базиса $\{\mathbf{E}_a\}$, дуального к базису $\{\mathbf{E}^a\}$. Иначе говоря, матрица (g_{ab}) тоже неособенна, и тензоры g^{ab}, g_{ab} можно использовать для установления изоморфизма между любым ковариантным тензорным аргументом и любым контравариантным аргументом, иначе говоря, для «поднятия» и «опускания индексов». Таким образом, если X^a — компоненты контравариантного вектора, ему однозначным образом сопоставляется ковариантный вектор с компонентами X_a, причем $X_a = g_{ab} X^b$, $X^a = g^{ab} X_b$. Аналогично, тензору T_{ab} типа $(0, 2)$ мы можем однозначно сопоставить тензоры $T^a{}_b = g^{ac} T_{cb}$, $T_a{}^b = g^{bc} T_{ac}$, $T^{ab} = g^{ac} g^{bd} T_{cd}$. Вообще мы будем рассматривать подобные взаимосвязанные ковариантные и контравариантные тензоры как различные представления одного и того же геометрического объекта; в частности, g_{ab}, $\delta_a{}^b$ и g^{ab} можно считать представлениями (относительно дуальных базисов) одного и того же геометрического объекта $\mathbf{g}$. Однако в некоторых случаях, когда мы будем иметь дело более чем с одной метрикой, нужно будет тщательно различать, какая из метрик используется для поднятия или опускания индексов.

Сигнатура $\mathbf{g}$ в точке p есть число, равное разности числа положительных и числа отрицательных собственных значений матрицы (g_{ab}) в p. Если метрика $\mathbf{g}$ не вырождена и непрерывна, то сигнатура постоянна на $\mathscr{M}$. Подходящим выбором базиса $\{\mathbf{E}_a\}$ компоненты метрического тензора в любой точке p можно при-

вести к виду

$$g_{ab} = \operatorname{diag}(\underbrace{+1, +1, \ldots, +1}_{(n+s)/2 \text{ членов}}, \underbrace{-1, -1, \ldots, -1}_{(n-s)/2 \text{ членов}}),$$

где s — сигнатура $\mathbf{g}$, n — размерность $\mathscr{M}$. В этом случае базисные векторы $\{\mathbf{E}_a\}$ образуют ортонормированную систему в p, т. е. каждый из них есть единичный вектор, ортогональный ко всем остальным базисным векторам.

Метрика, сигнатура которой равна n, называется *положительно определенной*; для такой метрики $g(\mathbf{X}, \mathbf{X}) = 0 \Rightarrow \mathbf{X} = 0$, а ее каноническая форма имеет вид

$$g_{ab} = \operatorname{diag}(\underbrace{+1, \ldots, +1}_{n \text{ членов}}).$$

Положительно определенная метрика является «метрикой» в пространстве в топологическом смысле этого слова.

Метрика с сигнатурой, равной $n - 2$, называется *лоренцевой метрикой*; ее каноническая форма:

$$g_{ab} = \operatorname{diag}(\underbrace{+1, \ldots, +1}_{(n-1) \text{ членов}}, -1).$$

Если $\mathscr{M}$ наделено лоренцевой метрикой, нулевые векторы в $p \in \mathscr{M}$ можно разделить на три класса: вектор $\mathbf{X} \in T_p$ называется *времениподобным*, *изотропным* или *пространственноподобным* в зависимости от того, отрицателен, равен нулю или положителен скалярный квадрат $g(\mathbf{X}, \mathbf{X})$. Изотропные векторы образуют в T_p двухполостный конус, отделяющий времениподобные векторы от пространственноподобных (рис. 8). Если $\mathbf{X}$, $\mathbf{Y}$ — какие-либо два непространственноподобных (т. е. времениподобных или изотропных) вектора, лежащие в одной и той же полости изотропного конуса точки p, то $g(\mathbf{X}, \mathbf{Y}) \leqslant 0$, причем равенство имеет место, только когда $\mathbf{X}$ и $\mathbf{Y}$ — параллельные изотропные векторы [т. е. при $\mathbf{X} = \alpha\mathbf{Y}$, $g(\mathbf{X}, \mathbf{X}) = 0$].

Любое паракомпактное C^r-многообразие допускает положительно определенную C^{r-1}-метрику (т. е. такая метрика определена на всем $\mathscr{M}$). Чтобы убедиться в этом, возьмем разложение единицы $\{f_\alpha\}$ для локально-конечного атласа $\{\mathscr{U}_\alpha, \phi_\alpha\}$. Тогда g можно задать равенством

$$g(\mathbf{X}, \mathbf{Y}) = \sum_\alpha f_\alpha \langle (\phi_\alpha)_* \mathbf{X}, (\phi_\alpha)_* \mathbf{Y} \rangle,$$

где $\langle\ ,\ \rangle$ — естественное скалярное произведение в евклидовом пространстве R^n; таким образом атлас используется для задания метрики путем отображения евклидовой метрики в $\mathscr{M}$. Ясно, что

такое введение метрики неинвариантно относительно изменения атласа и, следовательно, на $\mathscr{M}$ имеется много подобных положительно определенных метрик.

В противоположность этому паракомпактное C^r-многообразие допускает лоренцеву C^{r-1}-метрику тогда и только тогда, когда

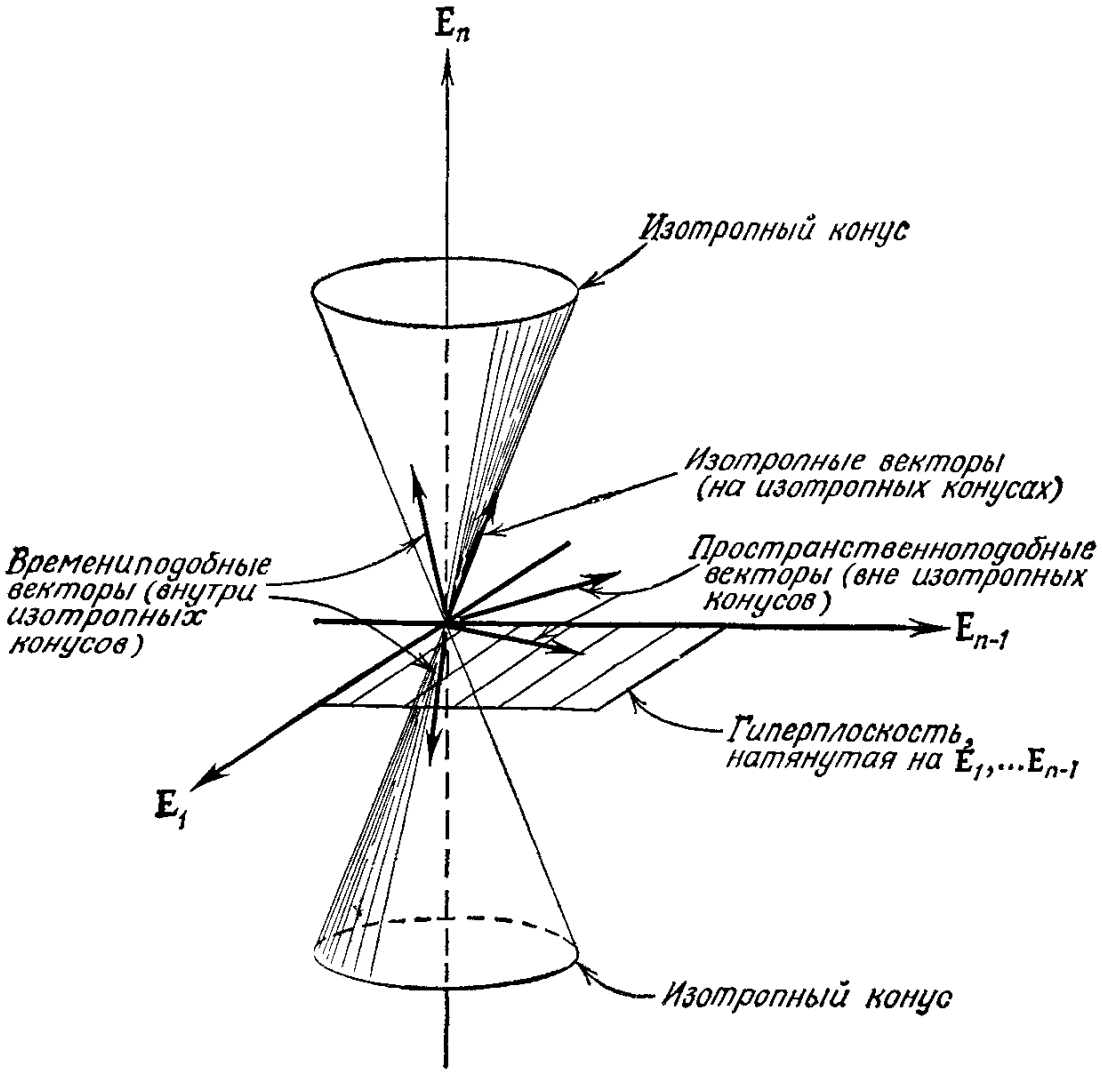


Рис. 8. Изотропные конусы, определяемые лоренцевой метрикой.

оно допускает неисчезающее C^{r-1}-поле линейных элементов; под полем линейных элементов подразумевается сопоставление каждой точке $p \in \mathscr{M}$ пары равных и противоположно направленных векторов $(\mathbf{X}, -\mathbf{X})$, т. е. поле линейных элементов подобно векторному полю, но не имеет определенного знака. Чтобы убедиться в этом, введем на многообразии положительно определенную C^{r-1}-метрику $\hat{\mathbf{g}}$. Тогда в каждой точке p можно следующим образом определить лоренцеву метрику $\mathbf{g}$:

$$g(\mathbf{Y}, \mathbf{Z}) = \hat{g}(\mathbf{Y}, \mathbf{Z}) - 2\frac{\hat{g}(\mathbf{X}, \mathbf{Y})\,\hat{g}(\mathbf{X}, \mathbf{Z})}{\hat{g}(\mathbf{X}, \mathbf{X})},$$

где $\mathbf{X}$ — один из векторов пары $(\mathbf{X}, -\mathbf{X})$ в p. (Отметим, что $\mathbf{X}$ появляется в этом выражении четное число раз, и поэтому несущественно, подставлять $\mathbf{X}$ или $-\mathbf{X}$.) Очевидно, $g(\mathbf{X}, \mathbf{X}) = -\hat{g}(\mathbf{X}, \mathbf{X})$, и если $\mathbf{Y}$, $\mathbf{Z}$ — ортогональны к $\mathbf{X}$ относительно $\hat{\mathbf{g}}$, то они ортогональны к $\mathbf{X}$ также относительно $\mathbf{g}$, причем $g(\mathbf{Y}, \mathbf{Z}) = \hat{g}(\mathbf{Y}, \mathbf{Z})$. Следовательно, ортонормированный относительно $\hat{g}$ базис будет ортонормированным и относительно $\mathbf{g}$. Поскольку метрика $\hat{\mathbf{g}}$ не единственна, то на $\mathscr{M}$ имеется множество лоренцевых метрик, если есть хотя бы одна. Обратно, пусть дана лоренцева метрика $\mathbf{g}$; рассмотрим уравнение $g_{ab}X^b = \lambda \hat{g}_{ab}X^b$, где $\hat{\mathbf{g}}$ — какая-либо положительно определенная метрика. При этом получим одно отрицательное и $(n-1)$ положительных собственных чисел λ. Поле собственного вектора $\mathbf{X}$, соответствующего отрицательному собственному значению, будет локально определено с точностью до знака нормирующего множителя; его можно нормировать, наложив условия $g_{ab}X^aX^b = -1$, и получить таким образом поле линейных элементов на $\mathscr{M}$.

В действительности любое некомпактное многообразие допускает поле линейных элементов, а чтобы такое поле допускало компактное многообразие, необходимо и достаточно, чтобы характеристика Эйлера многообразия равнялась нулю (т. е. тор T^2 допускает, а сфера S^2 не допускает поле линейных элементов). Позднее будет установлено, что многообразие может быть разумной моделью пространства-времени, только если оно некомпактно и в силу этого на нем существует множество лоренцевых метрик.

До сих пор мы рассматривали метрический тензор и связность как две не зависимые друг от друга структуры на $\mathscr{M}$. Однако при заданной метрике $\mathbf{g}$ на $\mathscr{M}$ имеется единственная связность без кручения, для которой ковариантная производная $\mathbf{g}$ равна нулю, т. е.

$$g_{ab;c} = 0. \tag{2.25}$$

При этой связности параллельный перенос векторов сохраняет скалярное произведение, задаваемое метрикой $\mathbf{g}$; поэтому, в частности, длины векторов инвариантны. Например, если $\partial/\partial t$ — вектор, касательный к геодезической, то вдоль этой геодезической $g(\partial/\partial t, \partial/\partial t) = \text{const}$.

Из (2.25) следует, что для производных векторных C^1-полей $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$

$$X(g(\mathbf{Y}, \mathbf{Z})) = \nabla_{\mathbf{X}}(g(\mathbf{Y}, \mathbf{Z})) = \nabla_{\mathbf{X}}g(\mathbf{Y}, \mathbf{Z}) + g(\nabla_{\mathbf{X}}\mathbf{Y}, \mathbf{Z}) + g(\mathbf{Y}, \nabla_{\mathbf{X}}\mathbf{Z}) = g(\nabla_{\mathbf{X}}\mathbf{Y}, \mathbf{Z}) + g(\mathbf{Y}, \nabla_{\mathbf{X}}\mathbf{Z}).$$

Складывая это выражение с аналогичным выражением для $Y(g(\mathbf{Z}, \mathbf{X}))$ и вычитая аналогичное выражение для $Z(g(\mathbf{X}, \mathbf{Y}))$,

получим

$$g(\mathbf{Z}, \nabla_{\mathbf{X}}\mathbf{Y}) = \frac{1}{2}\{-Z(g(\mathbf{X}, \mathbf{Y})) + Y(g(\mathbf{Z}, \mathbf{X})) + X(g(\mathbf{Y}, \mathbf{Z})) + g(\mathbf{Z}, [\mathbf{X}, \mathbf{Y}]) + g(\mathbf{Y}, [\mathbf{Z}, \mathbf{X}]) - g(\mathbf{X}, [\mathbf{Y}, \mathbf{Z}])\}.$$

Выбирая в качестве $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ базисные векторы, выразим компоненты связности через производные компонент метрики $g_{ab} = g(\mathbf{E}_a, \mathbf{E}_b)$ и производные Ли базисных векторов:

$$\Gamma_{abc} = g\left(\mathbf{E}_a, \nabla_{\mathbf{E}_b}\mathbf{E}_c\right) = g_{ad}\Gamma^{d}{}_{bc}.$$

В частности, при использовании координатного базиса эти производные Ли равны нулю и получаются обычные соотношения Кристоффеля для координатных компонент связности:

$$\Gamma_{abc} = \frac{1}{2}\left\{\frac{\partial g_{ab}}{\partial x^c} + \frac{\partial g_{ac}}{\partial x^b} - \frac{\partial g_{bc}}{\partial x^a}\right\}. \tag{2.26}$$

Далее мы везде будем предполагать, что на $\mathscr{M}$ введена эта единственная C^{r-1}-связность без кручения, определяемая C^r-метрикой $\mathbf{g}$. С ее помощью можно ввести нормальные координаты (разд. 2.5), используя, например, ортонормированный базис в q. Впредь под нормальными координатами мы будем подразумевать нормальные координаты, заданные с помощью ортонормированного базиса.

Тензор Римана для связности, задаваемой метрикой, есть C^{r-2}-тензор, обладающий свойством симметрии

$$R_{(ab)cd} = 0 \Leftrightarrow R_{abcd} = -R_{bacd} \tag{2.27а}$$

в дополнение к симметриям (2.21). Из равенств (2.21) и (2.27а) следует, что тензор Римана симметричен также по парам индексов $\{ab\}$, $\{cd\}$:

$$R_{abcd} = R_{cdab}. \tag{2.27б}$$

Отсюда видно, что тензор Риччи симметричен:

$$R_{ab} = R_{ba}. \tag{2.27в}$$

Свертка тензора Риччи есть по определению *скалярная кривизна*:

$$R = R^{a}{}_{a} = R^{a}{}_{bad}g^{bd}.$$

Вследствие этих симметрий R_{abcd} имеет $(1/12)n^2(n^2-1)$ алгебраически независимых компонент, где n — размерность $\mathscr{M}$. Из них $n(n+1)/2$ компонент можно выразить через компоненты тензора Риччи. Если $n = 1$, то $R_{abcd} = 0$. При $n = 2$ R_{abcd} имеет одну независимую компоненту, которая по существу совпадает с R. При $n = 3$ тензор кривизны полностью определяется тензором Риччи. Если $n > 3$, остальные компоненты тензора кривизны

можно выразить через *тензор Вейля* C_{abcd}, определяемый равенством

$$C_{abcd} = R_{abcd} + \frac{2}{n-2}\{g_{a[d}R_{c]b} + g_{b[c}R_{d]a}\} + \frac{2}{(n-1)(n-2)} R g_{a[c}g_{d]b}.$$

Из того, что последние два слагаемых в правой части обладают свойствами симметрии тензора кривизны, следует, что C_{abcd} также обладает этими свойствами. Легко убедиться, что в дополнение к этому

$$C^a{}_{bad} = 0,$$

т. е. тензор Вейля — это та часть тензора кривизны, все свертки которой равны нулю.

Иначе тензор Вейля можно охарактеризовать тем, что он является конформным инвариантом. Две метрики $\mathbf{g}$ и $\hat{\mathbf{g}}$ называются *конформными*, если можно подобрать некоторую ненулевую функцию Ω, такую, что

$$\hat{\mathbf{g}} = \Omega^2 \mathbf{g}. \tag{2.28}$$

Тогда для любых векторов $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{V}$, $\mathbf{W}$ в точке p

$$\frac{g(\mathbf{X}, \mathbf{Y})}{g(\mathbf{V}, \mathbf{W})} = \frac{\hat{g}(\mathbf{X}, \mathbf{Y})}{\hat{g}(\mathbf{V}, \mathbf{W})},$$

т. е. углы и отношения длин при конформных преобразованиях сохраняются; в частности, конформные преобразования сохраняют структуру изотропного конуса в T_p, поскольку

$$g(\mathbf{X}, \mathbf{X}) > 0, = 0, < 0 \Rightarrow \hat{g}(\mathbf{X}, \mathbf{X}) > 0, = 0, < 0$$

соответственно. Поскольку компоненты метрики связаны соотношениями

$$\hat{g}_{ab} = \Omega^2 g_{ab}, \quad \hat{g}^{ab} = \Omega^{-2} g^{ab},$$

то координатные компоненты связностей, определяемых метриками (2.28), связаны между собой равенством

$$\hat{\Gamma}^a{}_{bc} = \Gamma^a{}_{bc} + \Omega^{-1}\left(\delta^a{}_b \frac{\partial \Omega}{\partial x^c} + \delta^a{}_c \frac{\partial \Omega}{\partial x^b} - g_{bc} g^{ad} \frac{\partial \Omega}{\partial x^d}\right). \tag{2.29}$$

Вычисляя тензор Римана для $\hat{\mathbf{g}}$, получаем

$$\hat{R}^{ab}{}_{cd} = \Omega^{-2} R^{ab}{}_{cd} + \delta^{[a}{}_{[c}\Omega^{b]}{}_{d]},$$

где

$$\Omega^a{}_b = 4\Omega^{-1}(\Omega^{-1})_{;be} g^{ae} - 2(\Omega^{-1})_{;c}(\Omega^{-1})_{;d} g^{cd}\delta^a{}_b,$$

причем ковариантные производные в этом уравнении взяты относительно метрики $\mathbf{g}$. Отсюда (при условии $n > 2$)

$$\hat{R}^b{}_d = \Omega^{-2} R^b{}_d + (n-2)\,\Omega^{-1}(\Omega^{-1})_{;dc}\,g^{bc} - (n-2)^{-1}\Omega^{-n}(\Omega^{n-2})_{;ac}\,g^{ac}\delta^b{}_d$$

и

$$\hat{C}^a{}_{bcd} = C^a{}_{bcd}.$$

Это последнее равенство выражает тот факт, что тензор Вейля конформно-инвариантен. Из полученных соотношений следует, что

$$\hat{R} = \Omega^{-2}R - 2(n-1)\,\Omega^{-3}\Omega_{;cd}g^{cd} - (n-1)(n-4)\,\Omega_{;c}\Omega_{;d}g^{cd}. \quad (2.30)$$

Расщепив тензор Римана на две части, выражаемые соответственно через тензор Риччи и тензор Вейля, можно с помощью тождеств Бианки (2.22) получить дифференциальные соотношения между тензором Вейля и тензором Риччи; свертывая (2.22), имеем

$$R^a{}_{bcd;a} = R_{bd;c} - R_{bc;d}, \quad (2.31)$$

а после вторичной свертки

$$R^a{}_{c;a} = \frac{1}{2}R_{;c}.$$

Пользуясь определением тензора Вейля (при $n > 3$), можно переписать (2.31) в виде

$$C^a{}_{bcd;a} = 2\frac{n-3}{n-2}\left(R_{b[d;c]} - \frac{1}{2(n-1)}\,g_{b[d}R_{;c]}\right). \quad (2.32)$$

Если $n \leqslant 4$, равенство (2.31) содержит всю информацию, имеющуюся в тождествах Бианки (2.22) и, таким образом, при $n = 4$ соотношение (2.32) эквивалентно этим тождествам.

Диффеоморфизм $\phi: \mathscr{M} \to \mathscr{M}$ будем называть *изометрией*, если он переводит метрику в себя, иначе говоря, если отображенная метрика $\phi_*\mathbf{g}$ равна $\mathbf{g}$ в каждой точке. Тогда отображение $\phi_*: T_p \to T_{\phi(p)}$ сохраняет скалярные произведения

$$g(\mathbf{X}, \mathbf{Y})|_p = \phi_* g(\phi_*\mathbf{X}, \phi_*\mathbf{Y})|_{\phi(p)} = g(\phi_*\mathbf{X}, \phi_*\mathbf{Y})|_{\phi(p)}.$$

Если данная локальная однопараметрическая группа диффеоморфизмов ϕ_t, порожденная векторным полем $\mathbf{K}$, есть группа изометрий (т. е. при каждом t преобразование ϕ_t есть изометрия), то $\mathbf{K}$ будем называть *векторным полем Киллинга*. Производная Ли метрики относительно $\mathbf{K}$ равна нулю:

$$L_{\mathbf{K}}\mathbf{g} = \lim_{t\to 0}\frac{1}{t}(\mathbf{g} - \phi_{t*}\mathbf{g}) = 0,$$

так как $\mathbf{g} = \phi_{t*}\mathbf{g}$ при каждом t; но, согласно (2.17), $L_{\mathbf{K}}\,g_{ab} = 2K_{(a;b)}$, и, следовательно, векторное поле $\mathbf{K}$ Киллинга удовлетворяет

уравнению Киллинга

$$K_{a;b} + K_{b;a} = 0. \tag{2.33}$$

Обратно, если $\mathbf{K}$ — векторное поле, удовлетворяющее уравнению Киллинга, то $L_{\mathbf{K}}\mathbf{g} = 0$ и

$$\phi_{t*}\mathbf{g}\,|_p = \mathbf{g}\,|_p + \int_0^t \frac{\mathrm{d}}{\mathrm{d}t'}(\phi_{t'*}\mathbf{g})\,|_p\,\mathrm{d}t' = \mathbf{g}\,|_p + \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}(\phi_{t'*}\phi_{s*}\mathbf{g})_{s=0}\,|_p\,\mathrm{d}t' =$$

$$= \mathbf{g}\,|_p + \int_0^t \left(\phi_{t'*}\,\frac{\mathrm{d}}{\mathrm{d}s}\,\phi_{s*}\mathbf{g}\right)_{s=0}\Big|_p \mathrm{d}t' =$$

$$= \mathbf{g}\,|_p - \int_0^t \phi_{t'*}\left(L_{\mathbf{K}}\mathbf{g}\,|_{\phi_{-t'}(p)}\right) dt' = \mathbf{g}\,|_p.$$

Итак, для того, чтобы $\mathbf{K}$ было векторным полем Киллинга, необходимо и достаточно, чтобы оно удовлетворяло уравнению Киллинга. Тогда локально можно выбрать координаты $x^a = (x^\nu, t)$ $(\nu = 1, \ldots, n-1)$, в которых $K^a = \partial x^a/\partial t = \delta^a{}_n$; в этих координатах уравнения Киллинга принимают вид

$$\partial g_{ab}/\partial t = 0 \Leftrightarrow g_{ab} = g_{ab}(x^\nu).$$

В общем случае пространство не обладает какими-либо симметриями и поэтому не допускает никаких векторных полей Киллинга. Пространства частного вида могут допускать r линейно-независимых векторных полей $\mathbf{K}_a$ $(a = 1, \ldots, r)$. Можно показать, что совокупность всех векторных полей Киллинга на таких пространствах образует r-мерную алгебру Ли над R, $0 \leqslant r \leqslant \frac{1}{2}n(n+1)$ (верхний предел может быть меньше, если метрика вырождена), причем умножение в этой алгебре задается скобками Ли $[\ ,\ ]$ [см. (2.16)]. Локальная группа диффеоморфизмов, порожденных этими векторными полями, есть r-мерная группа Ли изометрий на многообразии $\mathscr{M}$. Полная группа изометрий на $\mathscr{M}$ может включать некоторые дискретные изометрии (такие, как отражения в плоскости), которые не порождаются векторными полями Киллинга; свойства симметрии пространства целиком определяются этой полной группой изометрий.

2.7. Гиперповерхности

Пусть $\mathscr{S}$ есть $(n-1)$-мерное многообразие и $\theta: \mathscr{S} \to \mathscr{M}$ — некоторое вложение; тогда образ $\theta(\mathscr{S})$ многообразия $\mathscr{S}$ называют *гиперповерхностью* в $\mathscr{M}$. Если $p \in \mathscr{S}$, образ T_p в $T_{\theta(p)}$ при отображении θ_* будет $(n-1)$-мерной плоскостью, проходящей через

начало. Следовательно, существует ненулевая линейная форма $\mathbf{n} \in T^*_{\theta(p)}$, такая, что для любого вектора $\mathbf{X} \in T_p$ $\langle \mathbf{n}, \theta_* \mathbf{X} \rangle = 0$. Форма $\mathbf{n}$ единственна с точностью до знака и нормировочного множителя, и если локально задать $\theta(\mathscr{S})$ уравнением $f = 0$, так что $df \neq 0$, то в качестве n можно локально взять df. Если $\theta(\mathscr{S})$ — двустороннее подмногообразие в $\mathscr{M}$, то $\mathbf{n}$ можно выбрать так, что полученное поле 1-формы на $\theta(\mathscr{S})$ будет всюду отлично от нуля. Это будет иметь место, если и $\mathscr{S}$, и $\mathscr{M}$ — ориентируемые многообразия. В таком случае выбор направления $\mathbf{n}$ связывает ориентации $\theta(\mathscr{S})$ и $\mathscr{M}$: пусть $\{x^i\}$ — такие локальные координаты из ориентированного атласа на $\mathscr{M}$, в которых локально уравнение $\theta(\mathscr{S})$ имеет вид $x^1 = 0$ и $\mathbf{n} = \alpha dx^1$, где $\alpha > 0$, тогда $(x^2, \ldots, x^n)$ будут ориентированными локальными координатами на $\theta(\mathscr{S})$.

Если $\mathbf{g}$ есть метрика на $\mathscr{M}$, вложение индуцирует метрику $\theta^*\mathbf{g}$ на $\mathscr{S}$, так что для $\mathbf{X}, \mathbf{Y} \in T_p \theta^* g(\mathbf{X}, \mathbf{Y}) \mid_p = g(\theta_* \mathbf{X}, \theta_* \mathbf{Y}) \mid_{\theta(p)}$. Иногда эту метрику называют первой фундаментальной формой на $\mathscr{S}$. Если $\mathbf{g}$ — положительно определенная метрика, то и $\theta^*\mathbf{g}$ будет положительно определенной. Если же $\mathbf{g}$ — лоренцева метрика, то $\theta^*\mathbf{g}$ будет

а) лоренцевой при $g^{ab} n_a n_b > 0$,

б) вырожденной при $g^{ab} n_a n_b = 0$,

в) положительно определенной при $g^{ab} n_a n_b < 0$.

В случае (а) гиперповерхность $\theta(\mathscr{S})$ называется *времениподобной*, в случае (б) — *изотропной* и в случае (в) *пространственноподобной*.

Чтобы убедиться в существовании этих возможностей, рассмотрим вектор $N^b = n_a g^{ab}$; он ортогонален ко всем векторам, касательным к $\theta(\mathscr{S})$, т. е. ко всем векторам подпространства $H = \theta_*(T_p)$ в $T_{\theta(p)}$. Предположим сначала, что сам вектор $\mathbf{N}$ не лежит в этом подпространстве. Тогда, если $(\mathbf{E}_2, \ldots, \mathbf{E}_n)$ образуют базис в T_p, векторы $\mathbf{N}, \theta_*(\mathbf{E}_2), \ldots, \theta_*(\mathbf{E}_n)$ линейно независимы и составляют базис в $T_{\theta(p)}$. Компоненты $\mathbf{g}$ относительно этого базиса равны

$$g_{ab} = \begin{pmatrix} g(\mathbf{N}, \mathbf{N}) & 0 \\ 0 & g(\theta_*(\mathbf{E}_i), \theta_*(\mathbf{E}_j)) \end{pmatrix} = \begin{pmatrix} g(\mathbf{N}, \mathbf{N}) & 0 \\ 0 & \theta^* g(\mathbf{E}_i, \mathbf{E}_j) \end{pmatrix}.$$

В силу предположения о невырожденности $\mathbf{g}$ отсюда следует, что $g(\mathbf{N}, \mathbf{N}) \neq 0$. Если $\mathbf{g}$ положительно определена, то $g(\mathbf{N}, \mathbf{N}) > 0$ и, следовательно, индуцированная метрика $\theta^*\mathbf{g}$ тоже должна быть положительно определенной. Если $\mathbf{g}$ лоренцева и $g(\mathbf{N}, \mathbf{N}) = g^{ab} n_a n_b < 0$, то $\theta^*\mathbf{g}$ необходимо положительно определена, так как матрица из компонент $\mathbf{g}$ имеет только одно отрицательное собственное число. Аналогично, если $g(\mathbf{N}, \mathbf{N}) = g^{ab} n_a n_b > 0$, то $\theta^*\mathbf{g}$ — лоренцева метрика. Предположим теперь, что вектор $\mathbf{N}$

касателен к $\theta(\mathscr{S})$. Тогда найдется ненулевой вектор $\mathbf{X} \in T_p$, такой, что $\theta_*(\mathbf{X}) = \mathbf{N}$; но $g(\mathbf{N}, \theta_*\mathbf{Y}) = 0$ для всех $\mathbf{Y} \in T_p$, и, следовательно, $\theta^* g(\mathbf{X}, \mathbf{Y}) = 0$. Таким образом, метрика $\theta^*\mathbf{g}$ вырождена. Кроме того, полагая $\mathbf{Y} = \mathbf{X}$, имеем $g(\mathbf{N}, \mathbf{N}) = g^{ab} n_a n_b = 0$.

Если $g^{ab} n_a n_b \neq 0$, нормальную линейную форму $\mathbf{n}$ можно нормировать так, чтобы она имела единичную „длину“, т. е. $g^{ab} n_a n_b = \pm 1$. В этом случае отображение $\theta^*: T^*_{\theta(p)} \to T^*_p$ будет однозначным в $(n-1)$-мерном подпространстве $H^*_{(\theta, p)}$ пространства $T^*_{\theta(p)}$, состоящем из всех линейных форм $\boldsymbol{\omega}$ в $\theta(p)$, для которых $g^{ab} n_a \omega_b = 0$; это следует из того, что $\theta^*\mathbf{n} = 0$ и $\mathbf{n}$ не лежит в H^*. Поэтому обратное отображение $(\theta^*)^{-1}$ будет отображением $\tilde{\theta}_*$ пространства T^*_p на $H^*_{\theta(p)}$ и, таким образом, на $T^*_{\theta(p)}$.

Это отображение можно расширить обычным образом до отображения ковариантных тензоров в $\mathscr{S}$ в ковариантные тензоры в $\theta(\mathscr{S}) \subset \mathscr{M}$; поскольку у нас есть отображение θ_* контравариантных тензоров в $\mathscr{S}$ в ковариантные тензоры в $\theta(\mathscr{S})$, мы можем расширить θ_* до отображения $\tilde{\theta}_*$ произвольных тензоров в $\mathscr{S}$ в тензоры в $\theta(\mathscr{S})$. Это отображение обладает тем свойством, что свертка $\tilde{\theta}_*\mathbf{T}$ с $\mathbf{n}$ по любому индексу равна нулю, т. е.

$$(\tilde{\theta}_* T)^{a \cdots b}{}_{c \ldots d} n_a = 0 \quad \text{и} \quad (\tilde{\theta}_* T)^{a \cdots b}{}_{c \ldots d} g^{ce} n_e = 0$$

для любого тензора $\mathbf{T} \in T^r_s(\mathscr{S})$.

Рассмотрим тензор $\mathbf{h}$ на $\theta(\mathscr{S})$, определяемый как $\mathbf{h} = \tilde{\theta}_*(\theta^*\mathbf{g})$. Его можно записать через нормированную линейную форму $\mathbf{n}$ (напомним, что $g^{ab} n_a n_b = \pm 1$) в виде

$$h_{ab} = g_{ab} \mp n_a n_b,$$

поскольку отсюда следует, что $\theta^*\mathbf{h} = \theta^*\mathbf{g}$ и $h_{ab} g^{bc} n_c = 0$.

Тензор $h^a{}_b = g^{ac} h_{cb}$ есть оператор проектирования, т. е. $h^a{}_b h^b{}_c = h^a{}_c$. Он проектирует вектор $\mathbf{X} \in T_{\theta(p)}$ на ту его часть, которая лежит в подпространстве $H = \theta_*(T_p)$ касательного к $\theta(\mathscr{S})$ пространства $T_{\theta(p)}$:

$$X^a = h^a{}_b X^b \pm n^a n_b X^b,$$

где второе слагаемое представляет собой часть вектора $\mathbf{X}$, ортогональную к $\theta(\mathscr{S})$. Тензор $h^a{}_b$ проектирует и линейную форму $\boldsymbol{\omega} \in T^*_{\theta(p)}$ на ее часть, лежащую в подпространстве $H^*_{\theta(p)}$:

$$\omega_a = h^b{}_a \omega_b \pm n_a n^b \omega_b.$$

Подобным же образом можно спроектировать любой тензор $\mathbf{T} \in T^r_s(\theta(p))$ на его часть, принадлежащую пространству

$$H^r_s(\theta(p)) = \underbrace{H_{\theta(p)} \otimes \ldots \otimes H_{\theta(p)}}_{r \text{ множителей}} \otimes \underbrace{H^*_{\theta(p)} \otimes \ldots \otimes H^*_{\theta(p)}}_{s \text{ множителей}},$$

т. е. часть, ортогональную к $\mathbf{n}$ по всем индексам.

θ_* является взаимно однозначным отображением из T_p в $H_{\theta(p)}$. Поэтому отображение θ^* из $T_{\theta(p)}$ в T_p можно задать, проектируя сначала $T_{\theta(p)}$ на $H_{\theta(p)}$ посредством $h^a{}_b$ и выполняя затем обратное отображение $(\theta)^{-1}_*$. Поскольку мы уже определили отображение θ^* линейных форм в $\theta(\mathscr{S})$, можно расширить определение θ^* до отображения $\tilde{\theta}^*$ тензоров на $\theta(\mathscr{S})$ любого типа в тензоры на многообразии $\mathscr{S}$. Это отображение обладает тем свойством, что $\tilde{\theta}^*(\tilde{\theta}_*\mathbf{T}) = \mathbf{T}$ для любого тензора $\mathbf{T} \in T^r_s(p)$ и $\tilde{\theta}_*(\tilde{\theta}^*\mathbf{T}) = \mathbf{T}$ для любого тензора $\mathbf{T} \in H^r_s(\theta(p))$. Мы будем отождествлять тензоры на $\mathscr{S}$ с тензорами из H^r_s на $\theta(\mathscr{S})$, если они переходят друг в друга при отображениях $\tilde{\theta}_*$, $\tilde{\theta}^*$. В частности, $\mathbf{h}$ можно рассматривать как индуцированную метрику на $\theta(\mathscr{S})$.

Пусть $\bar{\mathbf{n}}$ есть какое-либо расширение единичной нормали $\mathbf{n}$ на некоторую открытую окрестность гиперповерхности $\theta(\mathscr{S})$; тогда тензор $\boldsymbol{\chi}$, определенный на $\theta(\mathscr{S})$ формулой

$$\chi_{ab} = h^c{}_a h^d{}_b \bar{n}_{c;d},$$

называется *второй фундаментальной формой* на $\mathscr{S}$. Поскольку проектирование с помощью $h^a{}_b$ ограничивает ковариантные производные касательными к $\theta(\mathscr{S})$ направлениями, $\boldsymbol{\chi}$ не зависит от способа расширения. Локально поле можно представить в виде $\bar{\mathbf{n}} = \alpha\, \mathrm{d}f$, где f и α — функции на $\mathscr{M}$, причем на $\theta(\mathscr{S})$ $f = 0$. Поскольку

$$f_{;ab} = f_{;ba} \quad \text{и} \quad f_{;a} h^a{}_b = 0,$$

тензор χ_{ab} симметричен.

Индуцированная на $\mathscr{S}$ метрика $\mathbf{h} = \theta^*\mathbf{g}$ задает на $\mathscr{S}$ связность. Ковариантное дифференцирование относительно этой связности мы будем обозначать двойной чертой $\|$. Для любого тензора $\mathbf{T} \in H^r_s$ справедливо равенство

$$T^{a \ldots b}{}_{c \ldots d \| e} = \bar{T}^{i \ldots j}{}_{k \ldots l;m} h^a{}_i \ldots h^b{}_j h^k{}_c \ldots h^l{}_d h^m{}_e,$$

где $\bar{\mathbf{T}}$ — любое расширение $\mathbf{T}$ на некоторую окрестность $\theta(\mathscr{S})$. Это определение не зависит от выбора расширения, так как свертывание с $h^a{}_i$ ограничивает ковариантное дифференцирование направлениями, касательными к $\theta(\mathscr{S})$. Чтобы убедиться в пра-

вильности приведенной формулы, достаточно показать, что ковариантная производная индуцированной метрики равна нулю и кручение отсутствует. Это следует из равенств:

$$h_{ab\|c} = (g_{ef} \mp \bar{n}_e \bar{n}_f)_{;g} h^e{}_a h^f{}_b h^g{}_c = 0$$

и

$$f_{\|ab} = h^e{}_a h^g{}_b f_{;eg} = h^e{}_a h^g{}_b f_{;ge} = f_{\|ba}.$$

Тензор кривизны $R'^a{}_{bcd}$ относительно индуцированной метрики можно следующим образом связать с тензором кривизны $R^a{}_{bcd}$ на $\theta(\mathscr{S})$ и второй фундаментальной формой $\boldsymbol{\chi}$. Пусть $\mathbf{Y} \in H$ — векторное поле на $\theta(\mathscr{S})$; тогда

$$R'^a{}_{bcd} Y^b = Y^a{}_{\|dc} - Y^a{}_{\|cd}.$$

Имеем

$$Y^a{}_{\|dc} = (Y^a{}_{\|d})_{\|c} = (\bar{Y}^e{}_{;f} h^g{}_e h^f{}_i)_{;k} h^a{}_g h^i{}_d h^k{}_c =$$

$$= \bar{Y}^e{}_{;fk} h^a{}_e h^f{}_d h^k{}_c \mp \bar{Y}^e{}_{;f} \bar{n}_e \bar{n}^g{}_{;k} h^f{}_d h^a{}_g h^k{}_c \mp \bar{Y}^e{}_{;f} \bar{n}^f \bar{n}_{i;k} h^a{}_e h^i{}_d h^k$$

и, поскольку на $\theta(\mathscr{S})$ $\bar{Y}^e \bar{n}_e = 0$,

$$\bar{Y}^e{}_{;f} \bar{n}_e h^f{}_d = (\bar{Y}^e \bar{n}_e)_{;f} h^f{}_d - \bar{Y}^e \bar{n}_{e;f} h^f{}_d = - \bar{Y}^e \bar{n}_{e;f} h^f{}_d.$$

Отсюда

$$R'^a{}_{bcd} Y^b = (R^e{}_{bkf} h^a{}_e h^k{}_c h^f{}_d \pm \chi_{bd} \chi^a{}_c \mp \chi_{bc} \chi^a{}_d) Y^b,$$

и так как это равенство справедливо для всех $\mathbf{Y} \in H$, то

$$R'^a{}_{bcd} = R^e{}_{fgh} h^a{}_e h^f{}_b h^g{}_c h^h{}_d \pm \chi^a{}_c \chi_{bd} \mp \chi^a{}_d \chi_{bc}. \tag{2.34}$$

Это соотношение называют *уравнением Гаусса.*

Свертывая это уравнение по a и c и умножая на h^{bd}, получаем скалярную кривизну R' относительно индуцированной метрики

$$R' = R \mp 2R_{ab} n^a n^b \pm (\chi^a{}_a)^2 \mp \chi^{ab} \chi_{ab}. \tag{2.35}$$

Можно вывести иное соотношение между второй фундаментальной формой и тензором кривизны $R^a{}_{bcd}$ на $\theta(\mathscr{S})$, если из равенства

$$(\chi^a{}_a)_{\|b} = (\bar{n}^a{}_{;d} h^d{}_a)_{;e} h^e{}_b$$

вычесть равенство

$$(\chi^a{}_b)_{\|a} = (\bar{n}^c{}_{;d} h^a{}_c h^d{}_e)_{;f} h^f{}_a h^e{}_b.$$

В результате получаем уравнение

$$\chi^a{}_{b\|a} - \chi^a{}_{a\|b} = R_{ef} n^f h^e{}_b, \tag{2.36}$$

называемое *уравнением Кодацци.*

2.8. Элемент объема и теорема Гаусса

Пусть $\{\mathbf{E}^a\}$ — базис 1-форм; из него можно образовать n-форму

$$\boldsymbol{\varepsilon} = n!\mathbf{E}^1 \wedge \mathbf{E}^2 \wedge \ldots \wedge \mathbf{E}^n.$$

Если $\{\mathbf{E}^{a'}\}$ — другой базис, связанный с $\{\mathbf{E}^a\}$ соотношением $\mathbf{E}^{a'} = \Phi^{a'}{}_a \mathbf{E}^a$, то $\boldsymbol{\varepsilon}$ и n-форма $\boldsymbol{\varepsilon}'$, определяемая базисом $\{\mathbf{E}^{a'}\}$, связаны равенством

$$\boldsymbol{\varepsilon}' = \det\left(\Phi^{a'}{}_a\right)\boldsymbol{\varepsilon},$$

и, следовательно, эта форма не единственна. Однако наличие метрики можно использовать для введения n-формы

$$\boldsymbol{\eta} = |g|^{1/2}\boldsymbol{\varepsilon},$$

где $g \equiv \det(g_{ab})$. Компоненты этой n-формы равны

$$\eta_{ab\ldots d} = n!\,|g|^{1/2}\delta^1{}_{[a}\delta^2{}_b \ldots \delta^n{}_{d]}.$$

Согласно закону преобразования g, определитель $\det\Phi^{a'}{}_a$ сокращается при условии, что он положителен. Поэтому, если $\mathscr{M}$ ориентируемо, n-формы $\boldsymbol{\eta}$, определенные координатными базисами некоторого ориентированного атласа, будут тождественны. Иначе говоря, если задана ориентация $\mathscr{M}$, можно однозначно определить поле n-форм $\boldsymbol{\eta}$ (каноническую n-форму) на $\mathscr{M}$.

Контравариантный антисимметричный тензор

$$\eta^{ab\ldots d} = g^{ae}g^{bf}\ldots g^{dh}\eta_{ef\ldots h}$$

имеет компоненты

$$\eta^{ab\ldots d} = (-1)^{(n-s)/2}n!\,|g|^{1/2}\delta^{[a}{}_1\delta^b{}_2\ldots\delta^{d]}{}_n,$$

где s — сигнатура $\mathbf{g}$ (и, следовательно, число $(n-s)/2$ равно числу отрицательных собственных значений матрицы (g_{ab}) из компонент метрики). Поэтому эти тензоры удовлетворяют соотношениям

$$\eta^{ab\ldots d}\eta_{ef\ldots h} = (-1)^{\frac{n-s}{2}}n!\delta^a{}_{[e}\delta^b{}_f\ldots\delta^d{}_{h]}. \tag{2.37}$$

Из соотношений Кристоффеля следует, что ковариантные производные $\eta_{ab\ldots d}$ и $\eta^{ab\ldots d}$ относительно связности, определяемой метрикой, равны нулю, т. е.

$$\eta^{ab\ldots d}{}_{;e} = 0 = \eta_{ab\ldots d;e}.$$

С помощью канонической n-формы можно определить объем (относительно метрики $\mathbf{g}$) n-мерного подмногообразия $\mathscr{U} \subset \mathscr{M}$ как

$\int_{\mathcal{U}} \boldsymbol{\eta}$. Таким образом, $\boldsymbol{\eta}$ можно рассматривать как положительно определенную меру объема на $\mathcal{M}$. Мы будем часто пользоваться ею в этом смысле, обозначая ее $\mathrm{d}v$. Отметим, что здесь d не означает оператор внешнего дифференцирования; $\mathrm{d}v$ есть просто мера на $\mathcal{M}$. Если f — функция на $\mathcal{M}$, можно определить ее интеграл по $\mathcal{U}$ относительно этой меры:

$$\int_{\mathcal{U}} f \, \mathrm{d}v = \int_{\mathcal{U}} f\boldsymbol{\eta}.$$

Относительно локальных ориентированных координат $\{x^a\}$ этот интеграл можно представить в виде кратного интеграла

$$\int_{\mathcal{U}} f |g|^{1/2} \mathrm{d}x^1 \mathrm{d}x^2 \ldots \mathrm{d}x^n,$$

который инвариантен относительно замены координат.

Свертка векторного поля $\mathbf{X}$ на $\mathcal{M}$ с $\boldsymbol{\eta}$ будет полем 1-формы $\mathbf{X}\cdot\boldsymbol{\eta}$, причем

$$(\mathbf{X}\cdot\boldsymbol{\eta})_{b\ldots d} = X^a \eta_{ab\ldots d}.$$

Эту $(n-1)$-форму можно интегрировать по любому $(n-1)$-мерному компактному ориентируемому подмногообразию $\mathcal{V}$. Этот интеграл запишем в виде

$$\int_{\mathcal{V}} X^a \, \mathrm{d}\sigma_a = \int_{\mathcal{V}} \mathbf{X}\cdot\boldsymbol{\eta},$$

причем мы считаем, что форма $\boldsymbol{\eta}$, являющаяся мерой на $\mathcal{V}$, определяется канонической формой $\mathrm{d}\sigma_a$. Если ориентация $\mathcal{V}$ задана направлением нормальной формы n_a, то $\mathrm{d}\sigma_a$ можно представить в виде $n_a \mathrm{d}\sigma$, где $\mathrm{d}\sigma$ — положительно определенная мера объема на подмногообразии $\mathcal{V}$. Пока нормаль n_a не нормирована, мера объема $\mathrm{d}\sigma$ определена неоднозначно. При нормировке на n_a на единицу по метрике $\mathbf{g}$ на $\mathcal{M}$, т. е. если $n_a n_b g^{ab} = \pm 1$, $\mathrm{d}\sigma$ становится равной мере объема на $\mathcal{V}$, которую задает на $\mathcal{V}$ индуцированная метрика (чтобы убедиться в этом, достаточно выбрать ортонормированный базис с $n_a g^{ab}$ в качестве одного из базисных векторов).

С помощью канонической формы из теоремы Стокса можно вывести формулу Гаусса: для любого компактного n-мерного подмногообразия $\mathcal{U} \subset \mathcal{M}$

$$\int_{\partial\mathcal{U}} X^a \, \mathrm{d}\sigma_a = \int_{\partial\mathcal{U}} \mathbf{X}\cdot\boldsymbol{\eta} = n \int_{\mathcal{U}} \mathrm{d}(\mathbf{X}\cdot\boldsymbol{\eta}).$$

Однако

$$(\mathrm{d}(\mathbf{X}\cdot\boldsymbol{\eta}))_{a\ldots de}=(-1)^{n-1}\left(X^{g}\eta_{g[a\ldots d)};e]=$$

$$=(-1)^{n-1}\delta^{s}{}_{[a}\ldots\delta^{t}{}_{d}\delta^{u}{}_{e]}\eta_{gs\ldots t}X^{g}{}_{;u}=$$

$$=(-1)^{(n-1)-\frac{1}{2}(n-s)}\frac{1}{n!}\eta^{s\ldots tu}\eta_{a\ldots de}\eta_{gs\ldots t}X^{g}{}_{;u}=$$

$$=\eta_{a\ldots de}\delta^{s}{}_{[s}\ldots\delta^{t}{}_{t}\delta^{u}{}_{g]}X^{g}{}_{;u}=n^{-1}\eta_{a\ldots de}X^{g}{}_{;g},$$

где дважды использовано равенство (2.37). Отсюда для любого векторного поля $\mathbf{X}$

$$\int\limits_{\partial\mathcal{U}}X^{a}\,\mathrm{d}\sigma_{a}=\int\limits_{\mathcal{U}}X^{g}{}_{;g}\,\mathrm{d}v;$$

это и есть теорема Гаусса. Отметим, что эта теорема справедлива при такой ориентации в $\mathcal{U}$, задаваемой нормальной формой $\boldsymbol{\eta}$, при которой $\langle\mathbf{n},\mathbf{X}\rangle>0$ для вектора $\mathbf{X}$, направленного наружу $\mathcal{U}$. Если метрика $\mathbf{g}$ такова, что $g^{ab}n_{a}n_{b}<0$, вектор $g^{ab}n_{b}$ направлен внутрь $\mathcal{U}$.

2.9. Расслоенные пространства

Некоторые из геометрических свойств многообразия $\mathcal{M}$ легче всего рассматривать, построив многообразие, называемое расслоенным пространством. Локально оно представляет собой прямое произведение $\mathcal{M}$ и некоторого подходящего пространства. В этом разделе мы дадим определение расслоенного пространства и рассмотрим четыре примера, которые будут полезны в дальнейшем: касательное расслоение $T(\mathcal{M})$, тензорное расслоение $T^{r}_{s}(\mathcal{M})$, расслоение линейных реперов $L(\mathcal{M})$ и расслоение ортонормированных реперов $O(\mathcal{M})$.

C^{k}-расслоение над C^{s}-многообразием $\mathcal{M}$ $(s\geqslant k)$ есть C^{k}-многообразие $\mathcal{E}$ и сюръективное C^{k}-отображение $\pi:\mathcal{E}\to\mathcal{M}$. Многообразие $\mathcal{E}$ называется пространством расслоения, $\mathcal{M}$ — базой и π — проекцией. Там, где это не приведет к путанице, мы будем обозначать расслоение просто через $\mathcal{E}$. Вообще говоря, образ $\pi^{-1}(p)$ точки $p\in\mathcal{M}$ при обратном отображении не обязательно гомеоморфен $\pi^{-1}(q)$ для другой точки $q\in\mathcal{M}$. Простейший пример расслоения — *прямое произведение* $(\mathcal{M}\times\mathcal{A},\mathcal{M},\pi)$, в котором $\mathcal{A}$ — некоторое многообразие, а проекция π определяется как $\pi(p,v)=p$ для всех $p\in\mathcal{M}$, $v\in\mathcal{A}$. Например, выбирая в качестве $\mathcal{M}$ окружность и в качестве $\mathcal{A}$ действительную прямую R^{1}, мы получим прямое произведение над S^{1} — цилиндр C^{2}.

Расслоение называется расслоенным пространством, если оно локально является прямым произведением. Следовательно, рас-

слоение есть *расслоенное пространство* со слоем $\mathcal{F}$, если существует такая окрестность $\mathcal{U}$ каждой точки $q \in \mathcal{M}$, что окрестность $\pi^{-1}(\mathcal{U})$ изоморфна $\mathcal{U} \times \mathcal{F}$ в том смысле, что для каждой точки $p \in \mathcal{U}$ найдется диффеоморфизм ϕ_p точки $\pi^{-1}(p)$ на $\mathcal{F}$, такой, что отображение ψ, определяемое как $\psi(u) = (\pi(u), \phi_{\pi(u)})$, есть диффеоморфизм $\psi: \pi^{-1}(\mathcal{U}) \to \mathcal{U} \times \mathcal{F}$. Поскольку $\mathcal{M}$ паракомпактно, мы можем выбрать локально конечное покрытие $\mathcal{M}$ такими открытыми множествами $\mathcal{U}_\alpha$. Если $\mathcal{U}_\alpha$ и $\mathcal{U}_\beta$ — два множества такого покрытия, отображение

$$(\phi_{\alpha, p}) \circ (\phi_{\beta, p})^{-1}$$

есть диффеоморфизм $\mathcal{F}$ на себя для каждого $p \in (\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$. Поэтому образы $\pi^{-1}(p)$ точек $p \in \mathcal{M}$ все необходимо диффеоморфны $\mathcal{F}$ (и, значит, друг другу). Например, лист Мёбиуса есть расслоенное пространство на S^1 со слоем R^1; нужны два открытых множества $\mathcal{U}_1$, $\mathcal{U}_2$, чтобы получить покрытие множествами вида $\mathcal{U}_i \times R^1$. Из этого примера видно, что многообразие, являющееся локально прямым произведением двух других многообразий, в общем случае тем не менее не будет прямым произведением многообразий. Именно по этой причине понятие расслоенного пространства столь полезно.

Касательное расслоение $T(\mathcal{M})$ есть расслоенное пространство над C^k-многообразием, получающееся при наделении множества $\mathcal{E} = \bigcup_{p \in \mathcal{M}} T_p$ естественной для него структурой многообразия и естественной проекцией на $\mathcal{M}$. Таким образом, проекция π отображает каждую точку T_p на p. Структура многообразия на $\mathcal{E}$ определяется локальными координатами $\{z^A\}$ следующим образом. Пусть $\{x^i\}$ — локальные координаты на открытом множестве $\mathcal{U} \subset \mathcal{M}$. Тогда любой вектор $\mathbf{V} \in T_p$ можно представить в виде $\mathbf{V} = V^i \partial/\partial x^i|_p$. Координатами $\{z^A\}$ в $\pi^{-1}(\mathcal{U})$ будут $\{z^A\} = \{x^i, V^a\}$. Если выбрано покрытие $\mathcal{M}$ некоторыми координатными окрестностями $\mathcal{U}_\alpha$, соответствующие карты задают C^{k-1}-атлас на $\mathcal{E}$, при этом $\mathcal{E}$ становится C^{k-1}-многообразием (n^2-мерным). Чтобы убедиться в этом, нужно только заметить, что в любом пересечении $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ координаты $\{x^i{}_\alpha\}$ некоторой точки являются C^k-функциями координат $\{x^i{}_\beta\}$ той же точки, а компоненты $\{V^a{}_\alpha\}$ некоторого векторного поля будут C^{k-1}-функциями компонент $\{V^a{}_\beta\}$ того же поля. Тогда координаты $\{z^A{}_\alpha\}$ будут в $\pi^{-1}(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$ C^{k-1}-функциями координат $\{z^A{}_\beta\}$.

Слой $\pi^{-1}(p)$ есть T_p и, следовательно, является векторны пространством размерности n. Структура векторного пространства

сохраняется при отображении $\phi_{\alpha, p}: T_p \to R^n$, которое определяется как $\phi_{\alpha, p}(u) = V^a{}_\alpha(u)$, т. е. $\phi_{\alpha, \beta}$ отображает вектор в точке p на его компоненты относительно координат $\{x^a{}_\alpha\}$. Если $\{x^a{}_\beta\}$ — другая система локальных координат, отображение $(\phi_{\alpha, p}) \circ (\phi_{\beta, p})^{-1}$ есть линейное отображение R^n на себя и, следовательно, оно является элементом общей линейной группы $GL(n, R)$.

Тензорное расслоение типа (r, s) над $\mathcal{M}$, обозначаемое через $T^r_s(\mathcal{M})$, определяется подобным же образом. Строим множество $\mathcal{E} = \bigcup_{p \in \mathcal{M}} T^r_s(p)$, задаем проекцию π как отображение каждой точки $T^r_s(p)$ на p и для любой координатной окрестности $\mathcal{U} \subset \mathcal{M}$ вводим в $\pi^{-1}(\mathcal{U})$ локальные координаты $\{z^A\}$ как $\{z^A\} = \{x^i, T^{a \dots b}{}_{c \dots d}\}$, где $\{x^i\}$ — координаты точки p, $\{T^{a \dots b}{}_{c \dots d}\}$ — координатные компоненты $\mathbf{T}$ (т. е. $\mathbf{T} = T^{a \dots b}{}_{c \dots d} \partial/\partial x^a \otimes \dots \otimes \mathrm{d}x^d|_p$). Этим мы превращаем $\mathcal{E}$ в C^{k-1}-многообразие размерности n^{r+s+1}; любая точка $u \in T^r_s(\mathcal{M})$ соответствует единственному тензору $\mathbf{T}$ типа (r, s) в точке $\pi(u)$.

Расслоение линейных реперов $L(\mathcal{M})$ есть расслоенное C^{k-1}-пространство, определяемое следующим образом. Полное пространство $\mathcal{E}$ состоит из всех базисов во всех точках $\mathcal{M}$, иначе говоря, из всех множеств систем n линейно независимых нулевых векторов $\{\mathbf{E}_a\} \mathbf{E}_a \in T_p$ для каждого $p \in \mathcal{M}$ $(a = 1, \dots, n)$. Проекцией π является естественная проекция, отображающая базис в точке p в саму эту точку. Пусть $\{x^i\}$ — локальные координаты на открытом множестве $\mathcal{M}$, тогда

$$\{z^A\} = \{x^a, E_1{}^j, E_2{}^k, \dots, E_n{}^m\},$$

где $E_a{}^j$ есть j-я компонента вектора $\mathbf{E}_a$ в координатном базисе $\partial/\partial x^i$, являются локальными координатами в $\pi^{-1}(\mathcal{U})$. Общая линейная группа $GL(n, R)$ действует на $L(\mathcal{M})$ следующим образом: если $\{\mathbf{E}_a\}$ — базис в $p \in \mathcal{M}$, то $\mathbf{A} \in GL(n, R)$ отображает $u = \{p, \mathbf{E}_a\}$ в

$$A(u) = \{p, A_{ab}\mathbf{E}_b\}.$$

Если $\mathcal{M}$ наделена метрикой $\mathbf{g}$ с сигнатурой s, можно ввести подпространство расслоения $L(\mathcal{M})$ — *расслоение ортонормированных базисов* $O(\mathcal{M})$, которое состоит из ортонормированных (относительно $\mathbf{g}$) базисов во всех точках $\mathcal{M}$. На $O(\mathcal{M})$ действует подгруппа $O\left(\frac{1}{2}(n+s), \frac{1}{2}(n-s)\right)$ группы $GL(n, R)$. Эта подгруппа состоит из несингулярных вещественных матриц A_{ab}, удовлетворяющих условию

$$A_{ab} G_{bc} A_{dc} = G_{ad},$$

где G_{bc} — матрица вида

$$\operatorname{diag}(\underbrace{+1,\ +1,\ \ldots,\ +1}_{\frac{n+s}{2}\text{ членов}},\ \underbrace{-1,\ -1,\ \ldots,\ -1}_{\frac{n-s}{2}\text{ членов}}).$$

Матрица A_{ab} отображает $(p, \mathbf{E}_a) \in O(\mathscr{M})$ в $(p, A_{ab}\mathbf{E}_b) \in O(\mathscr{M})$. В случае лоренцевой метрики (т. е. $s = n-2$) группу $O(n-1, 1)$ называют n-мерной группой Лоренца.

C^r-сечение расслоения есть C^r-отображение $\Phi: \mathscr{M} \to \mathscr{E}$, такое, что $\pi \circ \Phi$ есть тождественное отображение на $\mathscr{M}$; таким образом, сечение есть установление C^r-соответствия между каждой точкой $p \in \mathscr{M}$ и некоторым элементом $\Phi(p)$ слоя $\pi^{-1}(p)$. Сечение касательного расслоения $T(M)$ есть векторное поле на $\mathscr{M}$; сечение $T^r_s(\mathscr{M})$ есть тензорное поле $T(\mathscr{M})$ типа (r, s) на $\mathscr{M}$; сечение $L(\mathscr{M})$ есть набор n ненулевых векторных полей $\{\mathbf{E}_a\}$, которые линейно независимы в каждой точке; наконец, сечение $O(\mathscr{M})$ есть набор ортонормированных векторных полей на $\mathscr{M}$.

Поскольку нулевые векторы и тензоры задают сечения в $T(\mathscr{M})$ и $T^r_s(\mathscr{M})$, эти расслоенные пространства всегда допускают сечения. Если $\mathscr{M}$ ориентируемо и некомпактно или компактно с характеристикой Эйлера, равной нулю, на нем будут везде существовать ненулевые векторные поля и, следовательно, везде ненулевые сечения $T(\mathscr{M})$. Расслоения $L(\mathscr{M})$ и $O(\mathscr{M})$ могут как допускать, так и не допускать сечений; например, $L(R^n)$ их допускает, а $L(S^2)$ — нет. Если $L(\mathscr{M})$ допускает сечение, $\mathscr{M}$ называется *параллелизуемым*. Герок показал [55], что некомпактное четырехмерное лоренцево многообразие $\mathscr{M}$ допускает спинорную структуру в том и только в том случае, когда оно параллелизуемо.

В терминах расслоенного пространства $L(\mathscr{M})$ можно дать изящное геометрическое описание связности на $\mathscr{M}$. Ее можно рассматривать как правило параллельного переноса векторов вдоль любой кривой $\gamma(t)$ в $\mathscr{M}$. Таким образом, если $\{\mathbf{E}_a\}$ есть базис в точке $p = \gamma(t_0)$, т. е. $\{p, \mathbf{E}_a\}$ есть точка u в $L(\mathscr{M})$, то параллельным переносом $\{\mathbf{E}_a\}$ вдоль $\gamma(t)$ можно однозначно построить базис в любой другой точке кривой $\gamma(t)$, т. е. указать единственную точку $\bar{\gamma}(t)$ в слое $\pi^{-1}(\gamma(t))$. Следовательно, существует в $L(\mathscr{M})$ единственная кривая $\bar{\gamma}(t)$, называемая *лифтом* $\gamma(t)$, для которой

1) $\bar{\gamma}(t_0) = u$;

2) $\pi(\bar{\gamma}(t)) = \gamma(t)$;

3) базис, изображаемый точкой $\bar{\gamma}(t)$, в $\mathscr{M}$ переносится параллельно вдоль кривой $\bar{\gamma}(t)$.

В терминах локальных координат $\{z^A\}$ кривая $\bar{\gamma}(t)$ записывается как $\{x^a(\gamma(t)),\ E_m{}^i(t)\}$; при этом

$$\frac{dE_m{}^i(t)}{dt} + E_m{}^j \Gamma^i{}_{aj} \frac{dx^a(\gamma(t))}{dt} = 0.$$

Рассмотрим касательное пространство $T_u(L(\mathcal{M}))$ к расслоенному пространству $L(\mathcal{M})$ в точке u. Его координатным базисом является $\{\partial/\partial z^A|_u\}$. n-мерное подпространство, натянутое на касательные векторы $\{(\partial/\partial t)_{\bar{\gamma}(t)}|_u\}$ к лифтам всех кривых, проходящих через p, называется *горизонтальным подпространством* H_u пространства $T_u(L(\mathcal{M}))$. Через локальные координаты

$$\left(\frac{\partial}{\partial t}\right)_{\bar{\gamma}} = \frac{dx^a(\gamma(t))}{dt}\frac{\partial}{\partial x^a} + \frac{dE_m{}^i}{dt}\frac{\partial}{\partial E_m{}^i} =$$
$$= \frac{dx^a(\gamma(t))}{dt}\left(\frac{\partial}{\partial x^a} - E_m{}^j \Gamma^i{}_{aj}\frac{\partial}{\partial E_m{}^i}\right),$$

так что координатным базисом в H_u является $\{\partial/\partial x^a - E_m{}^j\Gamma^i{}_{aj}\partial/\partial E_m{}^i\}$. Таким образом, связность в $\mathcal{M}$ определяет горизонтальные подпространства в касательном пространстве каждой точки $L(\mathcal{M})$. Обратно, в $\mathcal{M}$ можно ввести связность, задав для этого в $T_u(L(M))$ при каждой $u \in L(\mathcal{M})$ n-мерное подпространство со свойствами:

1) если $\mathbf{A} \in GL(n, R^1)$, то отображение A_*: $T_u(L(\mathcal{M})) \to T_{A(u)}(L(\mathcal{M}))$ переводит горизонтальное подпространство H_u в $H_{A(u)}$;

2) H_u не содержит ненулевых векторов, принадлежащих вертикальному подпространству V_u.

По определению *вертикальное подпространство* V_u есть n^2-мерное подпространство в $T_u(L(\mathcal{M}))$, натянутое на векторы, касательные к кривым в слое $\pi^{-1}(\pi(u))$; в терминах локальных координат V_u натянуто на векторы $\{\partial/\partial E_m{}^i\}$. Свойство (2) означает, что T_u есть прямая сумма H_u и V_u.

Проектирующее отображение π: $L(\mathcal{M}) \to \mathcal{M}$ индуцирует сюръективное линейное отображение π_*: $T_u(L(\mathcal{M})) \to T_{\pi(u)}(\mathcal{M})$, такое, что $\pi_*(V_u) = 0$, а сужение π_* на H_u есть взаимно однозначное отображение H_u на $T_{\pi(u)}$. Следовательно, π_*^{-1} есть линейное отображение $T_{\pi(u)}(\mathcal{M})$ на H_u. Поэтому для любого вектора $\mathbf{X} \in T_p(\mathcal{M})$ и точки $u \in \pi^{-1}(p)$ существует такой единственный вектор $\overline{\mathbf{X}} \in H_u$, называемый *горизонтальным лифтом* вектора $\mathbf{X}$, для которого $\pi_*(\overline{\mathbf{X}}) = \mathbf{X}$. Пусть даны кривая $\gamma(t)$ в $\mathcal{M}$ и начальная точка u в $\pi^{-1}(\gamma(t_0))$, тогда в $L(\mathcal{M})$ можно построить единственную кривую $\bar{\gamma}(t)$, проходящую через u, касательный вектор которой есть горизонтальный лифт касательного вектора кривой $\gamma(t)$. Таким образом, зная горизонтальные подпространства в каждой точке $L(\mathcal{M})$, можно определить

параллельное перенесение базисов вдоль любой кривой $\gamma(t)$ в $\mathscr{M}$. Затем можно определить ковариантные производные вдоль кривой $\gamma(t)$ для любого тензорного поля $\mathbf{T}$ как обыкновенные производные по t компонент $\mathbf{T}$ относительно параллельно перенесенного базиса.

Если на $\mathscr{M}$ введена метрика, ковариантная производная которой равна нулю, то ортонормированность параллельно переносимых реперов сохраняется. Следовательно, горизонтальные подпространства касательны к $O(\mathscr{M}) \subset L(\mathscr{M})$ и задают связность в $O(\mathscr{M})$.

Аналогично, через параллельный перенос векторов и тензоров связность на $\mathscr{M}$ задает n-мерные горизонтальные подпространства в касательных пространствах к расслоениям $T(\mathscr{M})$ и $T^r_s(\mathscr{M})$. Координатными базисами этих горизонтальных подпространств являются соответственно

$$\left\{\frac{\partial}{\partial x^a} - V^e \Gamma^f{}_{ae} \frac{\partial}{\partial V^f}\right\}$$

и

$$\left\{\frac{\partial}{\partial x^e} - (T^f{}_{\cdots}{}^b{}_{c \ldots}{}_d \Gamma^a{}_{ef} + (\text{и т. д. для каждого нижнего индекса}) - \right.$$

$$- T^{a \cdots}{}^b{}_{f \ldots}{}_d \Gamma^f{}_{ec} - (\text{и т. д. для каждого верхнего индекса}) \left. \times \frac{\partial}{\partial T^{a \cdots}{}^b{}_{c \ldots}{}_d} \right\}.$$

Как и в случае $L(\mathscr{M})$, π_* взаимно однозначно отображает эти горизонтальные подпространства в $T_{\pi(u)}(M)$. Таким образом, снова можно обратить π_* так, чтобы для любого вектора $\mathbf{X} \in T_{\pi(u)}$ получить единственный горизонтальный лифт $\overline{\mathbf{X}} \in T_u$. В частном случае $T(\mathscr{M})$ единственному вектору $\mathbf{W} \in T_{\pi(u)}(\mathscr{M})$ соответствует сама точка u и, следовательно, в $T(\mathscr{M})$ связностью внутренним образом определено горизонтальное векторное поле $\overline{\mathbf{W}}$; в локальных координатах $\{x^a, V^b\}$

$$\overline{W} = V^a \left(\frac{\partial}{\partial x^a} - V^e \Gamma^f{}_{ae} \frac{\partial}{\partial V^f}\right).$$

Это векторное поле можно интерпретировать следующим образом: его интегральная кривая, проходящая через $u = (p, \mathbf{X}) \in \in T(\mathscr{M})$, есть горизонтальный лифт той геодезической в $\mathscr{M}$, касательный вектор которой в p есть $\mathbf{X}$. Значит, векторное поле $\overline{\mathbf{W}}$ представляет все геодезические в $\mathscr{M}$. В частности, семейство всех геодезических, проходящих через $p \in \mathscr{M}$, есть семейство интегральных кривых $\overline{\mathbf{W}}$, проходящих через слой $\pi^{-1}(p) \subset T(\mathscr{M})$; кривые в $\mathscr{M}$ имеют пересечения по крайней мере в p, а кривые в $T(\mathscr{M})$ нигде не пересекаются.

Глава 3

Общая теория относительности

Чтобы понять появление сингулярностей и возможное крушение общей теории относительности, нам важно иметь точную формулировку этой теории и установить, в какой степени она единственна. Поэтому мы представим теорию в виде нескольких постулатов, касающихся математической модели пространства-времени.

В разд. 3.1 мы введем эту математическую модель, а в разд. 3.2 сформулируем первые два постулата — локальной причинности и локального сохранения энергии. Эти постулаты являются общими и для специальной и для общей теорий относительности, и поэтому можно считать, что они экспериментально обоснованы теми опытами, которые проводились для проверки специальной теории относительности. В разд. 3.3 мы выведем из лагранжиана уравнения материальных (т. е. негравитационных) полей и получим из него тензор энергии-импульса.

Третий постулат, определяющий уравнения поля, приведен в разд. 3.4. Экспериментально он обоснован не так хорошо, как первые два, но, как мы увидим позднее, любые иные уравнения приводили бы по крайней мере к одному или даже нескольким нежелательным следствиям или же требовали существования дополнительных полей, которые экспериментально не обнаружены.

3.1. Пространственно-временное многообразие

Математическая модель пространства-времени (т. е. совокупности всех событий), которой мы будем пользоваться, есть пара $(\mathscr{M}, \mathbf{g})$, где $\mathscr{M}$ — связное четырехмерное хаусдорфово C^∞-многообразие, $\mathbf{g}$ — лоренцева метрика (т. е. метрика сигнатуры $+2$) на $\mathscr{M}$.

Две модели $(\mathscr{M}, \mathbf{g})$ и $(\mathscr{M}', \mathbf{g}')$ будем считать эквивалентными, если они *изометричны*, иначе говоря, если существует диффеоморфизм $\theta: \mathscr{M} \to \mathscr{M}'$, который переводит метрику $\mathbf{g}$ в метрику $\mathbf{g}'$, т. е. $\theta_* \mathbf{g} = \mathbf{g}'$. Тогда, строго говоря, моделью пространства-времени является не просто одна пара $(\mathscr{M}, \mathbf{g})$, а весь класс эквивалентности пар $(\mathscr{M}', \mathbf{g}')$, эквивалентных $(\mathscr{M}, \mathbf{g})$. Обычно мы будем иметь дело только с одним представителем $(\mathscr{M}, \mathbf{g})$

данного класса эквивалентности, но тот факт, что эта пара определена лишь с точностью до эквивалентности, важен в некоторых ситуациях, в частности при рассмотрении задачи Коши в гл. 7.

Многообразие $\mathscr{M}$ берется связным, поскольку нам недоступна никакая информация относительно несвязных частей. Мы считаем его хаусдорфовым, поскольку это, по-видимому, согласуется с накопленным опытом. Однако в гл. 5 мы рассмотрим пример, в котором можно поступиться этим условием. Вместе с условием лоренцевости метрики условие хаусдорфовости означает, что $\mathscr{M}$ паракомпактно [55].

Понятие многообразия естественным образом отвечает нашим интуитивным представлениям о непрерывности пространства и времени. К настоящему времени такая непрерывность установлена до расстояний порядка 10^{-15} см экспериментами по рассеянию пионов [49]. Наверное, будет трудно распространить понятие непрерывности до много меньших расстояний, поскольку для этого понадобится частица такой большой энергии, что родится несколько других частиц, и они исказят эксперимент. Возможно поэтому, что модель, в которой пространство-время рассматривается как многообразие, неприемлема для расстояний меньше 10^{-15} см, и нам следовало бы пользоваться такой теорией, в которой в таких масштабах пространство-время обладает какой-то иной структурой. Однако можно ожидать, что подобные отклонения от описания пространства-времени как многообразия не повлияют на общую теорию относительности до тех пор, пока характерный гравитационный масштаб длины не станет такого же порядка. Это могло бы случиться лишь при плотности порядка 10^{58} г/см^3, т. е. в таких предельных состояниях вещества, относительно которых в настоящее время ничего не известно. Тем не менее, рассматривая пространство-время как многообразие и делая определенные разумные предположения, мы покажем в гл. 8—10, что какое-то нарушение общей теории относительности должно иметь место. Им могли бы быть или несостоятельность уравнений поля, или необходимость квантования метрики, или нарушение структуры многообразия.

Метрика g позволяет делить ненулевые векторы в точке $p \in \mathscr{M}$ на три класса; именно, ненулевой вектор $\mathbf{X} \in T_p$ называется *времениподобным*, *пространственноподобным* или *изотропным*, смотря по тому, является ли $g(\mathbf{X}, \mathbf{X})$ отрицательным, положительным или нулевым (ср. рис. 5).

Порядок r дифференцируемости метрики выбирается достаточным для того, чтобы уравнения поля были определены. Они могут быть определены в смысле обобщенных функций, если координатные компоненты метрики g_{ab} и g^{ab} непрерывны, имеют локально квадратично-интегрируемые обобщенные первые

производные по локальным координатам. Набор n функций $f_{;a}$ на R^n называются обобщенными производными функции ψ на R^n, если для любой функции C^∞-функции ψ на R^n с компактным носителем

$$\int f_{;a}\psi \mathrm{d}^n x = -\int f\,(\partial\psi/\partial x^a)\,\mathrm{d}^n x.$$

Однако это условие слишком слабое, поскольку оно не гарантирует ни существования, ни единственности геодезических; для этого требуется метрика класса C^{2-} (C^{2-}-метрика по определению обладает первыми производными координатных компонент, удовлетворяющими локальному условию Липшица; см. разд. 2.1). Фактически в большей части книги мы будем предполагать, что метрика принадлежит по крайней мере классу C^2. Это позволяет определить уравнения поля (а они содержат вторые производные метрики) в каждой точке. В разд. 8.4 мы ослабим требования к метрике до класса C^{2-} и покажем, что это не влияет на наличие сингулярностей.

В гл. 7 мы пользуемся иным условием дифференцируемости для того, чтобы показать, что развитие во времени полевых уравнений определяется подходящими начальными условиями. Мы потребуем, чтобы компоненты метрики и их обобщенные первые производные до порядка m ($m \geqslant 4$) были локально квадратично-интегрируемы. Это условие заведомо выполняется, если метрика принадлежит классу C^4.

Фактически порядок дифференцируемости метрики с точки зрения физики, скорее всего, не имеет значения. Поскольку метрику никогда нельзя измерить точно, а лишь в пределах некоторой ошибки, то никогда нельзя установить, существует ли действительно разрыв в производных того или иного порядка. Таким образом, всегда можно представить результаты измерений как C^∞-метрику.

Если предполагается, что метрика принадлежит классу C^r, то атлас многообразия должен быть порядка C^{r+1}. Можно, однако, всегда найти аналитический податлас в любом C^s-атласе ($s \geqslant 1$) ([174], цит. в [112]). Поэтому, не ограничивая общности, можно с самого начала предположить, что атлас аналитичен, даже в том случае, когда из-за того, что метрика класса C^r, физически можно задать только C^{r+1}-атлас.

Чтобы гарантировать включение в нашу модель $(\mathscr{M}, \mathbf{g})$ всех несингулярных точек пространства-времени, мы должны наложить на нее определенные условия. Будем говорить, что C^r-пара $(\mathscr{M}', \mathbf{g}')$ является *C^r-расширением* $(\mathscr{M}, \mathbf{g})$, если существует изометрическое C^r-вложение μ: $\mathscr{M} \to \mathscr{M}'$. Если бы такое расширение существовало, то нам пришлось бы рассматривать в качестве точек пространства-времени также и точки $\mathscr{M}'$. Поэтому мы

потребуем, чтобы модель $(\mathcal{M}, \mathbf{g})$ была *C^r-нерасширяемой*, т. е. чтобы не существовало ее C^r-расширения $(\mathcal{M}', \mathbf{g}')$, при котором $\mu(\mathcal{M})$ не совпадает с $\mathcal{M}'$.

В качестве примера пары $(\mathcal{M}_1, \mathbf{g}_1)$, которая не является нерасширяемой, рассмотрим двумерное евклидово пространство с осью x, вырезанной между $x_1 = -1$ и $x_1 = +1$. Очевидным способом его расширения является «возвращение на место» вырезанных точек, но можно также расширить его, взяв еще один экземпляр $(\mathcal{M}_2, \mathbf{g}_2)$ такого пространства и отождествив нижний берег разреза при $|x_1| < 1$ с верхним берегом разреза при $|x_2| < 1$, а верхний берег разреза при $|x_1| < 1$ с нижним берегом разреза при $|x_2| < 1$. Полученное таким образом пространство $(\mathcal{M}_3, \mathbf{g}_3)$ нерасширяемо, но не полно, поскольку мы остались без точек $(x_1 = \pm 1, y_1 = 0)$. Мы не можем поместить эти точки обратно, так как мы стеснены условиями продолжения верхней и нижней части оси x на различные листы. Однако, если взять подмножество $\mathcal{U} \subset \mathcal{M}_3$, определенное условиями $1 < x_1 < 2$, $-1 < y_1 < 1$, то можно расширить пару $(\mathcal{U}, \mathbf{g}_3|_{\mathcal{U}})$ и поместить обратно точку $x_1 = 1$, $y_1 = 0$. Это побуждает к более строгому определению нерасширяемости: пара $(\mathcal{M}, \mathbf{g})$ называется *локально C^r-нерасширяемой*, если не существует какого-либо открытого множества $\mathcal{U} \subset \mathcal{M}$ с некомпактным замыканием в $\mathcal{M}$, такого, что пара $(\mathcal{U}, \mathbf{g}|_{\mathcal{U}})$ имеет расширение $(\mathcal{U}', \mathbf{g}')$, в котором замыкание образа $\mathcal{U}$ компактно *).

3.2. Материальные (негравитационные) поля

Мы будем рассматривать на $\mathcal{M}$ различные поля, например такие, как электромагнитное поле, нейтринное поле и т. д., которые описывают «материальное содержимое» пространства-времени. Эти поля будут подчинены уравнениям, которые можно представить в виде таких соотношений между тензорами на $\mathcal{M}$, что все производные по положению будут ковариантными производными относительно симметричной связности, определяемой метрикой $\mathbf{g}$. Этот вывод следует из того, что только тензорные соотношения определяются структурой многообразия, а связность, задаваемая метрикой, — пока единственная связность, которую мы вводили. Если бы была задана другая связность на $\mathcal{M}$, то разность двух связностей была бы тензором, и его можно было бы рассматривать как новое физическое поле. Аналогично, другую метрику на $\mathcal{M}$ можно было бы рассматривать как еще одно физическое поле. Уравнения полей иногда выражаются в виде соотношений между спинорами на $\mathcal{M}$. В этой книге мы не

*) Если $\mathcal{M}$ — одномерное многообразие, т. е. $\mathcal{M} \equiv \mathcal{M}_1$, то пару $(\mathcal{M}, \mathbf{g})$, удовлетворяющую указанному условию, будем называть *локально C^r-непродолжимой* (см. стр. 205). — *Прим. перев.*

будем иметь дела с такого рода соотношениями, поскольку для тех задач, которые мы собираемся рассматривать, в них нет надобности. В действительности все спинорные уравнения можно заменить несколько более сложными тензорными уравнениями (см., например, [143]).

Мы получим ту или иную форму теории в зависимости от того, какие материальные поля включим в нее. Можно ввести в теорию все поля, которые экспериментально наблюдаются, причем можно было бы постулировать и существование полей, которые еще не обнаружены. Так, например, Бранс и Дикке ([39], приложение 7) постулировали существование дальнодействующего скалярного поля, которое связано слабо со следом тензора энергии-импульса. В форме, приданной ей Дикке ([39], приложение 2), теорию Бранса — Дикке можно рассматривать просто как общую теорию относительности с дополнительным скалярным полем. В настоящее время идут споры о том, обнаружено экспериментально такое поле или нет.

Мы будем обозначать материальные поля, включаемые в теорию, через $\Psi_i{}^{a\ldots b}{}_{c\ldots d}$, где индекс (i) нумерует рассматриваемые поля. Следующие два постулата относительно природы тех уравнений, которым подчиняются поля $\Psi_i{}^{a\ldots b}{}_{c\ldots d}$, являются общими для частной и общей теорий относительности.

Постулат (а). Локальная причинность

Пусть $\mathcal{U}$ — выпуклая нормальная окрестность, p, q — любые две точки $\mathcal{U}$; уравнения, описывающие материальные поля, должны быть такими, чтобы передача сигнала от одной из этих точек до другой была возможна тогда и только тогда, когда p и q могут быть соединены C^1-кривой, лежащей целиком в $\mathcal{U}$ и имеющей касательным вектором всюду ненулевой времениподобный или изотропный вектор; такую кривую мы будем называть *непространственноподобной.* (Наша формулировка теории относительности исключает возможность существования частиц вроде тахионов, движущихся по пространственноподобным кривым.) Идет ли сигнал от p к q или от q к p, зависит от направления времени в $\mathcal{U}$. Вопрос о том, можно ли во всех точках пространства-времени непротиворечивым образом задать направление времени, будет рассмотрен в разд. 6.2.

Более точную формулировку этого постулата можно дать через формулировку задачи Коши для материальных полей. Пусть точка $p \in \mathcal{U}$ такова, что каждая непространственноподобная кривая, проходящая через p, пересекает пространственноподобную поверхность $x^4 = 0$ в пределах $\mathcal{U}$. Пусть $\mathcal{F}$ — множество тех точек на поверхности $x^4 = 0$, которых можно достигнуть, двигаясь из p в пределах $\mathcal{U}$ по непространственноподобным кривым. Тогда мы потребуем, чтобы значения материальных полей в точ-

ке p однозначно определялись значениями на $\mathscr{F}$ этих полей и их производных до некоторого конечного порядка и в то же время не определялись однозначно значениями на любом собственном подмножестве множества $\mathscr{F}$, к которому $\mathscr{F}$ может быть непрерывным образом стянуто. (Более полное обсуждение задачи Коши см. в гл. 7.)

Именно этот постулат выделяет метрику $\mathbf{g}$ среди других полей на $\mathscr{M}$ и придает ей четко выраженный геометрический характер. Если $\{x^a\}$ — нормальные координаты в $\mathscr{U}$ вблизи p, интуитивно довольно очевидно (доказательство см. в гл. 4), что точками, которых можно достигнуть из p по непространственноподобным кривым в $\mathscr{U}$, являются точки с координатами, удовлетворяющими условию

$$(x^1)^2 + (x^2)^2 + (x^3)^2 - (x^4)^2 \leqslant 0.$$

Граница этих точек представляет собой образ изотропного конуса в p при экспоненциальном отображении, т. е. множество всех изотропных геодезических, проходящих через p. Таким образом, установив, какие точки могут сообщаться сигналом с p, можно определить изотропный конус N_p в T_p. Коль скоро известно N_p, то с точностью до конформного множителя определяется метрика в p. В этом можно убедиться следующим образом. Пусть $\mathbf{X}, \mathbf{Y} \in T_p$ соответственно времениподобный и пространственноподобный векторы. Уравнение

$$g(\mathbf{X} + \lambda\mathbf{Y}, \mathbf{X} + \lambda\mathbf{Y}) = g(\mathbf{X},\mathbf{X}) + 2\lambda g(\mathbf{X}, \mathbf{Y}) + \lambda^2 g(\mathbf{Y}, Y) = 0$$

имеет два действительных корня λ_1 и λ_2, поскольку $g(\mathbf{X}, \mathbf{X}) < 0$ и $g(\mathbf{Y}, \mathbf{Y}) > 0$. Если мы знаем N_p, то можем определить λ_1 и λ_2; но

$$\lambda_1\lambda_2 = g(\mathbf{X}, \mathbf{X})/g(\mathbf{Y}, \mathbf{Y}).$$

Таким образом, по изотропному конусу можно установить отношение величин времениподобного и пространственноподобного векторов. Тогда, если $\mathbf{W}$ и $\mathbf{Z}$ — любые два неизотропных вектора в p, то

$$g(\mathbf{W}, \mathbf{Z}) = \frac{1}{2}\,(g(\mathbf{W}, \mathbf{W}) + g(\mathbf{Z}, \mathbf{Z}) - g(\mathbf{W} + \mathbf{Z}, \mathbf{W} + \mathbf{Z})).$$

Каждую величину в правой части можно сравнить с «длиной» $\mathbf{X}$ или $\mathbf{Y}$ и таким способом найти $g(\mathbf{W}, \mathbf{Z})/g(\mathbf{X}, \mathbf{X})$. (Если вектор $\mathbf{W} + \mathbf{Z}$ изотропен, то можно использовать соответствующее выражение, содержащее $\mathbf{W} + 2\mathbf{Z}$.) Итак, существование локальной причинности позволяет измерить метрику с точностью до конформного множителя. На практике это измерение удобнее всего осуществить, используя тот экспериментальный факт, что не обнаружено никакого сигнала, который распространялся бы быстрее электромагнитного излучения. Это означает, что свет должен

распространяться по изотропным геодезическим. Однако такое свойство света является следствием конкретных уравнений, которым подчиняется электромагнитное поле, а не самой теории относительности. Дальнейшее рассмотрение причинности мы продолжим в гл. 6. Среди прочих результатов мы получим, что отношения причинности могут быть использованы для определения топологической структуры $\mathscr{M}$. Конформный множитель метрики можно определить, пользуясь постулатом (б), который формулируется ниже; тогда все объекты теории будут физически наблюдаемыми.

Постулат (б): локальное сохранение энергии и импульса

Уравнения, описывающие материальные поля, таковы, что существует симметричный тензор T^{ab}, называемый тензором энергии-импульса, который зависит от полей, их ковариантных производных и метрики; он обладает следующими свойствами:

1) T^{ab} обращается в нуль на открытом множестве $\mathscr{U}$ в том и только в том случае, когда материальные поля исчезают на $\mathscr{U}$.

2) T^{ab} подчиняется уравнению

$$T^{ab}{}_{;b} = 0. \tag{3.1}$$

Условие (1) выражает тот принцип, что все поля обладают энергией. Не исключены возражения против формулировки «только в том случае» на том основании, что могут оказаться два ненулевых поля, тензоры энергии импульса которых компенсируют друг друга. Такая возможность связана с допустимостью существования отрицательной энергии и обсуждается в разд. 3.3.

Если метрика допускает векторное поле Киллинга **K**, то уравнение (3.1) можно проинтегрировать и получить закон сохранения. Чтобы убедиться в этом, введем вектор P^a, компоненты которого равны $P^a = T^{ab}K_b$. Тогда

$$P^a{}_{;a} = T^{ab}{}_{;a}K_b + T^{ab}K_{b;\,a}.$$

Первый член есть нуль по уравнению сохранения (3.1), а второй обращается в нуль, так как тензор T^{ab} симметричен и для вектора Киллинга **K** $2K_{a;\,b} = L_{\mathbf{K}}g_{ab} = 0$. Поэтому, если $\mathscr{D}$ — компактная ориентируемая область с границей $\partial\mathscr{D}$, то по теореме Гаусса (разд. 2.7)

$$\int\limits_{\partial\mathscr{D}} P^b \,\mathrm{d}\sigma_b = \int\limits_{\mathscr{D}} P^b{}_{;\,b}\,\mathrm{d}v = 0. \tag{3.2}$$

Это равенство можно описать словами как равенство нулю полного потока **K**-й компоненты энергии импульса через замкнутую поверхность.

Когда метрика плоская, как это имеет место в частной теории относительности, можно выбрать координаты $\{x^a\}$, в которых

компоненты метрики равны $g_{ab} = e_a \delta_{ab}$ (суммирования нет!), где δ_{ab} — символ Кронекера, $e = +1$, если $a = 1, 2, 3$, и $e_a = -1$, если $a = 4$. Тогда векторами Киллинга будут

$$\underset{\alpha}{\mathbf{L}} = \frac{\partial}{\partial x^\alpha} \qquad (\alpha = 1, 2, 3, 4)$$

(они порождают 4-трансляции) и

$$\underset{\alpha\beta}{\mathbf{M}} = e_\alpha x^\alpha \frac{\partial}{\partial x^\beta} - e_\beta x^\beta \frac{\partial}{\partial x^\alpha} \text{ (суммирования нет; } \alpha, \beta = 1, 2, 3, 4)$$

(порождающими шесть «вращений» в пространстве-времени). Эти изометрии образуют 10-параметрическую группу Ли изометрий плоского пространства-времени, называемую неоднородной группой Лоренца. Их можно использовать для определения десяти векторов $\underset{\alpha}{P^a}$ и $\underset{\alpha\beta}{P^a}$, подчиняющихся уравнению (3.2). Можно считать, что вектор $\underset{4}{\mathbf{P}}$ описывает поток энергии, а $\underset{1}{\mathbf{P}}$, $\underset{2}{\mathbf{P}}$, $\underset{3}{\mathbf{P}}$ — поток трех компонент импульса; $\underset{\alpha\beta}{\mathbf{P}}$ можно интерпретировать как поток момента импульса.

Если метрика не плоская, то в общем случае нет ни одного вектора Киллинга, и подобных законов сохранения не существует. Однако в подходящей окрестности точки q можно ввести нормальные координаты $\{x^a\}$. Тогда в точке q компоненты метрики $g_{ab} = e_a \delta_{ab}$ (суммирования нет!), а компоненты связности $\Gamma^a{}_{bc} = 0$.

Можно взять такую окрестность $\mathscr{D}$ точки q, в которой g_{ab} и $\Gamma^a{}_{bc}$ отличаются от их значений в q на произвольно малую величину; тогда $L_{(a;\,b)}$ и $\underset{\alpha\beta}{M}{}_{(a;\,b)}$ не будут в $\mathscr{D}$ точно равны нулю, но будут отличаться от нуля сколь угодно мало.

Таким образом, интегралы

$$\int\limits_{\partial\mathscr{D}} \underset{\alpha}{P^b}\, \mathrm{d}\sigma_b \qquad \text{и} \qquad \int\limits_{\partial\mathscr{D}} \underset{\alpha\beta}{P^b}\, \mathrm{d}\sigma_b$$

в первом приближении по-прежнему будут равны нулю; иначе говоря, в малой области пространства-времени в этом приближении энергия, импульс и угловой момент импульса все еще сохраняются. Используя это, можно показать, что малое изолированное тело приближенно движется по времениподобной геодезической линии независимо от своей внутренней структуры, если только плотность энергии неотрицательна (расчет движения малого тела в теории относительности см. в [42]). Этот факт можно рассматривать как галилеевский принцип падения всех тел с одинаковой скоростью. В ньютоновских терминах можно было бы сказать, что для всех тел инертная масса (m в $\mathbf{F} = m\mathbf{a}$)

равна пассивной гравитационной массе (массе, на которую действует гравитационное поле). Справедливость этого была установлена с высокой точностью экспериментами Этвеша и Дикке [39].

Постулат (а) позволяет измерить метрику в каждой точке с точностью до конформного множителя. Пользуясь постулатом (б), можно связать между собой эти множители в различных точках, так как уравнения сохранения $T^{ab}{}_{;b} = 0$, вообще говоря, не выполняются для связности, полученной из метрики $\hat{\mathbf{g}} = \Omega^2 \mathbf{g}$. Это можно было бы осуществить, наблюдая пути малых «пробных» частиц и определяя времениподобные геодезические. Пусть $\gamma(t)$ — одна из таких линий и $\mathbf{K} = (\partial/\partial t)_\gamma$ — ее касательный вектор; тогда из (2.29) имеем

$$\frac{\hat{\mathrm{D}}}{\partial t} K^a = \frac{\mathrm{D}}{\partial t} K^a + 2\Omega^{-1}\Omega_{;\,b} K^b K^a - \Omega^{-1}\left(K^b K^c \hat{g}_{bc}\right) \hat{g}^{ad}\Omega_{;d}.$$

Поскольку $\gamma(t)$ — геодезическая относительно пространственно-временной метрики $\mathbf{g}$, то $K^{[b}(D/\partial t)\,K^{a]} = 0$. Следовательно,

$$K^{[b}\,\frac{\hat{\mathrm{D}}}{\partial t} K^{a]} = -\left(K^c K^d \hat{g}_{cd}\right) K^{[b}\hat{g}^{a]\,e}\,(\ln\Omega)_{;e}. \tag{3.3}$$

Зная конформную структуру, мы можем выбрать какую-либо метрику $\hat{\mathbf{g}}$, представительницу данного класса конформно-эквивалентных метрик, и вычислить левую часть в (3.3) для любой пробной частицы. Тогда по правой части $(\ln\Omega)_{;b}$ определяется с точностью до множителя $K^a\hat{g}_{ab}$. Рассмотрев другую кривую $\gamma'(t)$, касательный вектор которой K'^a не параллелен вектору K^a, можно найти $(\ln\Omega)_{;b}$ и таким образом определить Ω везде с точностью до постоянного множителя. Этот постоянный множитель зависит от используемых единиц измерения и может быть выбран произвольно.

Конечно, на практике конформный множитель измеряется не таким способом; обычно используется тот факт, что существует большое число подобных друг другу систем (например, электронных состояний атомов), движение которых вдоль времениподобных геодезических сопровождается рядом событий, обусловленных внутренними движениями. Эти события отмечают положения систем в пространстве-времени. Интервалы между этими событиями, по-видимому, не зависят от предыстории систем в том смысле, что интервалы, измеренные для двух соседних систем, находятся в соответствии друг с другом. Если возможна эффективная изоляция подобных систем от внешних материальных полей (так что они должны будут двигаться по геодезическим) и если предположить, что их внутреннее движение не зависит от кривизны пространства-времени, то мет-

рика — единственное, от чего оно может зависеть. Следовательно, длина дуги между двумя последовательными событиями на кривой должна быть одинаковой для каждой пары последовательных событий на любой такой кривой. Если принять эту длину дуги в качестве единицы измерения, можно определить конформный фактор в любой точке пространства-времени.

В действительности может оказаться, что изоляция системы от внешних материальных полей невозможна. Так, например, в теории Бранса — Дикке присутствует скалярное поле, которое везде отлично от нуля. Однако конформный множитель все же можно определить, если потребовать, чтобы уравнение сохранения $T^{ab}{}_{;b} = 0$ удовлетворялось. Таким образом, зная тензор энергии-импульса T_{ab}, мы можем узнать и конформный множитель.

3.3. Лагранжева формулировка

Условия (1) и (2) постулата (б) ничего не говорят ни о способе построения тензора энергии-импульса для данной системы полей, ни о его единственности. На практике в основном полагаются на интуитивные представления о том, что такое энергия и импульс. Однако в тех случаях, когда уравнения полей могут быть выведены из некоторого лагранжиана, для тензора энергии-импульса существует определенная и однозначная формула.

Пусть L, некоторая скалярная функция полей $\Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d}$ их первых производных и метрики, есть лагранжиан. Уравнения полей получаются из требования, чтобы действие

$$I = \int_{\mathscr{D}} L \, dv$$

было стационарно при вариации полей внутри компактной четырехмерной области $\mathscr{D}$. Под *вариацией поля* $\Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d}$ в $\mathscr{D}$ мы подразумеваем однопараметрическое семейство полей $\boldsymbol{\Psi}_{(i)}(u, r)$, где $u \in (-\varepsilon, \varepsilon)$ и $r \in \mathscr{M}$ такие, что

1) $\boldsymbol{\Psi}_{(i)}(0, r) = \boldsymbol{\Psi}_{(i)}(r)$,

2) $\boldsymbol{\Psi}_{(i)}(u, r) = \boldsymbol{\Psi}_{(i)}(r)$, когда $r \in \mathscr{M} - \mathscr{D}$.

Обозначим $\partial \boldsymbol{\Psi}_{(i)}(u, r)/\partial u\,|_{u=0}$ через $\Delta \boldsymbol{\Psi}_{(i)}$; тогда

$$\left.\frac{\partial I}{\partial u}\right|_{u=0} = \sum_{(i)} \int_{\mathscr{D}} \left(\frac{\partial L}{\partial \Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d}} \Delta \Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d} + \right.$$

$$\left. + \frac{\partial L}{\partial \Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d;e}} \Delta \Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d;e} \right) dv,$$

где $\Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d;e}$ — компоненты ковариантных производных $\boldsymbol{\Psi}_{(i)}$; но $\Delta\left(\Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d;e}\right) = \left(\Delta \Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d}\right)_{;e}$, и потому второй член

можно записать в виде

$$\sum_{(i)} \int_{\mathscr{D}} \left[\left(\frac{\partial L}{\partial \Psi_{(i)}{}^{a \dots b}{}_{c \dots d;e}} \Delta \Psi_{(i)}{}^{a \dots b}{}_{c \dots d} \right)_{;e} - \right.$$

$$\left. - \left(\frac{\partial L}{\partial \Psi_{(i)}{}^{a \dots b}{}_{c \dots d;e}} \right)_{;e} \Delta \Psi_{(i)}{}^{a \dots b}{}_{c \dots d} \right] \mathrm{d}v.$$

Первый член в этом выражении может быть представлен как

$$\int_{\mathscr{D}} Q^a{}_{;a} \mathrm{d}v = \int_{\partial \mathscr{D}} Q^a \, \mathrm{d}\sigma_a,$$

причем $\mathbf{Q}$ — вектор, компоненты которого равны

$$Q^e = \sum_{(i)} \frac{\partial L}{\partial \Psi_{(i)}{}^{a \dots b}{}_{c \dots d;e}} \Delta \Psi_{(i)}{}^{a \dots b}{}_{c \dots d}.$$

Этот интеграл равен нулю, ибо условие (2) эквивалентно утверждению, что $\Delta \Psi_{(i)}$ обращается в нуль на границе $\partial \mathscr{D}$. Таким образом, чтобы $\partial I / \partial u \,|_{u=0} = 0$ при всех вариациях в любой области $\mathscr{D}$, необходимо и достаточно, чтобы при всех i удовлетворялись *уравнения Эйлера — Лагранжа*

$$\frac{\partial L}{\partial \Psi_{(i)}{}^{a \dots b}{}_{c \dots d}} - \left(\frac{\partial L}{\partial \Psi_{(i)}{}^{a \dots b}{}_{c \dots d;e}} \right)_{;e} = 0. \tag{3.4}$$

Мы получили уравнения полей.

Тензор энергии-импульса мы выведем из лагранжиана, рассматривая изменение действия, индуцированное изменением метрики. Предположим, что вариация $g_{ab}(u, r)$ оставляет поля $\Psi_{(i)}{}^{a \dots b}{}_{c \dots d}$ неизменными, но изменяет компоненты метрики g_{ab}. Тогда

$$\left. \frac{\partial I}{\partial u} \right|_{u=0} = \int_{\mathscr{D}} \left(\sum_{(i)} \frac{\partial L}{\partial \Psi_{(i)}{}^{a \dots b}{}_{c \dots d;e}} \Delta \Psi_{(i)}{}^{a \dots b}{}_{c \dots d;e} + \frac{\partial L}{\partial g_{ab}} \Delta g_{ab} \right) \mathrm{d}v +$$

$$+ \int_{\mathscr{D}} L \frac{\partial (\mathrm{d}v)}{\partial g_{ab}} \Delta g_{ab}. \tag{3.5}$$

Последний член возникает из-за того, что мера объема $\mathrm{d}v$ зависит от метрики. Для вычисления этого члена вспомним, что $\mathrm{d}v$ есть каноническая 4-форма η с компонентами $\eta_{abcd} = (-g)^{1/2} 4! \delta_{[a}{}^1 \delta_b{}^2 \delta_c{}^3 \delta_{d]}{}^4$, где $g \equiv \det(g_{ab})$. Отсюда

$$\frac{\partial \eta_{abcd}}{\partial g_{ef}} = -\frac{1}{2}(-g)^{-1/2} \frac{\partial g}{\partial g_{ef}} 4! \delta_{[a}{}^1 \delta_b{}^2 \delta_c{}^3 \delta_{d]}{}^4 =$$

$$= -\frac{1}{2}(-g)^{-1/2} g^{ef} g 4! \delta_{[a}{}^1 \delta_b{}^2 \delta_c{}^3 \delta_{d]}{}^4 =$$

$$= \frac{1}{2} g^{ef} \eta_{abcd},$$

и, следовательно,

$$\frac{\partial(\mathrm{d}v)}{\partial g_{ab}} = \frac{1}{2} g^{ab}\,\mathrm{d}v.$$

Первый член в (3.5) появляется из-за того, что член $\Delta\Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d;e}$ не обязательно будет равен нулю, несмотря на то, что $\Delta\Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d} = 0$: вариация метрики индуцирует вариацию компонент связности $\Gamma^a{}_{bc}$. Поскольку разность двух связностей преобразуется как тензор, $\Delta\Gamma^a{}_{bc}$ можно считать компонентами тензора. Они связаны с вариацией компонент метрики соотношениями

$$\Delta\Gamma^a{}_{bc} = \frac{1}{2} g^{ad}\{(\Delta g_{d\,b})_{;c} + (\Delta g_{dc})_{;b} - (\Delta g_{bc})_{;d}\}.$$

(Вывести эту формулу проще всего, замечая тот факт, что она является тензорным соотношением и, следовательно, должна быть справедлива в любой системе координат. В частности, можно выбрать нормальные координаты с началом в точке p. В этих координатах в p обращаются в нуль компоненты $\Gamma^a{}_{bc}$ и производные компонент g_{ab} по координатам. Тогда можно убедиться, что данная формула справедлива в p.) Используя это соотношение, можно выразить $\Delta\Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d;e}$ через $(\Delta g_{bc})_{;d}$, после чего обычное интегрирование по частям приводит к подынтегральному выражению, содержащему только Δg_{ab}. Итак, мы можем записать $\partial I/\partial u$ в виде

$$\int_{\mathscr{D}} (T^{ab}\,\Delta g_{ab})\,\mathrm{d}v,$$

где T^{ab} — компоненты симметричного тензора, который мы примем за тензор энергии-импульса рассматриваемых полей. (Относительно связи между этим тензором и так называемым каноническим тензором энергии-импульса см. [142].)

Этот тензор энергии-импульса удовлетворяет уравнениям сохранения вследствие полевых уравнений, которым подчинены $\Psi_{(i)}{}^{a\ldots b}{}_{c\ldots d}$. Действительно, предположим, что имеется диффеоморфизм $\phi: \mathscr{M} \to \mathscr{M}$, который отличается от тождественного преобразования только внутри $\mathscr{D}$. Тогда вследствие инвариантности интегралов относительно дифференцируемого отображения имеем

$$I = \int_{\mathscr{D}} L\,\mathrm{d}v = \int_{\mathscr{D}} L\eta = \int_{\phi(\mathscr{D})} L\eta = \int_{\mathscr{D}} \phi^*(L\eta).$$

Следовательно,

$$\int_{\mathscr{D}} (L\eta - \phi^*(L\eta)) = 0.$$

Если диффеоморфизм ϕ порожден векторным полем $\mathbf{X}$ (отличным от нуля только внутри $\mathscr{D}$), то отсюда следует, что

$$\int_{\mathscr{D}} L_{\mathbf{X}}(L\eta) = 0;$$

однако

$$\int_{\mathscr{D}} L_{\mathbf{X}}(L\eta) = \sum_{(i)} \int_{\mathscr{D}} \left(\frac{\partial L}{\partial \Psi_{(i)}{}^{a \cdots b}{}_{c \ldots d}} - \left(\frac{\partial L}{\partial \Psi_{(i)}{}^{a \cdots b}{}_{c \ldots d;e}} \right)_{;e} \right) \times$$
$$\times L_{\mathbf{X}} \Psi_{(i)}{}^{a \cdots b}{}_{c \ldots d} \, dv + \int_{\mathscr{D}} T^{ab} L_{\mathbf{X}} g_{ab} \, dv.$$

Первый член обращается в нуль в силу полевых уравнений. Во втором члене $L_{\mathbf{X}} g_{ab} = 2X_{(a;b)}$. Таким образом,

$$\int_{\mathscr{D}} (T^{ab} L_{\mathbf{X}} g_{ab}) \, dv = 2 \int_{\mathscr{D}} ((T^{ab} X_a)_{;b} - T^{ab}{}_{;b} X_a) \, dv.$$

Вклад первого члена можно представить в виде интеграла по границе $\mathscr{D}$, который равен нулю из-за того, что $\mathbf{X} = 0$ на $\partial\mathscr{D}$. Отсюда следует, что $T^{ab}{}_{;b} = 0$, поскольку второй член должен обращаться в нуль при произвольном $\mathbf{X}$.

Теперь мы приведем в качестве примеров лагранжианы для некоторых полей, которые представляют интерес для последующего.

Пример 1. Скалярное поле ψ

Это поле может соответствовать, например, π^0-мезону. Его лагранжиан является

$$L = -\psi_{;a} \psi_{;b} g^{ab} - \frac{m^2}{\hbar^2} \psi^2,$$

где m, $\hbar$ — константы. Уравнение Эйлера — Лагранжа (3.4) имеет вид

$$\psi_{;ab} g^{ab} - \frac{m^2}{\hbar^2} \psi = 0.$$

Тензор энергии-импульса равен

$$T_{ab} = \psi_{;a} \psi_{;b} - \frac{1}{2} g_{ab} \left(\psi_{;c} \psi_{;a} g^{cd} + \frac{m^2}{\hbar^2} \psi^2 \right). \tag{3.6}$$

Пример 2. Электромагнитное поле

Оно описывается 1-формой $\mathbf{A}$, называемой потенциалом. Потенциал определен с точностью до добавления градиента какой-либо скалярной функции. Лагранжиан имеет вид

$$L = -\frac{1}{8\pi} F_{ab} F_{cd} g^{ac} g^{bd},$$

где тензор электромагнитного поля **F** по определению равен 2d**A**, т. е. $F_{ab} = 2A_{[b;a]}$. Варьируя по A_a, получаем уравнения Эйлера — Лагранжа (3.4) в виде

$$F_{ab;c}g^{bc} = 0.$$

Эти уравнения и уравнение $F_{[ab;c]} = 0$ (т. е. уравнение $\mathbf{dF} = \mathrm{d}(\mathrm{d}\mathbf{A}) = 0$) представляют уравнения Максвелла для электромагнитного поля без источников. Тензор энергии-импульса имеет вид

$$T_{ab} = \frac{1}{4\pi}\left(F_{ac}F_{bd}g^{cd} - \frac{1}{4}g_{ab}F_{ij}F_{kl}g^{ik}g^{jl}\right). \qquad (3.7)$$

Пример 3. Заряженное скалярное поле

Оно является фактически объединением двух вещественных скалярных полей ψ_1 и ψ_2 в одно комплексное скалярное поле $\psi = \psi_1 + i\psi_2$, которое может соответствовать, например, π^+- и π^--мезонам. Полный лагранжиан заряженного скалярного и электромагнитного полей равен

$$L = -(\psi_{;a} + ieA_a\psi)g^{ab}(\bar{\psi}_{;b} - ieA_b\bar{\psi}) - \frac{m^2}{\hbar^2}\psi\bar{\psi} - \frac{1}{8\pi}F_{ab}F_{cd}g^{ac}g^{bd},$$

где e — постоянная и $\bar{\psi}$ — поле, комплексно сопряженное ψ. Варьируя независимо по ψ, $\bar{\psi}$ и A_a, имеем уравнение

$$\psi_{;ab}g^{ab} - \frac{m^2}{\hbar^2}\psi + ieA_ag^{ab}(2\psi_{;b} + ieA_b\psi) + ieA_{a;b}g^{ab}\psi = 0,$$

получаемое отсюда комплексным сопряжением, и уравнение

$$\frac{1}{4\pi}F_{ab;c}g^{bc} - ie\psi(\bar{\psi}_{;a} - ieA_a\bar{\psi}) + ie\bar{\psi}(\psi_{;a} + ieA_a\psi) = 0.$$

Тензор энергии-импульса равен

$$T_{ab} = \frac{1}{2}(\psi_{;a}\bar{\psi}_{;b} + \bar{\psi}_{;a}\psi_{;b}) + \frac{1}{2}(-\psi_{;a}ieA_b\bar{\psi} + \bar{\psi}_{;b}ieA_a\psi +$$

$$+ \bar{\psi}_{;a}ieA_b\psi - \psi_{;b}ieA_a\bar{\psi}) + \frac{1}{4\pi}F_{ac}F_{bd}g^{cd} + e^2A_aA_b\psi\bar{\psi} + \frac{1}{2}Lg_{ab}.$$

Пример 4. Изэнтропическая идеальная жидкость

Здесь подход несколько отличается от предыдущих случаев. Жидкость описывают функцией ρ, называемой плотностью, и конгруэнцией времениподобных кривых, называемых линиями тока. Под конгруэнцией кривых подразумевается такое семейство кривых, что через каждую точку $\mathscr{M}$ проходит одна кривая. Если $\mathscr{D}$ достаточно малая компактная область, то конгруэнцию можно представить в виде диффеоморфизма $\gamma:[a, b]\times\mathscr{N} \to \mathscr{D}$, где $[a, b]$ — некоторый замкнутый интервал из R^1 и $\mathscr{N}$ — некоторое трехмерное многообразие с краем. Кривые называют времениподобными, если их касательные векторы $\mathbf{W} = (\partial/\partial t)_\gamma$, $t \in [a, b]$

всюду времениподобны. Вектор потока жидкости определяется как $\mathbf{j} = \rho\mathbf{V}$, где $\mathbf{V}$ — касательный вектор вида $\mathbf{V} = (-g(\mathbf{W}, \mathbf{W}))^{-1/2}\mathbf{W}$ и, следовательно, $g(\mathbf{V}, \mathbf{V}) = -1$. Требуется, чтобы поток сохранялся, т. е. $j^a{}_{;a} = 0$. Поведение жидкости определяется заданием потенциала упругих сил (или внутренней энергии) ε в функции ρ. Лагранжиан берется в виде

$$L = -2\rho(1 + \varepsilon),$$

и требуется, чтобы действие было стационарным при вариации линий тока, а ρ подбирается так, чтобы обеспечить сохранение потока j^a. Вариация линий тока есть дифференцируемое отображение

$$\gamma\colon (-\delta, \delta) \times [a, b] \times \mathscr{N} \to \mathscr{D},$$

удовлетворяющее условиям

$$\gamma(0, [a, b], \mathscr{N}) = \gamma([a, b]\ \mathscr{N}),$$

и

$$\gamma(u[a, b], \mathscr{N}) = \gamma([a, b], \mathscr{N}) \quad \text{на} \quad \mathscr{M} - \mathscr{D}, \quad (u \in (-\delta, \delta)).$$

Следовательно, $\Delta\mathbf{W} = L_{\mathbf{K}}\mathbf{W}$, где $\mathbf{K} = (\partial/\partial u)_\gamma$. $\mathbf{K}$ можно представить себе как вектор, задающий смещение некоторой точки линии тока при вариации. Отсюда следует, что

$$\Delta V^a = V^a{}_{;b}K^b - K^a{}_{;b}V^b - V^aV^bK_{b;c}V^c.$$

Используя тот факт, что $\Delta(j^a{}_{;a}) = 0 = (\Delta j^a)_{;a}$, имеем

$$(\Delta\rho)_{;a}V^a + \Delta\rho V^a{}_{;a} + \rho_{;a}\Delta V^a + \rho(\Delta V^a)_{;a} = 0.$$

Подставляя сюда ΔV^a и интегрируя вдоль линий тока, получаем

$$\Delta\rho = (\rho K^b)_{;b} + \rho K_{b;c}V^bV^c.$$

Следовательно, вариация интеграла действия равна

$$\frac{\partial I}{\partial u}\Big|_{u=0} = -2\int_{\mathscr{D}}\left\{\left((\rho K^b)_{;b} + \rho K_{b;c}V^bV^c\right)\left(1 + \frac{\mathrm{d}(\rho\varepsilon)}{\mathrm{d}\rho}\right)\right\}\mathrm{d}v.$$

После интегрирования по частям

$$\frac{\partial I}{\partial u}\Big|_{u=0} = 2\int_{\mathscr{D}}\left\{\left(\rho\left(1 + \frac{\mathrm{d}(\varepsilon\rho)}{\mathrm{d}\rho}\right)\dot{V}^a + \rho\left(\frac{\mathrm{d}(\varepsilon\rho)}{\mathrm{d}\rho}\right)_{;c}(g^{ca} + V^cV^a)\right)K_a\right\}\mathrm{d}v,$$

где $\dot{V}^a \equiv V^a{}_{;b}V^b$. Если это выражение равно нулю при всех $\mathbf{K}$, то

$$(\mu + p)\dot{V}^a = -p_{;b}(g^{ba} + V^bV^a),$$

где $\mu = \rho(1+\varepsilon)$ — плотность энергии и $p = \rho^2(d\varepsilon/d\rho)$ — давление. Итак, ортогональный к линиям тока градиент давления задает $\dot{V}^a$, ускорение линий тока.

Чтобы получить тензор энергии-импульса, проварьируем метрику. Вычисления можно упростить, если заметить, что сохранение тока может быть выражено в виде равенства

$$(j^a)_{;a} = \frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^a} (\sqrt{-g} j^a) = 0.$$

Если линии тока заданы, то из уравнений сохранения величина j^a однозначно определяется в каждой точке данной линии тока через свое начальное значение в некоторой фиксированной точке той же линии тока. Поэтому $\sqrt{-g} j^a$ не изменяется при вариации метрики. Но

$$\rho^2 = g^{-1} (\sqrt{-g} j^a \sqrt{-g} j^b) g_{ab},$$

и, следовательно,

$$2\rho \Delta\rho = (j^a j^b - j^c j_c g^{ab}) \Delta g_{ab},$$

а отсюда

$$T^{ab} = \left\{\rho(1+\varepsilon) + \rho^2 \frac{d\varepsilon}{d\rho}\right\} V^a V^b + \rho^2 \frac{d\varepsilon}{d\rho} g^{ab} =$$

$$= (\mu + p) V^a V^b + p g^{ab}. \qquad (3.8)$$

Любую форму материи, тензор энергии-импульса которой имеет вид (3.8) (независимо от того, выведен он из некоторого лагранжиана или нет), будем называть *идеальной жидкостью*. Уравнение сохранения энергии и импульса (3.1) в приложении к (3.8) дают

$$\mu_{;a} V^a + (\mu + p) V^a{}_{;a} = 0, \qquad (3.9)$$

$$(\mu + p) \dot{V}^a + (g^{ab} + V^a V^b) p_{;b} = 0. \qquad (3.10)$$

Эти уравнения совпадают с полученными из лагранжиана. Будем называть идеальную жидкость *изэнтропической*, если давление p является функцией одной лишь плотности энергии μ. В этом случае можно ввести сохраняющуюся плотность ρ и внутреннюю энергию ε и вывести эти уравнения и тензор энергии импульса (3.8) из некоторого лагранжиана.

Можно также рассмотреть заряженную жидкость с сохраняющимся электрическим зарядом e (т. е. $J^a{}_{;a} = 0$, где $\mathbf{J} = e\mathbf{V}$ — электрический ток). Лагранжиан заряженной жидкости и электромагнитного поля имеет вид

$$L = -\frac{1}{8\pi} F_{ab} F_{cd} g^{ac} g^{bd} - 2\rho(1+\varepsilon) - J^a A_a.$$

Последний член задает взаимодействие между жидкостью и полем. Варьируя **A**, линии тока и метрику соответственно, получаем

$$F^{ab}{}_{;b} = 4\pi J^a,$$

$$(\mu + p)\dot{V}^a = -p_{;b}(g^{ab} + V^a V^b) + F^a{}_b J^b,$$

$$T^{ab} = (\mu + p)V^a V^b + pg^{ab} + \frac{1}{4\pi}\left(F^a{}_c F^{bc} - \frac{1}{4} g^{ab} F_{cd} F_{cd}\right).$$

3.4. Уравнения поля

До сих пор мы не конкретизировали метрику **g**. В специальной теории относительности, которая не рассматривает гравитационные эффекты, принимается, что метрика плоская. Однако наблюдения показали, что лучи света, проходящие вблизи Солнца, отклоняются от своего первоначального направления. Поскольку световые лучи представляют собой изотропные геодезические, отсюда следует, что пространственно-временная метрика не может быть плоской или хотя бы конформной плоской метрике. Поэтому нам нужно иметь какой-то рецепт вычисления кривизны пространства-времени. Оказывается, его можно сформулировать так, чтобы в пределе малой, медленно изменяющейся кривизны воспроизводились результаты ньютоновской теории тяготения. Поэтому нет необходимости во введении какого-либо дополнительного поля для описания тяготения. Это не означает, что дополнительного поля, ответственного за часть гравитационных эффектов, вообще не может быть. Такого рода скалярное поле было введено Йорданом [86] и Брансом и Дикке [39]. Причем, как отмечено выше, такое дополнительное поле можно было бы рассматривать просто как еще один вид материального поля и включить в полный тензор энергии-импульса. Поэтому мы придерживаемся того взгляда, что поле тяготения описывается только пространственно-временной метрикой. Тогда возникает проблема нахождения полевых уравнений, связывающих метрику с распределением материи.

Эти уравнения должны быть тензорными уравнениями, учитывающими материю только через ее тензор энергии-импульса, т. е. они не выявляют различий между двумя различными материальными полями, которые имеют одинаковые распределения энергии и импульса. Данное условие можно рассматривать как обобщение ньютоновского принципа равенства активной гравитационной массы тела (массы, создающей поле тяготения) пассивной гравитационной массе (массе, на которую действует поле тяготения). Экспериментально это равенство было проверено Кройцером [93].

Чтобы установить, какими должны быть искомые уравнения поля, рассмотрим предельный ньютоновский случай. Поскольку в ньютоновское уравнение гравитационного поля время не входит, соответствие с ньютоновской теорией нужно устанавливать в статической метрике. Под статической метрикой подразумевается метрика, которая допускает времениподобное векторное поле Киллинга **K**, ортогональное к семейству пространственноподобных поверхностей. Эти поверхности можно рассматривать как поверхности постоянного времени и отмечать их параметром t. Введем единичный времениподобный вектор $\mathbf{V} = f^{-1}\mathbf{K}$, где $f^2 = -K^aK_a$. Тогда $V^a{}_{;a} = -\dot{V}^aV_b$, где $\dot{V}^a = V^a{}_{;b}V^b = f^{-1}f_{;b}g^{ab}$ описывает отклонение от геодезичности интегральных кривых вектора **V** (которые, конечно, одновременно являются интегральными кривыми вектора **K**). Заметим, что $\dot{V}^aV_a = 0$.

Эти интегральные кривые определяют статическую систему отсчета; иными словами, для частицы, история которой представлена одной из этих кривых, пространственно-временная метрика кажется не зависящей от времени. Частица, выведенная из состояния покоя и движущаяся по геодезической, выглядела бы так, как будто она получила начальное ускорение $-\dot{\mathbf{V}}$ относительно этой статической системы отсчета. Если f лишь незначительно отличается от единицы, начальное ускорение выведенной из покоя свободно движущейся частицы приблизительно равно градиенту f, взятому со знаком минус. Это наводит на мысль, что в качестве величины, аналогичной ньютоновскому гравитационному потенциалу, следует рассматривать $f - 1$.

Уравнение для этого потенциала можно вывести, рассматривая дивергенцию $\dot{V}^a$:

$$\dot{V}^a{}_{;a} = (V^a{}_{;b}V^b)_{;a} = V^a{}_{;b;a}V^b + V^a{}_{;b}V^b{}_{;a} =$$
$$= R_{ab}V^aV^b + (V^a{}_{;a})_{;b}V^b + (V_b\dot{V}^b)^2 = R_{ab}V^aV^b.$$

Однако

$$\dot{V}^a{}_{;a} = (f^{-1}f_{;b}g^{ab})_{;a} = -f^{-2}f_{;a}f_{;b}g^{ab} + f^{-1}f_{;ba}g^{ab}$$

и

$$f_{;ab}V^aV^b = -f_{;a}V^a{}_{;b}V^b = -f^{-1}f_{;a}f_{;b}g^{ab},$$

так что имеем

$$f_{;ab}(g^{ab} + V^aV^b) = fR_{ab}V^aV^b.$$

Левая часть есть лапласиан f относительно метрики, индуцированной на 3-поверхности $\{t = \text{const}\}$. Если эта метрика почти плоская, то мы получим ньютоновский лапласиан потенциала. Таким образом, мы получим согласие с теорией Ньютона в предельном случае слабого поля (т. е. когда $f \approx 1$), если правая часть будет равна плотности материи, умноженной на $4\pi G$ плюс

члены, которые малы в предельном случае слабого поля. Это будет иметь место при условии, что существует соотношение вида

$$R_{ab} = K_{ab}, \tag{3.11}$$

где K_{ab} — тензорная функция тензора энергии-импульса и метрики, такая, что $(4\pi G)^{-1}K_{ab}V^aV^b$ равно плотности материи плюс члены, которые малы в ньютоновском приближении. Допустим на время, что имеется соотношение такого вида.

Поскольку R_{ab} удовлетворяет свернутым тождествам Бианки $R_a{}^b{}_{;b} = \frac{1}{2}R_{;a}$, из (3.11) следует

$$K_a{}^b{}_{;b} = \tfrac{1}{2}K_{;b}. \tag{3.12}$$

Отсюда видно, что кажущееся естественным уравнение $K_{ab} = 4\pi G T_{ab}$ не может быть правильным, так как из (3.12) и уравнений сохранения $T_a{}^b{}_{;b} = 0$ мы получили бы, что $T_{;a} = 0$. Для идеальной жидкости это означало бы, что $\mu - 3p = \text{const}$ во всем пространстве-времени — условие, которому в общем случае жидкость удовлетворять не может.

В действительности из тех тождеств, которым удовлетворяет тензор энергии-импульса, законами сохранения являются в общем случае только тождества первого порядка. Отсюда следует, что тензорной функцией K_{ab} тензора энергии-импульса и метрики, для которой тождества (3.12) выполняются при любом тензоре энергии-импульса, может быть только функция

$$K_{ab} = \varkappa\left(T_{ab} - \tfrac{1}{2}Tg_{ab}\right) + \Lambda g_{ab}, \tag{3.13}$$

где $\varkappa$ и Λ — некоторые постоянные. Значения этих постоянных можно определить из ньютоновского приближения. Рассмотрим идеальную жидкость с плотностью энергии μ и давлением p, линии тока которой являются интегральными кривыми вектора Киллинга (т. е. жидкость покоится в статической системе отсчета). Ее тензор энергии-импульса имеет вид (3.8). Подставляя его в (3.13) и (3.11), получим

$$f_{;ab}(g^{ab} + V^aV^b) = f\left(\tfrac{1}{2}\varkappa(\mu + 3p) - \Lambda\right). \tag{3.14}$$

В ньютоновском пределе давление p обычно весьма мало в сравнении с плотностью энергии. (Мы пользуемся единицами, в которых скорость света равна единице. Если единицы таковы, что скорость света равна c, выражение $\mu + 3p$ следует заменить на $\mu + 3p/c^2$.) Следовательно, приближенное согласие с ньютоновской теорией будет достигнуто, если $\varkappa = 8\pi G$, а постоянная $|\Lambda|$ очень мала. Мы будем пользоваться единицами массы, в кото-

рых $G = 1$. В этих единицах масса 10^{28} г соответствует длине 1 см. Из наблюдений удаленных галактик, выполненных Сендейджем [145, 146], вытекают ограничения на $|\Lambda|$ порядка 10^{-56} см$^{-2}$. Обычно мы будем принимать $\Lambda = 0$, но следует помнить о возможности иных значений.

Теперь можно проинтегрировать (3.14) по компактной области $\mathscr{T}$ 3-поверхности $\{t = \text{const}\}$ и преобразовать левую часть в интеграл от градиента f по граничной 2-поверхности $\partial\mathscr{T}$:

$$\int_{\mathscr{T}} f\,(4\pi(\mu + 3p))\,\mathrm{d}\sigma = \int_{\mathscr{T}} f_{;ab}\left(g^{ab} + V^a V^b\right)\mathrm{d}\sigma = \int_{\partial\mathscr{T}} f_{;a}\left(g^{ab} + V^a V^b\right)\mathrm{d}\tau_b,$$

где $\mathrm{d}\sigma$ — элемент объема 3-поверхности $\{t = \text{const}\}$ в индуцированной метрике и $\mathrm{d}\tau_b$ — элемент площади 2-поверхности $\partial\mathscr{T}$ на этой 3-поверхности. Это приводит к аналогу формулы Ньютона для полной массы, заключенной внутри 2-поверхности. Есть, однако, два важных отличия от ньютоновского случая:

1. В интеграле в правой части появляется множитель f; это означает, что материя, находящаяся в области, где f значительно меньше единицы (большой отрицательный ньютоновский потенциал), дает меньший вклад в полную массу, чем материя из области, где f почти равно единице (малый отрицательный ньютоновский потенциал).

2. Давление дает вклад в полную массу; это означает, что в определенных условиях оно может способствовать, а вовсе не препятствовать гравитационному коллапсу.

Уравнения

$$R_{ab} = 8\pi\left(T_{ab} - \tfrac{1}{2}Tg_{ab}\right) + \Lambda g_{ab}$$

называются *уравнениями Эйнштейна*, и часто их записывают в следующей эквивалентной форме:

$$\left(R_{ab} - \tfrac{1}{2}Rg_{ab}\right) + \Lambda g_{ab} = 8\pi T_{ab}. \tag{3.15}$$

Поскольку тензоры в обеих частях симметричны, эти уравнения образуют систему десяти связанных нелинейных дифференциальных уравнений в частных производных для метрики и ее первых и вторых производных. Однако ковариантные дивергенции левой и правой частей тождественно равны нулю, т. е.

$$\left(R^{ab} - \tfrac{1}{2}Rg^{ab} + \Lambda g^{ab}\right)_{;b} = 0$$

и

$$T^{ab}{}_{;b} = 0$$

независимо от уравнений поля. Таким образом, в действительности уравнения поля содержат только шесть независимых дифференциальных уравнений для метрики. Это как раз то число уравнений, которое нужно для задания геометрии пространства-времени: четыре из десяти компонент метрики могут быть приведены к произвольным значениям при использовании четырех степеней свободы в виде координатных преобразований. К этому выводу можно прийти и другим путем: две метрики $\mathbf{g}_1$ и $\mathbf{g}_2$ на многообразии $\mathscr{M}$ задают ту же самую геометрию, если существует диффеоморфизм θ, переводящий $\mathbf{g}_1$ в $\mathbf{g}_2$; поэтому уравнения поля должны определять метрику с точностью до класса эквивалентности относительно диффеоморфизмов, в выборе которых имеются четыре степени свободы.

В гл. 7 мы рассмотрим задачу Коши для уравнений Эйнштейна и покажем, что последних вместе с уравнениями для материальных полей при подходящих начальных данных достаточно для определения эволюции пространства-времени и что они удовлетворяют постулату причинности (а) (стр. 72).

Уравнения Эйнштейна могут быть выведены из требования стационарности действия

$$I = \int_{\mathscr{D}} (A(R - 2\Lambda) + L)\, dv \tag{3.16}$$

при вариации g_{ab}. Здесь L — лагранжиан материи и A — подходящая постоянная. Действительно,

$$\Delta((R - 2\Lambda)\, dv) = \left((R - 2\Lambda)\frac{1}{2} g^{ab}\Delta g_{ab} + R_{ab}\,\Delta g^{ab} + g^{ab}\,\Delta R_{ab}\right) dv;$$

последний член можно переписать в виде

$$g^{ab}\,\Delta R_{ab}\, dv = g^{ab}\left((\Delta\Gamma^{c}{}_{ab})_{;c} - (\Delta\Gamma^{c}{}_{ac})_{;b}\right) dv = \\ = (\Delta\Gamma^{c}{}_{ab}g^{ab} - \Delta\Gamma^{d}{}_{ad}g^{ac})_{;c}\, dv.$$

Отсюда интеграл от этого члена можно преобразовать в интеграл по границе $\partial\mathscr{D}$, который обращается в нуль потому, что $\Delta\Gamma^{a}{}_{bc} = 0$ на $\partial\mathscr{D}$. Поэтому

$$\frac{\partial I}{\partial u}\Big|_{u=0} = \int_{\mathscr{D}} \left\{A\left(\left(\frac{1}{2}R - \Lambda\right) g^{ab} - R^{ab}\right) + T^{ab}\right\} \Delta g_{ab}\, dv, \tag{3.17}$$

и если $\partial I/\partial u$ обращается в нуль при всех Δg_{ab}, мы получаем уравнения Эйнштейна, положив $A = (8\pi)^{-1}$.

Может возникнуть вопрос, нельзя ли, варьируя действие, полученное из какой-либо другой комбинации метрического тензора и тензора кривизны, получить другую разумную систему уравнений. Однако скалярная кривизна R — единственный скаляр, который линеен по вторым производным метрического тен-

зора, так что только в этом случае можно выделить поверхностный интеграл и получить уравнение, содержащее производные метрики не выше второго порядка. Если же исходить из какого-либо другого скаляра, например $R_{ab}R^{ab}$ или $R_{abcd}R^{abcd}$, то получится уравнение, содержащее четвертые производные метрического тензора. Это обстоятельство нежелательно, ибо все прочие уравнения физики первого или второго порядка. Если бы уравнения были четвертого порядка, то для описания эволюции метрики пришлось бы задавать начальные значения не только метрики и ее первых производных, но также значения вторых и третьих производных.

Мы будем предполагать, что уравнения поля не содержат производных метрики выше второго порядка. Если эти уравнения выводятся из какого-либо лагранжиана, то действие обязательно должно иметь вид (3.16). Однако можно получить систему уравнений, отличную от эйнштейновской, если наложить ограничения на вид вариаций Δg_{ab}, при которых мы требуем стационарности действия. Скажем, можно было бы ограничиться метриками, конформными к плоской метрике, т. е. положить

$$g_{ab} = \Omega^2 \eta_{ab},$$

где η_{ab} — плоская метрика специальной теории относительности. Тогда

$$\Delta g_{ab} = 2\Omega^{-1}\,\Delta\Omega g_{ab},$$

и действие будет стационарным при условии, что

$$\left\{\left(A\left(\tfrac{1}{2}R - \Lambda\right)g^{ab} - R^{ab}\right) + T^{ab}\right\}\Delta\Omega\, g_{ab} = 0$$

при всех $\Delta\Omega$, т. е. если

$$R + A^{-1}T = 4\Lambda.$$

Из (2.30)

$$R = -\,6\Omega^{-3}\Omega_{|bc}\eta^{bc} = -\,6\Omega^{-1}\Omega_{;bc}g^{bc} + 12\Omega^{-2}\Omega_{;c}\Omega_{;d}g^{cd},$$

где $|$ означает ковариантное дифференцирование относительно плоской метрики η_{ab}. Для статической метрики $\Omega = \text{const}$ вдоль интегральных кривых вектора Киллинга $\mathbf{K}$ (т. е. Ω не будет зависеть от времени t). Модуль $\mathbf{K}$ будет пропорционален Ω. Поэтому

$$f_{;ab}\left(g^{ab} + V^aV^b\right)f^{-1} = \Omega_{;ab}\left(g^{ab} + V^aV^b\right)\Omega^{-1} =$$

$$= -\frac{1}{6}R + 2\Omega^{-2}\Omega_{;a}\Omega_{;b}g^{ab} - \Omega^{-1}\Omega_{;a}V^a{}_{;b}V^b =$$

$$= -\frac{1}{6}R + f^2 f_{;a}f_{;b}g^{ab}.$$

Таким образом, лапласиан f будет равен $-\frac{1}{6}R$ плюс член, пропорциональный квадрату градиента f. Этим последним членом в слабом поле можно пренебречь. Согласно уравнениям поля

$$-\frac{1}{6}R = \frac{1}{6}A^{-1}T - \frac{2}{3}\Lambda.$$

Для идеальной жидкости $T = -\mu + 3p$. Следовательно, согласие с ньютоновской теорией достигается, если Λ мало или равно нулю и $A^{-1} = -24\pi$.

Эта теория, в которой на метрику наложено условие, чтобы она была конформно-плоской, известна как теория Нордстрема. Ее можно переформулировать как теорию, в которой метрика $\boldsymbol{\eta}$ плоская, а гравитационное взаимодействие представлено добавочным скалярным полем ϕ. Как упоминалось ранее, такого рода теория не согласуется с наблюдаемым отклонением луча света массивными объектами и не может объяснить измеренное смещение перигелия Меркурия.

Наблюдаемые значения этих величин (отклонения луча и смещения перигелия Меркурия) можно было бы получить при следующем ограничении на вид метрики:

$$g_{ab} = \Omega^2(\eta_{ab} + W_a W_b),$$

где W_a — произвольное поле 1-формы. В этом случае ньютоновский предел имел бы место в статической метрике, в которой W_a параллельно времениподобному вектору Киллинга. Однако могла бы существовать и другая статическая метрика, в которой W_a не параллельно вектору Киллинга, и она не приводила бы к ньютоновскому пределу. Кроме того, это ограничение выглядит довольно искусственным. Представляется более естественным не накладывать иных ограничений на метрику, кроме требования, чтобы она была лоренцевой.

Поэтому в качестве нашего третьего постулата мы примем:

Постулат (в): Уравнения поля

На многообразии $\mathscr{M}$ удовлетворяются уравнения поля Эйнштейна (3.15).

Предсказания этих уравнений согласуются в пределах ошибок эксперимента с выполненными до настоящего времени наблюдениями отклонения луча света и сдвига перигелия Меркурия. В то же время пока остается открытым вопрос, существует ли дальнодействующее скалярное поле, которое нужно было бы включить в тензор энергии-импульса.

Глава 4

Физический смысл кривизны

В этой главе мы рассмотрим, как сказывается кривизна пространства-времени на семействах времениподобных и изотропных геодезических. Эти семейства могут изображать соответственно линии тока жидкости и(или) мировые линии фотонов. В разд. 4.1 и 4.2 мы выведем формулы для скорости вращения, поперечного сдвига (sheer) и расхождения (expansion) таких семейств кривых; уравнение для скорости расхождения (уравнение Райчаудхури) занимает центральное место в доказательствах теорем о сингулярностях в гл. 8. В разд. 4.3 обсуждаются общие неравенства для тензора энергии-импульса, из которых вытекает, что гравитационное воздействие материи всегда стремится вызвать сближение времениподобных и изотропных геодезических. Из результатов разд. 4.4 видно, что следствием этих энергетических условий будет появление сопряженных или фокальных точек в невращающихся семействах времениподобных или изотропных кривых общего вида. В разд. 4.5 показано, что из наличия сопряженных точек следует существование таких вариаций кривых между двумя точками, которые переводят изотропную геодезическую во времениподобную кривую или времениподобную геодезическую во времениподобную кривую большей длины.

4.1. Времениподобные кривые

Мы видели в гл. 3, что при статической метрике существует связь между длиной времениподобного вектора Киллинга и ньютоновским потенциалом. Находится ли тело в гравитационном поле, можно узнать по тому, ускоряется ли оно или нет, если его вывести из состояния покоя относительно статической системы отсчета, задаваемой этим вектором Киллинга. Однако в общем случае в пространстве-времени может не быть ни одного вектора Киллинга, и тогда не будет и какой-либо выделенной системы отсчета для измерения ускорения. Все, что еще можно сделать,— это взять два близких друг к другу тела и измерить их относительное ускорение. Такое измерение позволяет определить градиент гравитационного поля. Если рассматривать метрику как аналог ньютоновского потенциала, то градиенту ньютоновского поля будут соответствовать вторые производные метрики. Они

определяются тензором Римана. Поэтому можно ожидать, что относительное ускорение двух соседних тел будет связано с какими-то компонентами тензора Римана.

Чтобы установить эту связь более строго, изучим поведение конгруэнции времениподобных кривых с единичным касательным вектором $\mathbf{V}$ ($g(\mathbf{V}, \mathbf{V}) = -1$). Эти кривые могут представлять линии тока жидкости или (в том случае, когда они являются геодезическими) истории малых пробных частиц. В случае идеальной жидкости, согласно (3.10),

$$(\mu + p)\dot{V}^a = -p_{;b}h^{ab}, \tag{4.1}$$

где $\dot{V}^a = V^a{}_{;b}V_b$ — ускорение линий тока, а $h^a{}_b = \delta^a{}_b + V^aV_b$ — тензор, который проектирует вектор $\mathbf{X} \in T_q$ в подпространство H_q пространства T_q, ортогональное к V. Можно рассматривать h_{ab} и как метрику в H_q (ср. с разд. 2.7).

Пусть $\lambda(t)$ — некоторая кривая с касательным вектором $\mathbf{Z} = (\partial/\partial t)_\lambda$. Тогда можно построить семейство кривых $\lambda(t, s)$, смещая каждую точку кривой $\lambda(t)$ на расстояние s вдоль интегральных кривых $\mathbf{V}$. Если определить теперь $\mathbf{Z}$ как $(\partial/\partial t)_{\lambda(t, s)}$, то из определения производной Ли (см. разд. 2.4) следует, что $L_{\mathbf{V}}\mathbf{Z} = 0$, или, иначе говоря,

$$\frac{\mathrm{D}}{\partial s}Z^a = V^a{}_{;b}Z^b. \tag{4.2}$$

Величину $\mathbf{Z}$ можно интерпретировать как вектор, представляющий относительное расположение указанных выше двух точек на двух соседних кривых, которые находятся на равном расстоянии от некоторых произвольных начальных точек. Если к $\mathbf{Z}$ добавить вектор, кратный $\mathbf{V}$, то полученный вектор будет изображать относительное расположение точек тех же двух кривых, но находящихся на различных расстояниях. В действительности нас интересуют лишь девиация соседних кривых (separation of neighbouring curves), а не расположение отдельных точек на них. Поэтому нам приходится иметь дело с $\mathbf{Z}$ с точностью до компоненты, параллельной $\mathbf{V}$, т. е. только с проекцией $\mathbf{Z}$ в каждой точке q на пространство Q_q, которое состоит из классов эквивалентности векторов, отличающихся только добавлением вектора, кратного $\mathbf{V}$. Это пространство можно представлять себе как подпространство H_q пространства T_q, состоящее из векторов, ортогональных к $\mathbf{V}$. Проекцию $\mathbf{Z}$ на H_q будем обозначать через ${}_\perp Z^a = h^a{}_bZ^b$. В случае жидкости можно рассматривать ${}_\perp\mathbf{Z}$ как расстояние между двумя соседними частицами жидкости, измеренное в их системе покоя.

Из (4.2) следует, что

$${}_\perp\left(\frac{\mathrm{D}}{\partial s}\right)({}_\perp Z^a) = V^a{}_{;b}\,{}_\perp Z^b. \tag{4.3}$$

Эта формула дает скорость изменения девиации двух бесконечно близких кривых, измеренную в H_q. Применяя к (4.3) операцию $\mathrm{D}/\partial s$ еще раз и проектируя на H_q, получаем

$$h^a{}_b \frac{\mathrm{D}}{\partial s}\left(h^b{}_c \frac{\mathrm{D}}{\partial s} \perp Z^c\right) = h^a{}_b\left(V^b{}_{;cd} \perp Z^c V^d + V^b{}_{;c} V^c{}_{;d} V_e Z^e V^d + \right.$$
$$\left. + V^b{}_{;c} V^c V^e{}_{;d} Z_e V^d + V^b{}_{;c} h^c{}_e Z^e{}_{;d} V^d\right).$$

Изменяя порядок дифференцирования в первом члене и используя (4.2), приводим это равенство к виду

$$h^a{}_b\left(\frac{\mathrm{D}}{\partial s}\right)\left(h^b{}_c \frac{\mathrm{D}}{\partial s} \perp Z^c\right) = -R^a{}_{bcd} \perp Z^c V^b V^d +$$
$$+ h^a{}_b \dot{V}^b{}_{;c} \perp Z^c + \dot{V}^a \dot{V}_b \perp Z^b. \quad (4.4)$$

Это уравнение, известное как уравнение девиации или уравнение Якоби, дает относительное ускорение, т. е. вторую производную по времени девиации двух бесконечно близких кривых, измеренную в H_q. Мы видим, что если эти кривые — геодезические, то девиация зависит лишь от тензора Римана.

В ньютоновской теории ускорение каждой частицы определяется градиентом потенциала Φ, и поэтому относительное ускорение двух частиц на расстоянии Z^a равно $\Phi_{;ab} Z^b$. Таким образом, член с тензором Римана $R_{abcd} V^b V^d$ аналогичен ньютоновскому $\Phi_{;ac}$. Действие этой «приливной силы» хорошо демонстрирует пример сферы, составленной из свободно падающих на Землю частиц. Каждая частица движется по прямой, проходящей через центр Земли, но более близкие к Земле частицы падают быстрее тех, которые находятся дальше. Это означает, что сфера не сохраняет свою форму, а деформируется в эллипсоид того же объема.

Чтобы исследовать уравнение девиации, мы введем в произвольной точке q интегральной кривой $\gamma(s)$ вектора $\mathbf{V}$ дуальные ортонормированные базисы $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$, $\mathbf{E}_4$ и $\mathbf{E}^1$, $\mathbf{E}^2$, $\mathbf{E}^3$, $\mathbf{E}^4$ пространств T_q и T^*_q, причем возьмем $\mathbf{E}^4 = \mathbf{V}$. Было бы желательно перенести затем эти базисы вдоль кривой $\gamma(s)$, чтобы получить такие же базисы в каждой точке $\gamma(s)$. Однако при параллельном переносе базисов вдоль $\gamma(s)$ (т. е. когда $\mathrm{D}/\partial s$ для каждого вектора обращается в нуль), $\mathbf{E}_4$ не будет оставаться равным $\mathbf{V}$, а $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ не останутся ортогональными к $\mathbf{V}$, если $\gamma(s)$ — не геодезическая. Поэтому мы введем новую производную вдоль $\gamma(s)$, так называемую *производную Ферми* $\mathrm{D_F}/\partial s$. Для векторного поля $\mathbf{X}$ вдоль $\gamma(s)$ она определяется формулой

$$\frac{\mathrm{D_F}(\mathbf{X})}{\partial s} = \frac{\mathrm{D}\mathbf{X}}{\partial s} - g\left(\mathbf{X}, \frac{\mathrm{D}\mathbf{V}}{\partial s}\right)\mathbf{V} + g(\mathbf{X}, \mathbf{V})\frac{\mathrm{D}\mathbf{V}}{\partial s}$$

и обладает следующими свойствами:

1) $\frac{D_F}{\partial s} = \frac{D}{\partial s}$, если $\gamma(s)$ — геодезическая;

2) $\frac{D_F \mathbf{V}}{\partial s} = 0$;

3) если $\mathbf{X}$ и $\mathbf{Y}$ — векторные поля вдоль $\gamma(s)$, такие, что

$$\frac{D_F \mathbf{X}}{\partial s} = 0 = \frac{D_F \mathbf{Y}}{\partial s},$$

то $g(\mathbf{X}, \mathbf{Y}) = \text{const}$ вдоль $\gamma(s)$;

4) если $\mathbf{X}^e$ — векторное поле вдоль $\gamma(s)$, ортогональное к $\mathbf{V}$, то

$$\frac{D_F \mathbf{X}}{\partial s} = {}_\perp\left(\frac{D\mathbf{X}}{\partial s}\right).$$

(Последнее свойство показывает, что производная Ферми является естественным обобщением производной $D/\partial s$.)

Итак, если перенести ортонормированный базис пространства T_q вдоль $\gamma(s)$ так, чтобы производная Ферми каждого базисного вектора была равна нулю, получится ортонормированный базис в каждой точке $\gamma(s)$ с $\mathbf{E}_4 = \mathbf{V}$. Векторы $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ можно интерпретировать как орты невращающейся системы координатных осей вдоль $\gamma(s)$. Физически их можно реализовать с помощью маленьких гироскопов, оси которых направлены вдоль каждого из векторов.

Определение производной Ферми для векторных полей вдоль $\gamma(s)$ можно распространить на произвольные тензорные поля обычным образом, а именно:

I. $D_F/\partial s$ представляет собой линейное отображение тензорных полей типа (r, s) вдоль $\gamma(s)$ на тензорные поля типа (r, s), которое коммутирует со свертками;

II. $\frac{D_F}{\partial s}(\mathbf{K} \otimes \mathbf{L}) = \frac{D_F \mathbf{K}}{\partial s} \otimes \mathbf{L} + \mathbf{K} \otimes \frac{D_F \mathbf{L}}{\partial s}$;

III. $\frac{D_F f}{\partial s} = \frac{df}{ds}$,

где f — произвольная функция.

Из этих правил следует, что над дуальным базисом $\mathbf{E}^1$, $\mathbf{E}^2$, $\mathbf{E}^3$, $\mathbf{E}^4$ пространства T_q^* также совершается перенос Ферми вдоль $\gamma(s)$. Используя производные Ферми, можно переписать (4.3) и (4.4) в виде

$$\frac{D_F}{\partial s} {}_\perp Z^a = V^a{}_{;b} {}_\perp Z^b, \tag{4.5}$$

$$\frac{D_F^2}{\partial s^2} {}_\perp Z^a = - R^a{}_{bcd} {}_\perp Z^c V^b V^d + h^a{}_b \dot{V}^b{}_{;c} {}_\perp Z^c + \dot{V}^a \dot{V}_b {}_\perp Z^b. \tag{4.6}$$

Эти соотношения можно выразить через дуальные базисы, полученные переносом Ферми. Поскольку $_{\perp}\mathbf{Z}$ ортогонален к $\mathbf{V}$, у него есть компоненты лишь относительно $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$. Следовательно, его можно представить как $Z^{\alpha}\mathbf{E}_{\alpha}$, если мы условимся, что греческие индексы пробегают только значения 1, 2, 3. Тогда (4.5) и (4.6) можно записать через обыкновенные производные:

$$\frac{\mathrm{d}}{\mathrm{d}s} Z^{a} = V^{a}{}_{;\beta} Z^{\beta}, \tag{4.7}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} Z^{a} = \left(- R^{a}{}_{4\beta 4} + \dot{V}^{a}{}_{;\beta} + \dot{V}^{a}\dot{V}_{\beta}\right) Z^{\beta}, \tag{4.8}$$

где $V^{\alpha}{}_{;\beta}$ — те компоненты $V^{a}{}_{;b}$, для которых $a = \alpha$ и $b = \beta$. Поскольку компоненты Z^{α} удовлетворяют линейному обыкновенному дифференциальному уравнению 1-го порядка (4.7), эти компоненты можно выразить через их значения в некоторой точке:

$$Z^{\alpha}(s) = A_{\alpha\beta}(s)\, Z^{\beta}\big|_{q}, \tag{4.9}$$

где $A_{\alpha\beta}(s)$ — матрица 3×3, которая в точке q равна единичной матрице и удовлетворяет уравнению

$$\frac{\mathrm{d}}{\mathrm{d}s} A_{\alpha\beta}(s) = V_{\alpha;\gamma} A_{\gamma\beta}(s). \tag{4.10}$$

В случае жидкости можно считать, что матрица $A_{\alpha\beta}$ задает форму и ориентацию малого элемента жидкости, сферического в точке q. Эту матрицу можно записать в следующем виде:

$$A_{\alpha\beta} = O_{\alpha\beta} S_{\alpha\beta}, \tag{4.11}$$

где $O_{\alpha\beta}$ — ортогональная матрица с положительным определителем, а $S_{\alpha\beta}$ — симметричная матрица. $O_{\alpha\beta}$ и $S_{\alpha\beta}$ могут быть выбраны в точке q равными единичной матрице, причем $O_{\alpha\beta}$ можно рассматривать как матрицу, характеризующую вращение соседних кривых относительно базиса, полученного переносом Ферми, а $S_{\alpha\beta}$ — как матрицу, описывающую смещение этих кривых относительно $\gamma(s)$. При этом определитель $S_{\alpha\beta}$, который равен определителю $A_{\alpha\beta}$, характеризует трехмерный объем вырезаемого соседними кривыми элемента поверхности, ортогональной к $\gamma(s)$.

В точке q, где $A_{\alpha\beta}$ — единичная матрица, производная $\mathrm{d}O_{\alpha\beta}/ds$ антисимметрична, а $\mathrm{d}S_{\alpha\beta}/\mathrm{d}s$ симметрична. Таким образом, антисимметричная часть $V_{\alpha;\beta}$ дает скорость вращения соседних кривых в точке q, симметричная часть дает скорость изменения девиации от $\gamma(s)$, а скорость изменения объема определяется следом $V_{\alpha;\beta}$. Поэтому мы введем тензор вращения

$$\omega_{ab} = h_a{}^{c} h_b{}^{d} V_{[c;d]}, \tag{4.12}$$

тензор расхождения

$$\theta_{ab} = h_a{}^c h_b{}^d V_{(c;d)} \tag{4.13}$$

и объемное расхождение

$$\theta = \theta_{ab} h^{ab} = V_{a;b} h^{ab} = V^a{}_{;a}. \tag{4.14}$$

Далее мы определим тензор поперечного сдвига как часть θ_{ab} с нулевым следом:

$$\sigma_{ab} = \theta_{ab} - \frac{1}{3} h_{ab}\theta \tag{4.15}$$

и вектор вращения как

$$\omega^a = \frac{1}{2} \eta^{abcd} V_b \omega_{cd} = \frac{1}{2} \eta^{abcd} V_b V_{c;d}. \tag{4.16}$$

Ковариантную производную вектора **V** можно выразить через эти величины:

$$V_{a;b} = \omega_{ab} + \sigma_{ab} + \frac{1}{3}\theta h_{ab} - \dot{V}_a V_b. \tag{4.17}$$

Это разложение градиента вектора скорости жидкости аналогично соответствующему разложению в ньютоновской гидродинамике.

В ортонормированном базисе, полученном переносом Ферми, вращение и расхождение можно записать с помощью матрицы $A_{\alpha,\beta}$ и обратной ей матрицы $A^{-1}{}_{\alpha\beta}$:

$$\omega_{\alpha\beta} = -A^{-1}{}_{\gamma[\alpha} \frac{d}{ds} A_{\beta]\gamma}, \tag{4.18}$$

$$\theta_{\alpha\beta} = A^{-1}{}_{\gamma(\alpha} \frac{d}{ds} A_{\beta)\gamma}, \tag{4.19}$$

$$\theta = (\det \mathbf{A})^{-1} \frac{d}{ds} (\det \mathbf{A}). \tag{4.20}$$

Из уравнения девиации (4.8) следует, что

$$\frac{d^2}{ds^2} A_{\alpha\beta} = (-R_{\alpha 4 \gamma 4} + \dot{V}_{\alpha;\gamma} + \dot{V}_\alpha \dot{V}_\gamma) A_{\gamma\beta}. \tag{4.21}$$

Если известен тензор Римана, то это уравнение позволяет вычислить распространение вращения, сдвига и расхождения вдоль интегральных кривых **V**.

Умножив (4.21) на $A^{-1}{}_{\beta\gamma}$ и взяв антисимметричную часть, получим

$$\frac{d}{ds} \omega_{\alpha\beta} = 2\omega_{\gamma[\alpha}\theta_{\beta]\gamma} + \dot{V}_{[\alpha;\beta]}. \tag{4.22}$$

Таким образом, распространение вращения зависит от антисимметризованного градиента ускорения, но не от «приливной силы».

Запишем это уравнение в другой форме:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(A_{\gamma\alpha}\omega_{\gamma\delta}A_{\delta\beta}\right) = A_{\gamma\alpha}\dot{V}_{[\gamma;\delta]}A_{\delta\beta}. \tag{4.23}$$

Отсюда видно, что $A_{\gamma\alpha}\omega_{\gamma\delta}A_{\delta\beta}$ — постоянная матрица, если рассматриваемые кривые — геодезические; в частности, если, вращение обращается в нуль в одной точке какой-либо кривой, то в случае геодезических оно равно нулю во всех точках этой кривой. Если эти кривые представляют линии тока идеальной жидкости, то из (4.1) следует, что

$$\dot{V}_{[\alpha;\beta]} = -\frac{1}{\mu + p}\,\omega_{\alpha\beta}\frac{\mathrm{d}p}{\mathrm{d}s}.$$

Если жидкость изэнтропическая, то отсюда следует закон сохранения:

$$WA_{\gamma\alpha}\omega_{\gamma\delta}A_{\delta\beta} = \mathrm{const}, \tag{4.24}$$

где

$$\ln W = \int \frac{\mathrm{d}p}{\mu + p}.$$

Этот закон сохранения представляет собой релятивистскую форму ньютоновского закона сохранения момента количества движения. В геодезическом случае или при отсутствии давления он сводится к обычному утверждению, что величина вектора момента количества движения обратно пропорциональна площади сечения элемента жидкости, ортогонального к этому вектору. Когда давление не равно нулю, появляются дополнительный релятивистский эффект, возникающий из-за того, что давление совершает работу над жидкостью, вследствие чего возрастает масса, а следовательно, и инерция элемента жидкости [ср. с (3.9)]. Это означает, что вращение жидкости под давлением возрастает медленнее, чем можно было бы ожидать без учета давления.

Умножив (4.21) на $A^{-1}{}_{\beta\gamma}$ и взяв симметричную часть, находим

$$\frac{\mathrm{d}}{\mathrm{d}s}\theta_{\alpha\beta} = -R_{\alpha 4\beta 4} - \omega_{\alpha\gamma}\omega_{\gamma\beta} - \theta_{\alpha\gamma}\theta_{\gamma\beta} + \dot{V}_{(\alpha;\beta)} + \dot{V}_{\alpha}\dot{V}_{\beta}. \tag{4.25}$$

Уравнения (4.25) и (4.23) можно записать в общем базисе (не ортонормированном и не обязательно полученном путем переноса Ферми) при помощи замены обыкновенных производных на производные Ферми и проектированием на подпространство, ортогональное к **V**.

След уравнения (4.25) равен

$$\frac{\mathrm{d}}{\mathrm{d}s}\theta = -R_{ab}V^{a}V^{b} + 2\omega^{2} - 2\sigma^{2} - \frac{1}{3}\theta^{2} + \dot{V}^{a}{}_{;a}, \tag{4.26}$$

где

$$2\omega^2 = \omega_{ab}\omega^{ab} \geqslant 0,$$
$$2\sigma^2 = \sigma_{ab}\sigma^{ab} \geqslant 0.$$

Это уравнение, впервые полученное Ландау и независимо Райчаудхури, в дальнейшем будет играть весьма важную роль. Из него видно, что вращение порождает расхождение, как и следовало ожидать из аналогии с центробежной силой, в то время как сдвиг приводит к схождению. Согласно уравнениям поля для идеальной жидкости, касательным вектором линий тока которой является V^a, имеем

$$R_{ab}V^aV^b = 4\pi(\mu + 3p).$$

Поэтому можно ожидать, что этот член также приводит к схождению. Общее исследование знака этого члена будет проведено в разд. 4.3.

Вычитая из (4.25) его след, получим

$$\frac{D_F}{\partial s}\sigma_{ab} = -C_{acbd}V^cV^d + \frac{1}{2}h_a{}^ch_b{}^dR_{cd} - \omega_{ac}\omega^c{}_b - \sigma_{ac}\sigma^c{}_b -$$
$$- \frac{2}{3}\theta\sigma_{ab} + h_a{}^ch_b{}^d\dot{V}_{(c;d)} - \frac{1}{3}h_{ab}\left(2\omega^2 - 2\sigma^2 + \dot{V}^a{}_{;a} + \frac{1}{2}R_{cd}h^{cd}\right), \quad (4.27)$$

где C_{abcd} — тензор Вейля. Поскольку этот тензор имеет нулевой след, он не входит непосредственно в уравнение расхождения (4.26), но из-за присутствия члена $-2\sigma^2$ в правой части этого уравнения тензор Вейля, индуцируя сдвиг, косвенно вызывает схождение кривых. Тензор Римана выражается через тензор Вейля и тензор Риччи:

$$R_{abcd} = C_{abcd} - g_{a[d}R_{c]b} - g_{b[c}R_{d]a} - \frac{1}{3}Rg_{a[c}g_{d]b}.$$

Тензор Риччи определяется уравнениями Эйнштейна

$$R_{ab} - \frac{1}{2}g_{ab}R + \Lambda g_{ab} = 8\pi T_{ab}.$$

Следовательно, тензор Вейля представляет собой ту часть кривизны, которая локально не определяется распределением материи. Однако он не может быть совершенно произвольным, поскольку тензор Римана должен удовлетворять тождествам Бианки:

$$R_{ab[cd;e]} = 0.$$

Последние можно переписать в виде

$$C^{abcd}{}_{;d} = J^{abc}, \quad (4.28)$$

где

$$J^{abc} = R^{c[a;b]} + \frac{1}{6}g^{c[b}R^{;a]}. \quad (4.29)$$

Это соотношение похоже на уравнения Максвелла в электродинамике:

$$F^{ab}{}_{;b} = J^a,$$

где F^{ab} — тензор электромагнитного поля, а J^a — ток источников. Таким образом, тождества Бианки (4.28) в определенном смысле можно считать уравнениями поля для тензора Вейля, являющегося той частью кривизны в данной точке, которая зависит от распределения материи в других точках. (Этот подход был использован для анализа поведения гравитационного излучения в работах [71, 113, 116].)

4.2. Изотропные кривые

Как и в случае времениподобных кривых, тензор Римана будет влиять на скорость изменения относительного расположения изотропных кривых. Для простоты мы рассмотрим только изотропные геодезические. Они могут отображать мировые линии фотонов; влияние тензора Римана будет состоять в искривлении или фокусировке малых пучков световых лучей.

Чтобы исследовать это явление, рассмотрим уравнение девиации для конгруэнции изотропных геодезических с касательным вектором $\mathbf{K}$ ($g(\mathbf{K}, \mathbf{K}) = 0$). Здесь имеются два важных отличия от случая времениподобных кривых, рассмотренного в предыдущем разделе.

Во-первых, там мы располагали возможностью нормировать касательные векторы $\mathbf{V}$ к времениподобным кривым при помощи условия $g(\mathbf{V}, \mathbf{V}) = -1$. В сущности это означает, что кривые были параметризованы длиной дуги s. Ясно, что это невозможно для изотропных кривых, так как дуги их имеют нулевую длину. Самое большее, что можно сделать, — это воспользоваться аффинным параметром v; тогда касательный вектор $\mathbf{K}$ подчиняется уравнению

$$\frac{\mathrm{D}}{\mathrm{d}v} K^a = K^a{}_{;b} K^b = 0.$$

Однако v можно умножить на функцию f, постоянную вдоль каждой кривой. Тогда fv будет другим аффинным параметром, и соответствующим касательным вектором будет $f^{-1}\mathbf{K}$. Таким образом, если кривые заданы как точечные множества на многообразии, то касательный вектор будет однозначен лишь с точностью до множителя, постоянного вдоль каждой кривой.

Во-вторых, еще одно различие состоит в том, что Q_q, фактор-пространство T_q по $\mathbf{K}$, теперь не изоморфно H_q, т. е. подпространству T_q, ортогональному к $\mathbf{K}$: из-за того, что $g(\mathbf{K}, \mathbf{K}) = 0$, сам вектор $\mathbf{K}$ тоже содержится в H_q. На самом деле, как будет показано ниже, реальный интерес представляет не пространство

Q_q в целом, а только его подпространство S_q, состоящее из классов эквивалентности векторов из H_q, которые отличаются лишь на вектор, кратный $\mathbf{K}$. В случае лучей света элемент S_q можно рассматривать как вектор, описывающий девиацию двух близких лучей, излученных источником в одно и то же время.

Как и прежде, введем дуальные базисы $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$, $\mathbf{E}_4$ и $\mathbf{E}^1$, $\mathbf{E}^2$, $\mathbf{E}^3$, $\mathbf{E}^4$ пространств T_q и T_q^* в некоторой точке q на кривой $\gamma(v)$. Однако не будем брать их ортонормированными: выберем $\mathbf{E}_4$ равным $\mathbf{K}$, в качестве $\mathbf{E}_3$ возьмем некоторый другой изотропный вектор $\mathbf{L}$, скалярное произведение которого с $\mathbf{E}_4$ равно -1, т. е. $g(\mathbf{E}_3, \mathbf{E}_3)=0$, $g(\mathbf{E}_3, \mathbf{E}_4)=-1$, а $\mathbf{E}_1$ и $\mathbf{E}_2$ пусть будут единичными пространственноподобными векторами, ортогональными друг другу и векторам $\mathbf{E}_3$, $\mathbf{E}_4$:

$$g(\mathbf{E}_1, \mathbf{E}_1)=g(\mathbf{E}_2, \mathbf{E}_2)=1,\quad g(\mathbf{E}_1, \mathbf{E}_2)=g(\mathbf{E}_1, \mathbf{E}_3)=g(\mathbf{E}_1, \mathbf{E}_4)=0 \text{ и т. д.}$$

Отметим, что в силу неортонормированного характера этого базиса форма $\mathbf{E}^3$ будет равна форме $-K^a g_{ab}$, а $\mathbf{E}^4 = -L^a g_{ab}$. Можно убедиться, что $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_4$ образуют базис в H_q, в то время как проекции векторов $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ на Q_q образуют базис в Q_q; наконец, проекции $\mathbf{E}_1$ и $\mathbf{E}_2$ образуют базис в S_q. Как правило, мы не будем различать вектор $\mathbf{Z}$ и его проекции на Q_q или S_q. Базис $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$, $\mathbf{E}_4$, обладающий указанными свойствами, будем называть *псевдоортонормированным*. Параллельным переносом этих векторов вдоль геодезической $\gamma(v)$ мы получим псевдоортонормированный базис в каждой точке $\gamma(v)$.

Мы используем этот базис для анализа уравнения девиации изотропных геодезических. Если $\mathbf{Z}$ — вектор девиаций, изображающий относительное расположение соответствующих точек на соседних кривых, то, как и прежде, имеем

$$L_{\mathbf{K}}\mathbf{Z}=0,$$

так что

$$\frac{\mathrm{D}}{\mathrm{d}v} Z^a = K^a{}_{;b} Z^b \tag{4.30}$$

и

$$\frac{\mathrm{D}^2}{\mathrm{d}v^2} Z^a = -R^a{}_{bcd} Z^c K^b K^d. \tag{4.31}$$

В псевдоортонормированном базисе $K^a{}_{;4}=0$, поскольку вектор $\mathbf{K}$ — касательный к геодезической. Поэтому при $a=1, 2, 3$ уравнение (4.30) можно представить как систему обыкновенных дифференциальных уравнений:

$$\frac{\mathrm{d}}{\mathrm{d}v} Z^\alpha = K^\alpha{}_{;\beta} Z^\beta,$$

где, как и прежде, греческие индексы принимают значения 1, 2, 3. Отсюда видно, что проекция $\mathbf{Z}$ на пространство Q_q подчиняется

уравнению распространения, содержащему только саму эту проекцию, без компоненты, параллельной **K**. Далее, $K^3{}_{;c} = 0$, поскольку $(K^a g_{ab} K^b)_{;c} = 0$. Отсюда следует, что компонента $Z^3 = -Z^a K_a$ постоянна вдоль геодезической $\gamma(v)$. Это можно интерпретировать следующим образом: девиация лучей света, испущенных одним и тем же источником в разное время, остается постоянной во времени. Коль скоро это так, наибольший интерес представляет поведение соседних изотропных геодезических с чисто пространственным вектором девиации, т. е. **Z** с $Z^3 = 0$. Проекции таких векторов на Q_q будут лежать тогда в подпространстве S_q и будут подчиняться уравнению

$$\frac{\mathrm{d}}{\mathrm{d}v} Z^m = K^m{}_{;n} Z^n,$$

где m, n принимают только значения 1, 2. Оно подобно уравнению (4.7) для времениподобного случая, за тем исключением, что теперь мы имеем дело с двумерным пространством векторов **Z**, соединяющих соседние кривые.

Как и в предыдущем разделе, Z^m можно выразить через его значения в некоторой точке q:

$$Z^m(v) = \hat{A}_{mn}(v)\, Z^n|_q,$$

где $\hat{A}_{mn}(v)$ — матрица 2×2, удовлетворяющая уравнениям

$$\frac{\mathrm{d}}{\mathrm{d}v} \hat{A}_{mn}(v) = K_{m;p} \hat{A}_{pn}(v), \tag{4.32}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}v^2} \hat{A}_{mn}(v) = -R_{m4p4} \hat{A}_{ph}(v). \tag{4.33}$$

Как и прежде, будем называть $\hat{\omega}_{mn} = K_{[m;n]}$ вращением, $\hat{\theta}_{mn} = K_{(m;n)}$ — скоростью девиации, $\hat{\theta}$, след $\hat{\theta}_{mn}$ — расхождением; сдвиг $\hat{\sigma}_{mn}$ определим как бесследовую часть $\hat{\theta}_{mn}$. Уравнения для них подобны уравнениям для аналогичных величин в предыдущем разделе:

$$\frac{\mathrm{d}}{\mathrm{d}v} \hat{\omega}_{mn} = -\hat{\theta}\hat{\omega}_{mn}, \tag{4.34}$$

$$\frac{\mathrm{d}}{\mathrm{d}v} \hat{\theta} = -R_{ab}K^aK^b + 2\hat{\omega}^2 - 2\hat{\sigma}^2 - \frac{1}{2}\hat{\theta}^2, \tag{4.35}$$

$$\frac{\mathrm{d}}{\mathrm{d}v} \hat{\sigma}_{mn} = -C_{m4n4} - \hat{\theta}\hat{\sigma}_{mn}. \tag{4.36}$$

Уравнение (4.35) является аналогом уравнения Райчаудхури для времениподобных геодезических. Снова мы видим, что вращение вызывает расхождение, в то время как сдвиг приводит к схождению. В следующем разделе мы покажем, что в нормальных условиях член с тензором Риччи $(R_{ab}K^aK^b)$ отрицателен и,

следовательно, порождает фокусирование. Как и прежде, тензор Вейля непосредственно не влияет на расхождение, но вызывает искривление, которое в свою очередь приводит к схождению (ср. [127]).

4.3. Энергетические условия

В реальной Вселенной тензор энергии-импульса будет состоять из вкладов большого числа различных материальных полей. Поэтому, даже если нам были бы известны точный вид вклада каждого поля в отдельности и его уравнения движения, было бы слишком сложно выписать точное выражение для тензора энергии-импульса. В самом деле, мы весьма смутно представляем поведение материи в условиях предельно высоких давлений и плотностей. Таким образом, может показаться, что мало надежды на предсказание существования сингулярностей во Вселенной по уравнениям Эйнштейна, поскольку их правая часть нам неизвестна. Однако, по-видимому, имеются основания для того, чтобы потребовать выполнения определенных неравенств для тензора энергии-импульса. В этом разделе мы собираемся рассмотреть эти неравенства. Оказывается, что во многих ситуациях их достаточно для того, чтобы независимо от вида тензора энергии-импульса доказать наличие сингулярностей.

Первым из этих неравенств является

Слабое энергетическое условие

Тензор энергии импульса в каждой точке $p \in \mathscr{M}$ удовлетворяет неравенству $T_{ab}W^aW^b \geqslant 0$ при любом времениподобном векторе $\mathbf{W} \in T_p$. Вследствие непрерывности это условие будет справедливо и для любого изотропного вектора $\mathbf{W} \in T_p$.

Наблюдателю, мировая линия которого имеет единичным касательным вектором $\mathbf{V}$, локальная плотность энергии кажется равной $T_{ab}V^aV^b$. Следовательно, наше неравенство эквивалентно утверждению, что локальная плотность энергии, измеренная каким-либо наблюдателем, неотрицательна. С точки зрения физики это представляется весьма разумным. Для дальнейшего выяснения смысла этого неравенства воспользуемся тем обстоятельством, что в точке p компоненты тензора энергии-импульса T^{ab} в базисе $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$, $\mathbf{E}_4$ (с времениподобным $\mathbf{E}_4$) можно представить в одной из четырех канонических форм.

Тип I

$$T^{ab} = \begin{pmatrix} p_1 & & & \\ & p_2 & & \\ & & p_3 & \\ & & & \mu \end{pmatrix}.$$

Это тот общий случай, когда тензор энергии-импульса обладает времениподобным собственным вектором $\mathbf{E}_4$, и притом единственным, за исключением случая, когда $\mu = -p_\alpha$ ($\alpha = 1, 2, 3$). Собственное значение μ представляет собой плотность энергии по измерениям наблюдателя, мировая линия которого в точке p имеет касательным вектором $\mathbf{E}_4$; собственные значения p_α ($\alpha = 1, 2, 3$) являются главными значениями давления в трех пространственноподобных направлениях $\mathbf{E}_\alpha$. Таков вид энергии-импульса для всех наблюдаемых полей с ненулевой массой покоя, а также и для всех полей с нулевой массой покоя, за исключением одного особого случая, когда мы имеем каноническую форму типа II.

Тип II

$$T^{ab} = \begin{pmatrix} p_1 & 0 & & \\ 0 & p_2 & & \\ & & \nu - \varkappa & \nu \\ & & \nu & \nu + \varkappa \end{pmatrix}, \quad \nu = \pm 1.$$

В этом особом случае тензор энергии-импульса имеет двукратный собственный вектор $\mathbf{E}_3 + \mathbf{E}_4$. Единственный наблюдаемый случай такой формы — когда безмассовое поле представляет собой излучение, целиком распространяющееся в направлении $\mathbf{E}_3 + \mathbf{E}_4$. В этом случае p_1, p_2 и $\varkappa$ равны нулю.

Тип III

$$T^{ab} = \begin{pmatrix} p & 0 & 0 & 0 \\ 0 & -\nu & 1 & 1 \\ 0 & 1 & -\nu & 0 \\ 0 & 1 & 0 & \nu \end{pmatrix}.$$

Это — особый случай, в котором тензор энергии-импульса обладает трехкратным изотропным собственным вектором $\mathbf{E}_3 + \mathbf{E}_4$. Полей, обладающих тензором энергии такого вида, не обнаружено.

Тип IV

$$T^{ab} = \begin{pmatrix} p_1 & 0 & & \\ 0 & p_2 & & \\ & & -\varkappa & \nu \\ & & \nu & 0 \end{pmatrix}, \quad \varkappa^2 < 4\nu^2.$$

Это — общий случай, когда у тензора энергии-импульса нет времениподобных или изотропных собственных векторов. Наблюдаемых полей с тензорами энергии-импульса такого вида тоже не существует.

Для типа I слабое энергетическое условие будет выполняться при $\mu \geqslant 0$, $\mu + p_\alpha \geqslant 0$ $(\alpha = 1, 2, 3)$. Для типа II оно будет выполнено, если $p_1 \geqslant 0$, $p_2 \geqslant 0$, $\varkappa > 0$, $\nu = +1$. Эти неравенства являются весьма разумными, и им удовлетворяют все экспериментально обнаруженные поля. Для физически нереализующихся типов III и IV слабое энергетическое условие не выполняется.

Этому условию удовлетворяет также скалярное поле ϕ, постулированное Брансом и Дикке [39]. От этого поля требуется, чтобы оно всюду было положительным. Его тензор энергии-импульса имеет вид (3.6), причем $m = 0$. Тензор энергии-импульса всех других полей отличается в ϕ раз от того, что было бы, если бы это скалярное поле не существовало.

Слабое энергетическое условие не выполняется для «C-поля», предложенного Хойлом и Нарликаром [82], которое также является скалярным полем с $m = 0$; только на этот раз тензор энергии-импульса имеет противоположный знак и, следовательно, плотность энергии отрицательна. Ввиду этого возможно одновременное рождение квантов полей с положительной энергией и C-поля с отрицательной энергией. Этот процесс происходит в стационарной модели Вселенной, предложенной Хойлом и Нарликаром, в которой по мере разлета частиц вследствие общего расширения Вселенной непрерывно рождается новое вещество, так что поддерживается постоянная средняя плотность. Однако подобный процесс вызывает трудности с точки зрения квантовой механики. Даже если сечение такого процесса очень мало, наличие бесконечного фазового объема для квантов положительной и отрицательной энергии привело бы к рождению бесконечного числа пар в конечной области пространства-времени.

Такой катастрофы не случится, если выполняется слабое энергетическое условие. Существует также несколько более сильное условие, при котором невозможно рождение материи в том смысле, что пространство-время остается пустым все время, если оно пустое в какой-либо момент времени, и никакая материя не поступает из бесконечности; наоборот, материя, имеющаяся в какой-то момент времени, не может исчезнуть и, таким образом, присутствует всегда. Этим условием является

Условие энергодоминантности

Для любого времениподобного $W_a T^{ab} W_a W_b \geqslant 0$, а $T^{ab} W_a$ — непространственноподобный вектор.

Это условие можно интерпретировать следующим образом: любой наблюдатель видит, что локальная энергия неотрицательна, а локальный поток энергии непространственноподобен. Эквивалентным этому является утверждение, что в любом ортонормированном базисе энергия доминирует над другими компонентами T_{ab}, т. е.

$$T^{00} \geqslant |T^{ab}| \text{ для любых } a, b.$$

Это условие выполняется для типа I при $\mu \geqslant 0$, $-\mu \leqslant p_\alpha \leqslant \mu$ ($\alpha = 1, 2, 3$) и для типа II при $\nu = +1$, $\varkappa \geqslant 0$, $0 \leqslant p_i \leqslant \varkappa$

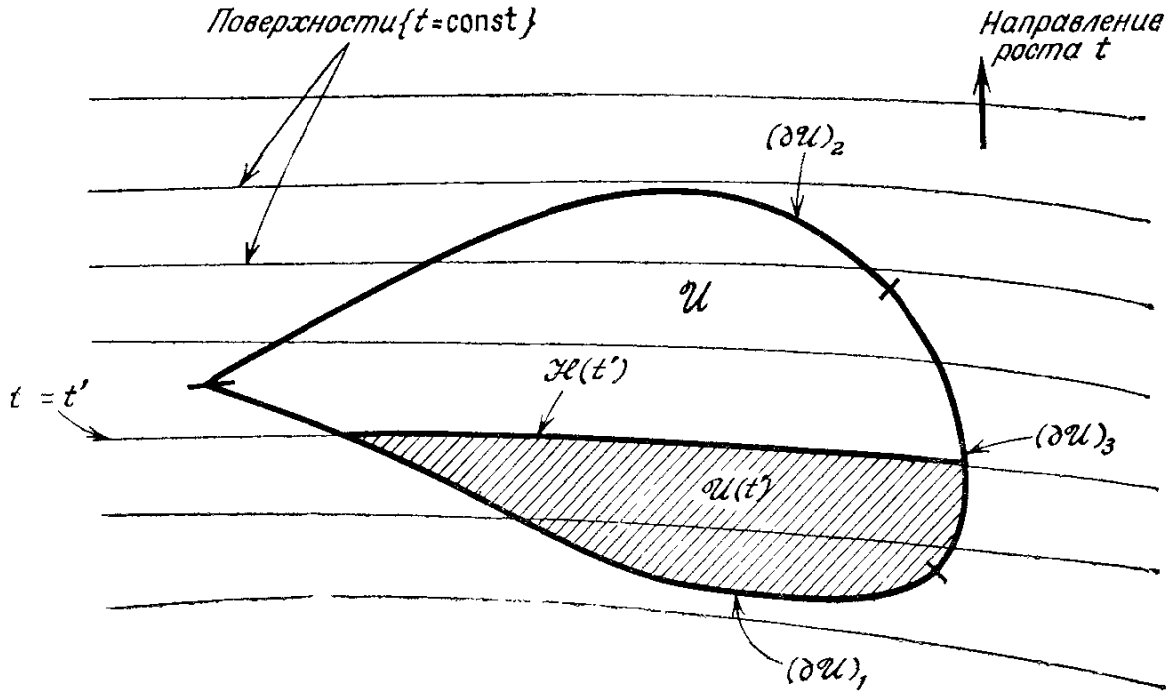


Рис. 9. Компактная область $\mathcal{U}$ пространства-времени с невременноподобными границами прошлого и будущего $(\partial\mathcal{U})_1$, $(\partial\mathcal{U})_2$ и временноподобной границей $(\partial\mathcal{U})_3$. $\mathcal{U}(t')$ — часть $\mathcal{U}$, лежащая в прошлом поверхности $\mathcal{H}(t')$ (определяемой значением $t = t'$).

($i = 1, 2$). Другими словами, условие энергодоминантности представляет собой слабое энергетическое условие, дополненное требованием, чтобы давление не превышало плотности энергии. Ему удовлетворяют все известные формы материи и имеются серьезные основания предполагать, что оно выполняется во всех ситуациях. Дело в том, что скорость звуковых волн, распространяющихся в направлении E_α, равна скорости света, умноженной на $(dp_\alpha/d\mu)_{\text{адиаб.}}$ Значит, должно быть справедливо неравенство $(dp_\alpha/d\mu) \leqslant 1$, ибо, согласно постулату (а) разд. 3.2, никакой сигнал не может распространяться быстрее света. Отсюда следует, что $p_\alpha \leqslant \mu$, поскольку для любой известной формы материи давления малы, когда малы плотности. (Блудман и Рудерман [8, 9] показали, что могут быть поля, для которых перенормировка массы может привести к тому, что давление будет больше

плотности энергии. Однако, на наш взгляд, это скорее говорит против теории перенормировок, чем в пользу возможности такой ситуации.)

Рассмотрим теперь ситуацию, изображенную на рис. 9, когда имеется C^2-функция t, градиент которой всюду времениподобен. (В разд. 6.4 будет показано, что такая функция существует, если только пространство-время не слишком близко к нарушению причинности.) Граница $\partial\mathcal{U}$ компактной области $\mathcal{U}$ состоит: 1) из части $(\partial\mathcal{U})_1$, нормальная форма $\mathbf{n}$ которой непространственноподобна и такая, что на $n_a t^{;b} g^{ab} > 0$; 2) из части $(\partial\mathcal{U})_2$, нормальная форма $\mathbf{n}$ которой непространственноподобна, но $n_a t^{;b} g^{ab} < 0$ и 3) из остальной части $(\partial\mathcal{U})_3$ (которая может быть пуста). Знак нормальной формы $\mathbf{n}$ получается из требования, чтобы $\langle \mathbf{n}, \mathbf{X} \rangle$ было положительным для всех векторов $\mathbf{X}$, которые направлены наружу $\mathcal{U}$ (ср. с разд. 2.8); $\mathcal{H}(t')$ означает поверхность $t = t'$, а $\mathcal{U}(t')$ — область в $\mathcal{U}$, в которой $t < t'$. Для дальнейшего использования в разд. 7.4 мы установим одно неравенство, которое выполняется не только для T^{ab}, но и для любого симметричного тензора S^{ab}, удовлетворяющего условию энергодоминантности. В приложении к тензору энергии-импульса это неравенство будет означать, что он обращается в нуль всюду в $\mathcal{U}$, если он равен нулю на $(\partial\mathcal{U})_3$ и на начальной поверхности $(\partial\mathcal{U})_1$.

Лемма 4.3.1.

Существует некоторая положительная постоянная P, такая, что для любого тензора S^{ab}, удовлетворяющего условию энергодоминантности и обращающегося в нуль на $(\partial\mathcal{U})_3$,

$$\int\limits_{\mathcal{H}(t)\cap\mathcal{U}} S^{ab} t_{;a}\, \mathrm{d}\sigma_b \leqslant -\int\limits_{(\partial\mathcal{U})_1} S^{ab} t_{;a}\, d\sigma_b + \\ + P\int^{t}\left(\int\limits_{\mathcal{H}(t')\cap\mathcal{U}} S^{ab} t_{;a}\, \mathrm{d}\sigma_b\right)\mathrm{d}t' + \int^{t}\left(\int\limits_{\mathcal{H}(t')\cap\mathcal{U}} S^{ab}{}_{;a}\, \mathrm{d}\sigma_b\right)\mathrm{d}t'.$$

Рассмотрим интеграл по объему

$$I(t) = \int\limits_{\mathcal{U}(t)} (S^{ab} t_{;a})_{;b}\, \mathrm{d}v = \int\limits_{\mathcal{U}(t)} S^{ab} t_{;ab}\, \mathrm{d}v + \int\limits_{\mathcal{U}(t)} S^{ab}{}_{;b} t_{;a}\, \mathrm{d}v.$$

По теореме Гаусса его можно преобразовать в интеграл по границе области $\mathcal{U}(t)$:

$$I(t) = \int\limits_{\partial\mathcal{U}(t)} S^{ab} t_{;a}\, \mathrm{d}\sigma_b.$$

Граница $\mathcal{U}(t)$ состоит из $\mathcal{U}(t) \cap \partial\mathcal{U}$ и $\mathcal{U} \cap \mathcal{H}(t)$. Поскольку $S^{ab} = 0$ на $(\partial\mathcal{U})_3$, то

$$I(t) = \int\limits_{\mathcal{U}(t)\cap(\partial\mathcal{U})_1} + \int\limits_{\mathcal{U}(t)\cap(\partial\mathcal{U})_2} + \int\limits_{\mathcal{U}\cap\mathcal{H}(t)} .$$

В силу условия энергодоминантности $S^{ab}t_{;a}$ существует непространственноподобный вектор и $S^{ab}t_{;a}t_{;b} \geqslant 0$. Так как нормальная форма на $(\partial\mathcal{U})_2$ непространственноподобна и $n_a t_{;b} g^{ab} < 0$, то второй член в правой части неотрицателен. Следовательно,

$$\int\limits_{\mathcal{U}\cap\mathcal{H}(t)} S^{ab}t_{;a}\,d\sigma_b \leqslant -\int\limits_{\mathcal{U}(t)\cap(\partial\mathcal{U})_1} S^{ab}t_{;a}\,d\sigma_b + \int\limits_{\mathcal{U}(t)} (S^{ab}t_{;ab} + S^{ab}{}_{;b}t_{;a})\,dv.$$

В силу компактности $\mathcal{U}$ в любом ортонормированном базисе с времениподобным вектором, направленным по $t_{;ab}$, существует некоторая верхняя граница для компонент $t_{;a}$. Значит, существует такое $P > 0$, что в $\mathcal{U}$

$$S^{ab}t_{;ab} \leqslant PS^{ab}t_{;a}t_{;b}$$

для любого S^{ab}, удовлетворяющего условию энергодоминантности. Интеграл по $\mathcal{U}(t)$ можно представить в виде поверхностного интеграла по $\mathcal{H}(t') \cap \mathcal{U}$ с последующим интегрированием по t':

$$\int\limits_{\mathcal{U}(t)} (PS^{ab}t_{;a}t_{;b} + S^{ab}{}_{;b}t_{;a})\,dv = \int^{t} \left\{ \int\limits_{\mathcal{H}(t')\cap\mathcal{U}} (PS^{ab}t_{;b} + S^{ab}{}_{;b})\,d\sigma_a \right\} dt',$$

где $d\sigma_a$ — элемент поверхности $\mathcal{H}(t')$. Итак,

$$\int\limits_{\mathcal{H}(t)\cap\mathcal{U}} S^{ab}t_{;a}\,d\sigma_b \leqslant -\int\limits_{\mathcal{U}(t)\cap(\partial\mathcal{U})_1} S^{ab}t_{;a}\,d\sigma_b + P\int^{t}\left(\int\limits_{\mathcal{H}(t')\cap\mathcal{U}} S^{ab}t_{;a}\,d\sigma_b\right)dt' + \int^{t}\left(\int\limits_{\mathcal{H}(t')\cap\mathcal{U}} S^{ab}{}_{;a}\,d\sigma_b\right)dt'. \quad \square$$

Из этого результата сразу следует

Теорема сохранения

Если тензор энергии-импульса удовлетворяет условию энергодоминантности и равен нулю на $(\partial\mathcal{U})_3$ и на начальной поверхности $(\partial\mathcal{U})_1$, тогда он равен нулю всюду в $\mathcal{U}$.

Допустим, что

$$x(t) = \int\limits_{\mathscr{U}(t)} T^{ab} t_{;a} t_{;b}\, \mathrm{d}v = \int^{t} \left(\int\limits_{\mathscr{H}(t') \cap \mathscr{U}} T^{ab} t_{;a}\, d\sigma_b \right) \mathrm{d}t' \geqslant 0.$$

Тогда по лемме 4.3.1 получаем $dx/\mathrm{d}t \leqslant Px$. Но при достаточно малых значениях t $\mathscr{H}(t)$ не пересекает $\mathscr{U}$ и, следовательно, $x = 0$. Тогда $x=0$ при всех t и, следовательно, $T^{ab}=0$ в $\mathscr{U}$. □

Из теоремы сохранения следует, что тензор энергии-импульса, равный нулю на множестве $\mathscr{S}$, равен нулю также на области Коши будущего, $D^+(\mathscr{S})$, которая по определению есть множество всех точек, обладающих тем свойством (рис. 10), что

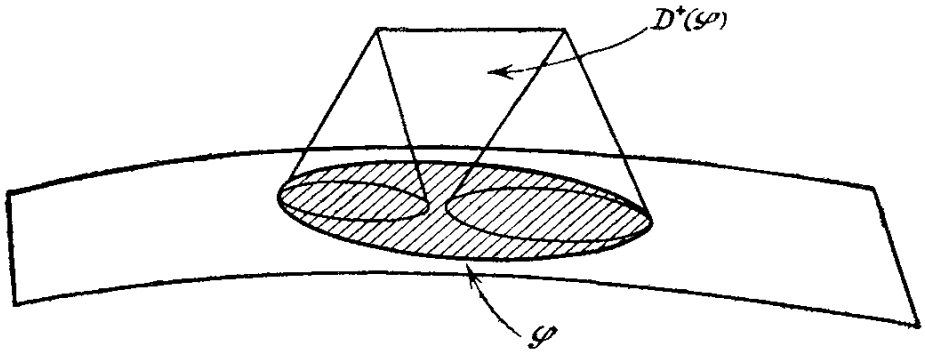


Рис. 10. Область Коши будущего D^+ $(\mathscr{S})$ пространственноподобного множества $(\mathscr{S})$.

любая проходящая через них непространственноподобная кривая, направленная в прошлое, пересекает $\mathscr{S}$ (ср. с разд. 6.5). Действительно, если q — какая-нибудь точка $D^+(\mathscr{S})$, то область $D^+(\mathscr{S})$, расположенная в прошлом q, компактна (предложение 6.6.6), и ее можно взять в качестве $\mathscr{U}$. Этот результат можно интерпретировать в том смысле, что из условия энергодоминантности следует невозможность движения материи быстрее света.

Значение слабого энергетического условия для нашего исследования сингулярностей состоит в том, что из него следует вывод о фокусирующем (точнее недефокусирующем) действии материи на конгруэнции изотропных геодезических. Если нет вращения, расхождение $\hat{\theta}$ подчиняется уравнению

$$\frac{\mathrm{d}}{\mathrm{d}v}\hat{\theta} = -R_{ab}K^aK^b - 2\hat{\sigma}^2 - \frac{1}{2}\hat{\theta}^2.$$

Таким образом, $\hat{\theta}$ в этом случае монотонно убывает вдоль изотропной геодезической, если $R_{ab}W^aW^b \geqslant 0$ для любого изотропного вектора $\mathbf{W}$. Назовем это условие *условием изотропного схождения*. Из уравнений Эйнштейна

$$R_{ab} - \frac{1}{2} g_{ab}R + \Lambda g_{ab} = 8\pi T_{ab}$$

следует, что условие изотропного схождения независимо от значения Λ содержится в слабом энергетическом условии.

Из (4.26) можно получить, что расхождение θ времениподобной геодезической конгруэнции будет монотонно убывать вдоль геодезической, если $R_{ab}W^aW^b \geqslant 0$ для любого времениподобного $\mathbf{W}$. Это условие назовем *условием времениподобного схождения*. В силу уравнений Эйнштейна оно удовлетворяется, если для тензора энергии-импульса выполняется неравенство

$$T_{ab}W^aW^b \geqslant W^aW_a\left(\tfrac{1}{2}T - \frac{1}{8\pi}\Lambda\right).$$

Для типа I это имеет место при

$$\mu + p_\alpha \geqslant 0, \quad \mu + \sum p_\alpha - \frac{1}{4\pi}\Lambda \geqslant 0,$$

а для типа II — при

$$\nu = +1, \quad \varkappa \geqslant 0, \quad p_1 \geqslant 0, \quad p_2 \geqslant 0 \quad \text{и} \quad p_1 + p_2 - \frac{1}{4\pi}\Lambda \geqslant 0.$$

Будем говорить, что тензор энергии-импульса удовлетворяет *сильному энергетическому условию*, если приведенное неравенство выполняется при $\Lambda = 0$. Это требование сильнее слабого энергетического условия, и все же в приложении к полному тензору энергии-импульса оно физически обосновано. Действительно, в общем случае (тип I) оно нарушается только наличием отрицательной плотности энергии или большого отрицательного давления (например, для идеальной жидкости плотностью 1 г/см^3 нарушение может произойти только при $p < -10^{15}$ атм). Это условие выполняется для электромагнитного поля и безмассового скалярного поля (в частности, для скалярного поля Бранса — Дикке). Для скалярного поля с ненулевой массой m тензор энергии-импульса имеет вид (разд. 3.3)

$$T_{ab} = \phi_{;a}\phi_{;b} - \frac{1}{2}g_{ab}(\phi_{;c}\phi_{;d}g^{cd} + m^2\phi^2).$$

Отсюда, если W^a — единичный времениподобный вектор, имеем

$$T_{ab}W^aW^b - \frac{1}{2}W_aW^aT = (\phi_{;a}W^a)^2 - \frac{1}{2}\frac{m^2}{\hbar^2}\phi^2, \qquad (4.37)$$

т. е. это выражение может быть отрицательным. Однако из уравнений скалярного поля следует

$$\frac{1}{2}\frac{m^2}{\hbar^2}\phi^2 = \frac{1}{2}\phi\phi_{;ab}g^{ab}.$$

Учитывая это соотношение в (4.37) и интегрируя по некоторой области $\mathcal{U}$, получаем

$$\frac{1}{2}\int_{\mathcal{U}}(g^{ab} + 2W^aW^b)\,\phi_{;a}\phi_{;b}\,d\sigma + \frac{1}{2}\int_{\partial\mathcal{U}}\phi\phi_{;a}g^{ab}\,d\sigma_b.$$

Первый член неотрицателен, ибо $g^{ab} + 2W^aW^b$ есть положительно-определенная метрика, а второй член мал по сравнению с первым, если область $\mathcal{U}$ велика по сравнению с длиной волны $\hbar/m$. Для π-мезонов, которые классически можно описывать скалярным полем, $m = 6 \cdot 10^{-25}$ г, $\hbar/m = 3 \cdot 10^{-13}$ см. Итак, хотя тензор энергии-импульса π-мезонов может и не удовлетворять сильному энергетическому условию в каждой точке, это не будет влиять на схождение времениподобных геодезических на расстояниях, превышающих 10^{-12} см. Возможно, такое нарушение сильного энергетического условия привело бы к тому, что теоремы о сингулярностях (гл. 8) теряли бы справедливость при радиусе кривизны меньше 10^{-12} см, но при этом кривизна столь велика, что ее можно было бы с успехом принять за сингулярность (разд. 10.2).

4.4. Сопряженные точки

В разд. 4.1 мы видели, что компоненты вектора девиации, характеризующего взаимное расположение $\gamma(s)$ и соседней кривой в конгруэнции времениподобных геодезических, подчиняются уравнению Якоби:

$$\frac{d^2}{ds^2} Z^\alpha = - R_{\alpha 4 \beta 4} Z^\beta \quad (\alpha, \beta = 1, 2, 3). \tag{4.38}$$

Решение этого уравнения будем называть *якобиевым полем* вдоль $\gamma(s)$. Поскольку решение может быть охарактеризовано заданием Z^α и dZ^α/ds в некоторой точке $\gamma(s)$, то вдоль $\gamma(s)$ будет шесть якобиевых полей. Существует три независимых якобиевых поля, равных нулю в некоторой точке $q \in \gamma(s)$. Их можно записать следующим образом:

$$Z^\alpha(s) = A_{\alpha\beta}(s) \frac{d}{ds} Z^\beta \Big|_q,$$

где

$$\frac{d^2}{ds^2} A_{\alpha\beta}(s) = - R_{\alpha 4 \gamma 4} A_{\gamma\beta}(s), \tag{4.39}$$

и $A_{\alpha\beta}(s)$ — матрица 3×3, равная нулю в q. Можно считать, что эти якобиевы поля описывают девиацию (взаимное разделение) соседних геодезических, проходящих через q. Как и прежде, можно определить вращение, поперечный сдвиг и расхождение якобиевых полей вдоль $\gamma(s)$, обращающихся в нуль в q:

$$\omega_{\alpha\beta} = A^{-1}{}_{\gamma[\beta} \frac{d}{ds} A_{\alpha]\gamma}, \tag{4.40}$$

$$\sigma_{\alpha\beta} = A^{-1}{}_{\gamma(\beta} \frac{d}{ds} A_{\alpha)\gamma} - \frac{1}{3} \delta_{\alpha\beta}\theta, \tag{4.41}$$

$$\theta = (\det \mathbf{A})^{-1} \frac{d}{ds} (\det \mathbf{A}). \tag{4.42}$$

Они будут подчиняться выведенным в разд. 4.1 уравнениям с $\dot{V}_a = 0$. В частности, матрица

$$A_{\gamma\alpha}\omega_{\gamma\delta}A_{\delta\beta} = \frac{1}{2}\left(A_{\gamma\alpha}\frac{\mathrm{d}}{\mathrm{d}s}A_{\gamma\beta} - A_{\gamma\beta}\frac{\mathrm{d}}{\mathrm{d}s}A_{\gamma\alpha}\right)$$

будет постоянной вдоль $\gamma(s)$. Однако она равна нулю в q, где $A_{\alpha\beta} = 0$. Следовательно, $\omega_{\alpha\beta} = 0$ везде, где $A_{\alpha\beta}$ неособенна.

Будем говорить, что точка p на $\gamma(s)$ *сопряжена точке q вдоль* $\gamma(s)$, если найдется якобиево поле, не равное тождественно нулю, которое исчезает в q и p. Можно представлять себе p как точку, в которой пересекаются бесконечно близкие геодезические, проходящие через q. (Отметим, однако, что в p будут пересекаться только *бесконечно близкие* геодезические; не обязательно, чтобы две различные геодезические, выходя из q, проходили через p.) Якобиевы поля, обращающиеся в нуль в q, определяются матрицами $A_{\alpha\beta}$. Поэтому точка p сопряжена точке q вдоль $\gamma(s)$ тогда и только тогда, когда $A_{\alpha\beta}$ — особенная матрица в p. Объемное расхождение θ определяется как $(\det \mathbf{A})^{-1}\mathrm{d}(\det \mathbf{A})/\mathrm{d}s$. Поскольку $A_{\alpha\beta}$ подчиняется уравнению (4.39) с конечным $R_{\alpha 4\gamma 4}$, $\mathrm{d}(\det\mathbf{A})/\mathrm{d}s$ должно быть конечно. Следовательно, p будет сопряжена q вдоль $\gamma(s)$, если θ становится бесконечным в p. Обратное тоже верно, ибо $\theta = \mathrm{d}\ln(\det\mathbf{A})/\mathrm{d}s$, а матрица $A_{\alpha\beta}$ может быть особенной только в изолированных точках, иначе она была бы особенной всюду.

Предложение 4.4.1

Пусть в некоторой точке $\gamma(s_1)$ $(s_1 > 0)$ объемное расхождение θ принимает значение $\theta_1 < 0$ и пусть всюду $R_{ab}V^aV^b \geqslant 0$; тогда между $\gamma(s_1)$ и $\gamma(s_1 + (3/-\theta_1))$ [при условии, что $\gamma(s)$ можно продолжать до этих значений параметра] существует точка, сопряженная q вдоль $\gamma(s)$. (Это последнее условие может оказаться невыполненным, если пространство-время геодезически неполно. В гл. 8 такая неполнота интерпретируется как проявление сингулярности.)

Объемное расхождение θ матрицы $A_{\alpha\beta}$ подчиняется уравнению Райчаудхури (4.26):

$$\frac{\mathrm{d}}{\mathrm{d}s}\theta = -R_{ab}V^aV^b - 2\sigma^2 - \frac{1}{3}\theta^2.$$

(Здесь мы пользуемся тем, что вращение отсутствует.) Все члены в правой части отрицательны. Следовательно, при $s > s_1$

$$\theta \leqslant \frac{3}{s - (s_1 + (3/-\theta_1))}.$$

Отсюда следует, что θ обращается в бесконечность, так что существует точка, сопряженная q, при значении s_0, заключенном между s_1 и $s_1 + (3/-\theta_1)$. □

Другими словами, если выполняется условие времениподобного схождения и если соседние геодезические, выходящие из q, начинают сходиться к $\gamma(s)$, то некоторые бесконечно близкие геодезические пересекут $\gamma(s)$, если только можно провести $\gamma(s)$ до достаточно больших значений параметра s.

Предложение 4.4.2
Если $R_{ab}V^aV^b \geqslant 0$ и если в некоторой точке $p = \gamma(s_1)$ приливная сила $R_{abcd}V^bV^d$ не равна нулю, найдутся такие значения s_0 и s_2, что $q = \gamma(s_0)$ и $r = \gamma(s_2)$ будут сопряжены вдоль $\gamma(s)$ при условии, что $\gamma(s)$ можно продолжить до этих значений s.

Решение уравнения (4.39) вдоль $\gamma(s)$ однозначно определяется значениями $A_{\alpha\beta}$ и $\mathrm{d}A_{\alpha\beta}/\mathrm{d}s$ в точке p. Рассмотрим множество P, состоящее из всех таких решений, для которых $A_{\alpha\beta}|_p = \delta_{\alpha\beta}$, а $(\mathrm{d}A_{\alpha\beta}/\mathrm{d}s)|_p$ — симметричная матрица со следом $\theta|_p \geqslant 0$. Для каждого решения из P найдется некоторое $s_3 > s_1$, при котором $A_{\alpha\beta}(s_3)$ — особенная матрица, поскольку или $\theta|_p < 0$, и тогда это следует из предыдущего результата, или $\theta|_p = 0$, и тогда $(\mathrm{d}\sigma_{\alpha\beta}/\mathrm{d}s)|_p \neq 0$, а, значит, $\sigma^2 > 0$, и вследствие этого θ становится отрицательным при $s > s_1$. Элементы множества P находятся во взаимно-однозначном соответствии с пространством S всех симметричных матриц 3×3 с неположительным следом [т. е. со всеми $(\mathrm{d}A_{\alpha\beta}/\mathrm{d}s)|_p$]. Следовательно, имеется отображение η из S в $\gamma(s)$, которое каждому начальному значению $(\mathrm{d}A_{\alpha\beta}/\mathrm{d}s)|_p$ ставит в соответствие ту точку на $\gamma(s)$, в которой $A_{\alpha\beta}$ в первый раз становится особенной. Это отображение непрерывно. Далее, если какая-либо из компонент $(\mathrm{d}A_{\alpha\beta}/\mathrm{d}s)|_p$ очень велика, соответствующая точка на $\gamma(s)$ будет находиться близко к p, поскольку член $R_{\alpha 4\gamma 4}$ в (4.39) в пределе становится пренебрежимо малым и решение приобретает тот же вид, что и в плоском случае. Таким образом, найдутся такие постоянные $C > 0$ и $s_4 > s_1$, что если какая-либо компонента $(\mathrm{d}A_{\alpha\beta}/\mathrm{d}s)|_p$ больше C, то соответствующая точка на $\gamma(s)$ ближе к p, чем $\gamma(s_4)$. Но в пространстве S подпространство, состоящее из всех матриц, элементы которых меньше или равны C, компактно. Отсюда видно, что существует такое $s_5 > s_1$, что $\eta(S)$ содержится в сегменте $[\gamma(s_1), \gamma(s_5)]$. Рассмотрим теперь точку $r = \gamma(s_2)$, причем $s_2 > s_5$. Если между r и p нет точек, сопряженных r, якобиевы поля, равные нулю в r, должны иметь в точке p положительное расхождение θ (иначе они принадлежали бы множеству P, представляющему собой все семейства якобиевых полей без вращения с неположительным объемным расхождением в p). Тогда из предыдущего предложения следует, что найдется точка $q = \gamma(s_0)$ $(s_0 < s_1)$, которая сопряжена r вдоль $\gamma(s)$. □

Для реалистического с точки зрения физики решения уравнений Эйнштейна можно ожидать, что каждая времениподобная геодезическая проходит сквозь какую-то материю или гравитационное излучение и, следовательно, содержит точку, где $R_{abcd}V^bV^d \neq 0$ (хотя этого может и не быть в точном решении с высокой симметрией). Было бы разумно предположить, что в таком решении каждая времениподобная геодезическая содержит пары сопряженных точек, при условии, что эта линия может быть продолжена достаточно далеко в обоих направлениях.

Мы рассмотрим также конгруэнцию времениподобных геодезических, нормальных к пространственноподобной трехмерной поверхности $\mathscr{H}$. *Пространственноподобной 3-поверхностью* мы называем вложенное трехмерное подмногообразие, определенное локально уравнением $f = 0$, где f есть C^2-функция и $g^{ab}f_{;a}f_{;b} < 0$ при $f = 0$. Введем **N**, единичный нормальный вектор к $\mathscr{H}$: $N^a = (-g^{bc}f_{;b}f_{;c})^{-1/2}g^{ad}f_{;d}$ и второй фундаментальный тензор χ на $\mathscr{H}$: $\chi_{ab} = h_a{}^c h_b{}^d N_{c;d}$, где $h_{ab} = g_{ab} + N_aN_b$ называется первым фундаментальным тензором (или индуцированным метрическим тензором) на $\mathscr{H}$ (ср. с разд. 2.7). Из этого определения следует, что тензор χ симметричен. Конгруэнция времениподобных геодезических, ортогональных к $\mathscr{H}$, будет состоять из времениподобных геодезических, у которых единичные касательные векторы **V** на $\mathscr{H}$ равны **N**. Тогда имеем

$$V_{a;b} = \chi_{ab} \quad \text{на} \quad \mathscr{H}. \tag{4.43}$$

Вектор девиации **Z**, изображающий расположение относительно $\gamma(s)$ соседних геодезических, нормальных к $\mathscr{H}$, будет подчиняться уравнению Якоби (4.38). На $\gamma(s)$ Z в точке $q \in \mathscr{H}$ удовлетворяет начальному условию

$$\frac{\mathrm{d}}{\mathrm{d}s}Z^\alpha = \chi_{\alpha\beta}Z^\beta. \tag{4.44}$$

Мы представим якобиевы поля вдоль $\gamma(s)$, удовлетворяющие этому условию, в виде

$$Z^\alpha(s) = A_{\alpha\beta}(s)\,Z^\beta|_q,$$

где

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}A_{\alpha\beta} = -R_{\alpha 4\gamma 4}A_{\gamma\beta}, \tag{4.45}$$

а в точке q $A_{\alpha\beta}$ — единичная матрица, и

$$\frac{\mathrm{d}}{\mathrm{d}s}A_{\alpha\beta} = \chi_{\alpha\gamma}A_{\gamma\beta}. \tag{4.46}$$

Будем говорить, что точка $p \in \gamma(s)$ *сопряжена* поверхности $\mathscr{H}$ *вдоль* $\gamma(s)$, если существует якобиево поле вдоль $\gamma(s)$, не равное тождественно нулю, удовлетворяющее в q начальному

условию 4.4.4 и обращающееся в нуль в p. Другими словами, p сопряженно $\mathscr{H}$ вдоль $\gamma(s)$, если и только если $A_{\alpha\beta}$ особенна в p. Можно представлять себе p как точку, в которой пересекаются соседние геодезические, нормальные к $\mathscr{H}$. Как и прежде, $A_{\alpha\beta}$ будет особенна там и только там, где расхождение становится бесконечным. Начальное значение $A_{\gamma\alpha}\omega_{\gamma\delta}A_{\delta\beta}$ в q равно нулю, и поэтому $\omega_{\alpha\beta} = 0$ на $\gamma(s)$. Начальное значение θ равно $\chi_{ab}q^{ab}$.

Предложение 4.4.3

Если $R_{ab}V^aV^b \geqslant 0$ и $\chi_{ab}g^{ab} < 0$, найдется точка, сопряженная $\mathscr{H}$ вдоль $\gamma(s)$ на расстоянии от $\mathscr{H}$ не более $3/(-\chi_{ab}g^{ab})$, при условии, что $\gamma(s)$ можно продолжить на такое расстояние.

Для доказательства можно использовать уравнение Райчаудхури (4.26), как и при доказательстве предложения 4.4.1. □

Будем называть решение уравнения

$$\frac{\mathrm{d}^2}{\mathrm{d}v^2} Z^m = -R_{m4n4}Z^n \quad (m,\ n = 1,\ 2)$$

вдоль изотропной геодезической $\gamma(v)$ *якобиевым полем вдоль* $\gamma(v)$. Компоненты Z^m можно рассматривать как компоненты вектора в пространстве S_q каждой точки q относительно базиса $\mathbf{E}_1$, $\mathbf{E}_2$. Будем говорить, что p сопряжено q вдоль изотропной геодезической $\gamma(v)$, если существует не равное тождественно нулю якобиево поле вдоль $\gamma(v)$, обращающееся в нуль в q и p. Если $\mathbf{Z}$ — вектор, соединяющий соседние изотропные геодезические, проходящие через q, компонента Z^3 будет равна нулю всюду. Таким образом, p можно представлять себе как точку, в которой пересекаются бесконечно близкие геодезические, проходящие через q. Опишем обращающиеся в нуль в q якобиевы поля вдоль $\gamma(s)$ матрицами $2 \times 2\hat{A}_{mn}$:

$$Z^m(v) = \hat{A}_{mn}\frac{\mathrm{d}}{\mathrm{d}v} Z^n|_q.$$

Как и прежде, имеем $\hat{A}_{lm}\hat{\omega}_{lk}\hat{A}_{kn} = 0$, и поэтому у якобиевых полей, равных нулю в p, вращение отсутствует. И снова p будет сопряженно q вдоль $\gamma(s)$, если и только если

$$\hat{\theta} = (\det \hat{\mathbf{A}})^{-1}\frac{\mathrm{d}}{\mathrm{d}v}(\det \hat{\mathbf{A}})$$

обращается в p в бесконечность. Аналогично предложению 4.4.1 имеем

Предложение 4.4.4

Если всюду $R_{ab}K^aK^b \geqslant 0$ и если в некоторой точке $\gamma(v_1)$ расхождение $\hat{\theta}$ имеет отрицательное значение $\hat{\theta}_1 < 0$, тогда между $\gamma(v_1)$ и $\gamma(v_1 + (2/-\hat{\theta}_1))$ найдется точка, сопряженная q вдоль

$\gamma(v)$, при условии, что $\gamma(v)$ можно продолжить на такое расстояние.

Расхождение $\hat{\theta}$ матрицы $\bar{A}_{mn}$ подчиняется уравнению (4.3.5):

$$\frac{\mathrm{d}}{\mathrm{d}v}\hat{\theta} = -R_{ab}K^aK^b - 2\hat{\sigma}^2 - \frac{1}{2}\hat{\theta}^2$$

и, следовательно, доказательство проводится, как прежде. □

Предложение 4.4.5

Если всюду $R_{ab}K^aK^b \geqslant 0$ и если $p = \gamma(v_1)$, а $K^cK^dK_{[a}R_{b]cd[e}K_{f]}$ отлично от нуля, то найдутся такие v_0 и v_2, что $q = \gamma(v_0)$ и $r = \gamma(v_2)$ будут сопряжены вдоль $\gamma(v)$ при условии, что $\gamma(v)$ можно продолжить до этих значений v.

Если $K^cK^dK_{[a}R_{b]cd[e}K_{f]}$ отлично от нуля, то же верно и для R_{m4n4}. Тогда доказательство аналогично доказательству предложения 4.4.2. □

Как и во времениподобном случае, условие $K^cK^dK_{[a}R_{b]cd[e}K_{f]} \neq 0$ выполняется для изотропных геодезических, проходящих сквозь некоторую материальную среду, за исключением случая, когда эта среда есть чистое излучение (тензор энергии-импульса типа II, разд. 4.3), распространяющееся в направлении вектора **K**, касательного к геодезической. В пустом пространстве это условие выполняется, если рассматриваемая изотропная геодезическая содержит точку, в которой тензор Вейля отличен от нуля, а **K** не принадлежит ни к одному из направлений, для которых $K^cK^dK_{[a}C_{b]cd[e}K_{f]} = 0$ (таких направлений самое большее четыре). Поэтому представляется разумным предположить, что в физически реалистическом решении уравнений Эйнштейна каждая времениподобная или изотропная геодезическая содержит точку, в которой $K^aK^bK_{[c}R_{d]ab[e}K_{f]}$ отлично от нуля. Будем говорить, что пространство-время, в котором выполняется это условие, удовлетворяет *типовому условию* (generic condition).

Подобным же образом мы можем рассмотреть изотропные геодезические, ортогональные к пространственноподобной 2-поверхности $\mathscr{S}$. Под *пространственноподобной 2-поверхностью* $\mathscr{S}$ мы имеем в виду вложенное двумерное подмногообразие, определяемое локально уравнениями $f_1 = 0$, $f_2 = 0$, где f_1 и f_2 суть такие O^2-функции, что при $f_1 = 0$, $f_2 = 0$ векторы $f_{1;a}$ и $f_{2;a}$ не равны нулю, не параллельны и

$$(f_{1;a} + \mu f_{2;a})(f_{1;b} + \mu f_{2;b})\,g^{ab} = 0$$

для двух разных действительных значений $\mu = \mu_1$ и $\mu = \mu_2$. Тогда каждый вектор, лежащий на этой двумерной поверхности, обязательно пространственноподобен. В качестве изотропных

векторов, нормальных к $\mathscr{S}$, введем $N_1{}^a$ и $N_2{}^a$, пропорциональные соответственно $g^{ab}(f_{1;b}+\mu_1 f_{2;b})$ и $g^{ab}(f_{1;b}+\mu_2 f_{2;b})$, и затем нормируем их так, чтобы

$$N_1{}^a N_2{}^b g_{ab} = -1.$$

Введя еще два единичных пространственноподобных вектора $Y_1{}^a$ и $Y_2{}^a$, ортогональных друг к другу и к $N_1{}^a$ и $N_2{}^a$, можно получить псевдоортогональный базис. Определим два вторых фундаментальных тензора на $\mathscr{S}$:

$${}_n\chi_{ab} = -N_{nc;d}(Y_1{}^c Y_{1a}+Y_2{}^c Y_{2a})(Y_1{}^d Y_{1b}+Y_2{}^d Y_{2b}),$$

где $n = 1, 2$; ${}_1\chi_{ab}$ и ${}_2\chi_{ab}$ симметричны.

Существуют два семейства изотропных геодезических, нормальных к $\mathscr{S}$, соответственно двум изотропным нормалям $N_1{}^a$ и $N_2{}^a$. Рассмотрим то семейство, касательный вектор которого $\mathbf{K}$ на $\mathscr{S}$ равен $\mathbf{N}_2$. Мы можем фиксировать наш псевдоортогональный базис $\mathbf{E}_1 = \mathbf{N}_2$, $\mathbf{E}_2$, $\mathbf{E}_3$, $\mathbf{E}_4$, приняв, что на $\mathscr{S}$ $\mathbf{E}_1 = \mathbf{Y}_1$, $\mathbf{E}_2 = \mathbf{Y}_2$, $\mathbf{E}_3 = \mathbf{N}_1$, $\mathbf{E}_4 = \mathbf{N}_2$, и перенеся его параллельно вдоль изотропных геодезических. Проекция на пространство S_q вектора девиации $\mathbf{Z}$, описывающего взаимное разделение соседних изотропных геодезических относительно изотропной геодезической $\gamma(v)$, будет удовлетворять уравнению (4.30) и начальным условиям

$$\frac{\mathrm{d}}{\mathrm{d}v} Z^m = {}_2\chi_{mn} Z^n \tag{4.47}$$

в точке $q \in \mathscr{S} \cap \gamma(v)$. Как и прежде, вращение этих полей будет равно нулю; начальное значение расхождения $\hat{\theta}$ равно ${}_2\chi_{ab}g^{ab}$. Аналогично предложению 4.4.3 имеем

Предложение 4.4.6

Если всюду $R_{ab}K^aK^b \geqslant 0$ и ${}_2\chi_{ab}g^{ab} < 0$, то существует точка, сопряженная $\mathscr{S}$ вдоль $\gamma(v)$, в пределах афинного расстояния $2/(-{}_2\chi_{ab}g^{ab})$ от $\mathscr{S}$. □

В силу определения сопряженных точек их существование означает наличие в семействах геодезических самопересечений или каустик. Дальнейшее обсуждение роли сопряженных точек будет дано в следующем разделе.

4.5. Вариация длины дуги

В этом разделе мы рассмотрим времениподобные и непространственноподобные кривые, которые кусочно класса C^3, но могут иметь точки, где их касательные векторы терпят разрыв. Мы будем требовать, чтобы в таких точках два касательных

вектора $(\partial/\partial t)|_-$ и $(\partial/\partial t)|_+$ удовлетворяли условию

$$g\left(\frac{\partial}{\partial t}\Big|_-, \frac{\partial}{\partial t}\Big|_+\right) = -1,$$

т. е. чтобы они были направлены внутрь одной и той же полости изотропного конуса.

Предложение 4.5.1

Пусть $\mathcal{U}$ — выпуклая координатная окрестность точки q. Тогда те точки в $\mathcal{U}$, которых можно достичь из q по времениподобным (соответственно непространственноподобным) кривым, имеют вид $\exp_q(\mathbf{X})$, $\mathbf{X} \in T_q$, где $g(\mathbf{X}, \mathbf{X}) < 0$ (соответственно $\leqslant 0$). (Здесь и в остальной части раздела мы полагаем, что отображение exp ограничено такой окрестностью начала в T_q, которая диффеоморфна $\mathcal{U}$ под действием $\exp_q$.)

Другими словами, изотропные геодезические, выходящие из q, образуют границу той области в $\mathcal{U}$, которую можно достичь из q по времениподобным или пространственноподобным кривым в $\mathcal{U}$. Интуитивно это утверждение достаточно очевидно, но поскольку оно является фундаментальным для понятия причинности, мы докажем его строго. Начнем с доказательства следующей леммы:

Лемма 4.5.2

В $\mathcal{U}$ времениподобные геодезические, проходящие через q, ортогональны к трехмерным поверхностям $\sigma = \text{const}$ $(\sigma < 0)$, где по определению при $p \in \mathcal{U}$ $\sigma(p) = g(\exp_q{}^{-1}p, \exp_q{}^{-1}p)$.

Доказательство опирается на тот факт, что вектор девиации точек равного расстояния вдоль соседних геодезических, первоначально ортогональный к этим геодезическим, остается ортогональным к ним. Точнее, пусть $\mathbf{X}(t)$ означает кривую в T_q, причем $g(\mathbf{X}(t), \mathbf{X}(t)) = -1$. Нужно показать, что соответствующие кривые $\lambda(t) = \exp_q(s_0\mathbf{X}(t))$ $(s_0 = \text{const})$ — там, где они определены, — ортогональны к времениподобным геодезическим $\gamma(s) = \exp_q(s\mathbf{X}(t_0))$ $(t_0 = \text{const})$. Следовательно, для двумерной поверхности α, определяемой $\alpha(s, t) = \exp_q(s\mathbf{X}(t))$, нужно доказать, что

$$g\left(\left(\frac{\partial}{\partial s}\right)_\alpha, \left(\frac{\partial}{\partial t}\right)_\alpha\right) = 0$$

(см. рис. 11). Имеем

$$\frac{\partial}{\partial s} g\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) = g\left(\frac{\mathrm{D}}{\partial s}\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) + g\left(\frac{\partial}{\partial s}, \frac{\mathrm{D}}{\partial s}\frac{\partial}{\partial t}\right).$$

Первый член в правой части равен нулю, потому что $\partial/\partial s$ — единичный касательный вектор к времениподобной геодезической, выходящей из q. Во втором члене согласно определению

производной Ли имеем

$$\frac{D}{\partial s}\frac{\partial}{\partial t}=\frac{D}{\partial t}\frac{\partial}{\partial s}.$$

Итак,

$$\frac{\partial}{\partial s}g\left(\frac{\partial}{\partial s},\ \frac{\partial}{\partial t}\right)=g\left(\frac{\partial}{\partial s},\ \frac{D}{\partial t}\frac{\partial}{\partial s}\right)=\frac{1}{2}\frac{\partial}{\partial t}g\left(\frac{\partial}{\partial s},\ \frac{\partial}{\partial s}\right)=0,$$

и, следовательно, $g(\partial/\partial s, \partial/\partial t)$ не зависит от s. Но при $s=0$ $(\partial/\partial t)_\alpha=0$; поэтому $g(\partial/\partial s,\ \partial/\partial t)$ тождественно равно нулю. □

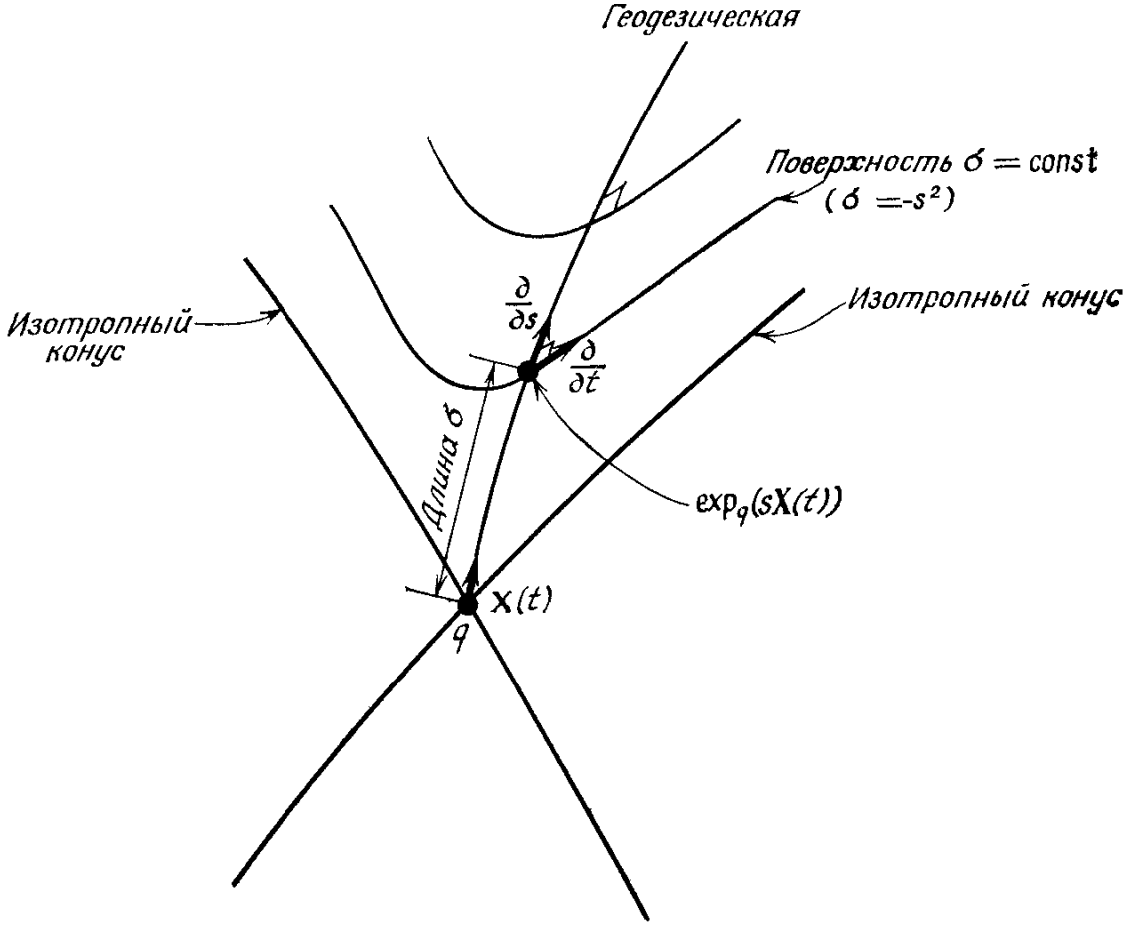


Рис. 11. В нормальной окрестности поверхности постоянного расстояния от q ортогональны к геодезическим, проходящим через q.

Доказательство предложения 4.5.1

Пусть C_q означает множество всех времениподобных векторов в q. Они заполняют внутренность конуса в T_q с вершиной в начале. Пусть $\gamma(t)$ — некоторая времениподобная кривая в $\mathcal{U}$ от q до p и пусть дана кусочно C^2-кривая в T_q: $\bar{\gamma}(t)=\exp_q^{-1}(\gamma(t))$. Тогда, отождествляя касательное пространство к T_q с самим T_q, имеем

$$(\partial/\partial t)_\gamma\big|_q=(\partial/\partial t)_{\bar{\gamma}}\big|_q.$$

Поэтому вектор $(\partial/\partial t)_{\bar\gamma}$ времениподобен в q. Отсюда видно, что кривая $\bar\gamma(t)$ входит в область C_q, но $\exp_q(C_q)$ — это та область $\mathcal{U}$, в которой $\sigma < 0$ и в которой, в силу предыдущей леммы, поверхности $\sigma = \text{const}$ пространственноподобны. Таким образом, σ должно монотонно убывать вдоль $\gamma(t)$, поскольку вектор $(\partial/\partial t)_\gamma$, будучи времениподобным, никогда не может быть касательным к поверхности $\sigma = \text{const}$ и поскольку в любой точке разрыва производных $\gamma(t)$ два касательных вектора направлены в одну и ту же полость изотропного конуса. Поэтому $p \in \exp_q(\bar C_q)$, что завершает доказательство для времениподобных кривых.

Чтобы доказать, что непространственноподобная кривая $\gamma(t)$ остается в $\exp_q(\bar C_q)$, рассмотрим такую малую вариацию $\gamma(t)$, которая переведет ее во времениподобную кривую. Пусть $\mathbf{Y}$ — векторное поле на T_q, такое, что индуцированное на $\mathcal{U}$ векторное поле $\exp_{q*}(\mathbf{Y})$ всюду времениподобно, и $g(\mathbf{Y}, (\partial/\partial t)_\gamma|_q) < 0$. Пусть при каждом $\varepsilon \geqslant 0$ $\beta(r, \varepsilon)$ есть кривая в T_q, которая выходит из начала, причем ее касательный вектор $(\partial/\partial r)_\beta = (\partial/\partial t)_{\bar\gamma}|_{t=r} + \varepsilon \mathbf{Y}|_{\beta(r,\varepsilon)}$. Тогда $\beta(r, \varepsilon)$ — дифференцируемая функция r и ε. Для каждого $\varepsilon > 0$ $\exp_q(\beta(r, \varepsilon))$ является времениподобной кривой на $\mathcal{U}$ и, таким образом, содержится в $\exp_q(C_q)$. Следовательно, непространственноподобная кривая $\exp_q(\beta(r, 0)) = \gamma(r)$ содержится в $\overline{\exp_q(C_q)} = \exp_q(\bar C_q)$. □

Следствие

Если $p \in \mathcal{U}$ можно достичь из q по непространственноподобной, но не по времениподобной кривой, то p лежит на изотропной геодезической, проходящей через q. □

Длина непространственноподобной кривой $\gamma(t)$ от q до p равна

$$L(\gamma, q, p) = \int_q^p \left[-g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right)\right]^{1/2} \mathrm{d}t,$$

где интеграл взят по дифференцируемым частям этой кривой.

При положительно определенной метрике можно искать кратчайшую кривую между двумя точками, но в лоренцевой метрике нет никакой кратчайшей кривой, поскольку любую кривую можно деформировать в изотропную, длина которой равна нулю. Однако при определенных условиях между двумя точками или между точкой и пространственноподобной трехмерной поверхностью найдется непространственноподобная кривая наибольшей длины. Сначала мы будем иметь дело с ситуацией, когда две точки близки друг к другу. Затем мы выведем необходимые условия для общего случая, когда две точки не обязательно близки. Достаточное условие для этого случая будет рассмотрено в разд. 6.7.

Предложение 4.5.3

Пусть q и p принадлежат нормальной выпуклой окрестности $\mathcal{U}$. Тогда, если можно соединить q и p непространственноподобной кривой в $\mathcal{U}$, то кривая наибольшей длины единственна в $\mathcal{U}$ и является непространственноподобной геодезической от q до p. Более того, если ввести функцию $\rho(q, p)$, равную длине этой кривой, когда она существует, и равную нулю, если такой кривой нет, то $\rho(q, p)$ непрерывна в $\mathcal{U} \times \mathcal{U}$.

По определению выпуклой нормальной окрестности (разд. 2.5) в $\mathcal{U}$ существует единственная геодезическая $\gamma(t)$, для которой $\gamma(0) = q$, $\gamma(1) = p$. Поскольку эта геодезическая есть дифференцируемая функция своих концов, интеграл

$$\sigma(q, p) = \int_0^1 g\left(\left(\frac{\partial}{\partial t}\right)_\gamma, \left(\frac{\partial}{\partial t}\right)_\gamma\right) \mathrm{d}t$$

будет дифференцируемой функцией на $\mathcal{U} \times \mathcal{U}$. (Это та же функция σ, что и в лемме 4.5.2.) Таким образом, функция $\rho(q, p)$ непрерывна в $\mathcal{U} \times \mathcal{U}$, поскольку она равна $[-\sigma(q, p)]^{1/2}$ при $\sigma < 0$ и нулю в иных случаях. Остается теперь показать, что в случае существования в $\mathcal{U}$ времениподобных кривых, соединяющих q и p, времениподобная геодезическая между ними имеет наибольшую длину. Пусть, как прежде, $\alpha(s, t) = \exp_q(s\mathbf{X}(t))$, причем $g(\mathbf{X}(t), \mathbf{X}(t)) = -1$. Если $\lambda(t)$ — времениподобная кривая в $\mathcal{U}$ от q до p, ее можно представить в виде $\lambda(t) = \alpha(f(t), t)$. Тогда

$$\left(\frac{\partial}{\partial t}\right)_\lambda = f'(t)\left(\frac{\partial}{\partial s}\right)_\alpha + \left(\frac{\partial}{\partial t}\right)_\alpha.$$

Поскольку два вектора в правой части взаимно ортогональны в силу леммы 4.5.2 и $g((\partial/\partial s)_\alpha, (\partial/\partial s)_\alpha) = -1$, то

$$g\left(\left(\frac{\partial}{\partial t}\right)_\lambda, \left(\frac{\partial}{\partial t}\right)_\lambda\right) = -(f'(t))^2 + g\left(\left(\frac{\partial}{\partial t}\right)_\alpha, \left(\frac{\partial}{\partial t}\right)_\alpha\right) \geqslant -(f'(t))^2,$$

причем равенство имеет место тогда и только тогда, когда $(\partial/\partial t)_\alpha = 0$, т. е. если и только если λ — геодезическая кривая. Итак,

$$L(\lambda, q, p) \leqslant \int_q^p f'(t)\, \mathrm{d}t = \rho(q, p),$$

причем равенство имеет место тогда и только тогда, когда λ — геодезическая кривая в $\mathcal{U}$ от q до p (она единственна). □

Теперь мы перейдем к случаю, когда q и p не обязательно принадлежат выпуклой нормальной окрестности $\mathcal{U}$. Необходимые условия для того, чтобы времениподобная кривая $\gamma(t)$, со-

единяющая q и p, была кривой наибольшей длины от q до p, мы выведем из рассмотрения малых вариаций. *Вариация* α кривой $\gamma(t)$ есть C^1-отображение $\alpha:(-\varepsilon, \varepsilon)\times[0, t_p]\to\mathscr{M}$ со следующими свойствами:

1) $\alpha(0, t)=\gamma(t)$;

2) существует такое подразбиение $0=t_1<t_2\ldots<t_n=t_p$ отрезка $[0, t_p]$, что α является C^3-отображением на каждом $(-\varepsilon, \varepsilon)\times[t_i, t_{i+1}]$;

3) $\alpha(u, 0)=q$, $\alpha(u, t_p)=p$;

4) для любой постоянной u $\alpha(u, t)$ — времениподобная кривая.

Вектор $\mathbf{Z}=(\partial/\partial u)_\alpha|_{u=0}$ будем называть *вектором вариации*. Обратно, если вдоль $\gamma(t)$ задано непрерывное векторное поле $\mathbf{Z}$ кусочно класса C^2, обращающееся в нуль в q и p, мы можем определить вариацию α, для которой $\mathbf{Z}$ будет вектором вариации:

$$\alpha(u, t)=\exp_r(u\mathbf{Z}|_r),$$

где $u\in(-\varepsilon, \varepsilon)$ при некотором $\varepsilon>0$ и $r=\gamma(t)$.

Лемма 4.5.4

Вариация расстояния от q до p при отображении α равна

$$\frac{\partial L}{\partial u}\Big|_{u=0}=\sum_{i=1}^{n-1}\int_{t_i}^{t_{i+1}} g\left(\frac{\partial}{\partial u},\ \left\{f^{-1}\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t}-f^{-2}\left(\frac{\partial f}{\partial t}\right)\frac{\partial}{\partial t}\right\}\right)\mathrm{d}t+$$

$$+\sum_{i=2}^{n-1} g\left(\frac{\partial}{\partial u},\ \left[f^{-1}\frac{\partial}{\partial t}\right]\right),$$

где $f^2=g(\partial/\partial t, \partial/\partial t)$ — квадрат касательного вектора, а $[f^{-1}\partial/\partial t]$ — разрыв в одной из сингулярных точек $\gamma(t)$.

Имеем

$$\begin{aligned}\frac{\partial L}{\partial u}\Big|_{u=0}&=\sum\frac{\partial}{\partial u}\int\left(-g\left(\frac{\partial}{\partial t},\frac{\partial}{\partial t}\right)\right)^{1/2}\mathrm{d}t=\\&=-\sum\int g\left(\frac{\mathrm{D}}{\partial u}\frac{\partial}{\partial t},\frac{\partial}{\partial t}\right)f^{-1}\mathrm{d}t=\\&=-\sum\int g\left(\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial u},\frac{\partial}{\partial t}\right)f^{-1}\mathrm{d}t=\\&=-\sum\int\left\{\frac{\partial}{\partial t}\left(g\left(\frac{\partial}{\partial u},\frac{\partial}{\partial t}\right)\right)f^{-1}-g\left(\frac{\partial}{\partial u},\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t}\right)f^{-1}\right\}\mathrm{d}t.\end{aligned}$$

Интегрируя по частям, получаем нужную формулу. □

Эту формулу можно упростить, выбирая в качестве параметра t длину дуги s. Тогда $g(\partial/\partial t, \partial/\partial t)=-1$. Обозначив через $\mathbf{V}$

единичный касательный вектор $\partial/\partial s$, имеем

$$\frac{\partial L}{\partial u}\Big|_{u=0} = \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} g(\mathbf{Z}, \dot{\mathbf{V}})\,ds + \sum_{i=2}^{n-1} g(\mathbf{Z}, [\mathbf{V}]),$$

где $\dot{\mathbf{V}} = D\mathbf{V}/\partial s$ — ускорение. Отсюда следует необходимое условие того, чтобы $\gamma(t)$ была кривой наибольшей длины от q до p: она *должна быть геодезической кривой без изломов*, в противном случае можно выбрать вариацию, которая даст более длинную кривую.

Можно также рассмотреть времениподобную кривую $\gamma(t)$ от некоторой пространственноподобной 3-поверхности $\mathcal{H}$ до точки p. Вариацию α этой кривой определим, как прежде, за тем исключением, что свойство (3) заменим на следующее:

3) $\alpha(u, 0)$ лежит на $\mathcal{H}$, $\alpha(u, t_p) = p$.

Отсюда следует, что на поверхности $\mathcal{H}$ вектор вариации $\mathbf{Z} = \partial/\partial u$ лежит в $\mathcal{H}$.

Лемма 4.5.5

$$\frac{\partial L}{\partial u}\Big|_{u=0} = \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} g(\dot{\mathbf{V}}, \mathbf{Z})\,ds + \sum_{i=2}^{n-1} g(\mathbf{Z}, [\mathbf{V}]) + g(\mathbf{Z}, \mathbf{V})|_{s=0}.$$

Она доказывается так же, как и лемма 4.5.4. □

Отсюда следует необходимое условие для того, чтобы $\gamma(t)$ была кривой наибольшой длины от $\mathcal{H}$ до p: она должна быть *геодезической кривой без изломов, ортогональной к* $\mathcal{H}$.

Путем варьирования α мы убедились, что первая производная длины времениподобной геодезической равна нулю. Следующим нашим шагом будет вычисление второй производной. Определим двухпараметрическую вариацию α геодезической линии $\gamma(t)$, соединяющей q и p, как C^1-отображение

$$\alpha\colon (-\varepsilon_1, \varepsilon_1) \times (-\varepsilon_2, \varepsilon_2) \times [0, t_p] \to \mathcal{M}$$

с такими свойствами:

1) $\alpha(0, 0, t) = \gamma(t)$;

2) существует такое подразбиение $0 = t_1 < t_2 < \ldots < t_n = t_p$ отрезка $[0, t_p]$, что α является C^3-отображением на каждом $(-\varepsilon_1, \varepsilon_1) \times (-\varepsilon_2, \varepsilon_2) \times [t_i, t_{i+1}]$;

3) $\alpha(u_1, u_2, 0) = q$, $\alpha(u_1, u_2, t_p) = p$;

4) для любых постоянных u_1, u_2 $\alpha(u_1, u_2, t)$ является времениподобной кривой.

В качестве двух векторов вариации мы вводим

$$\mathbf{Z}_1 = \left(\frac{\partial}{\partial u_1}\right)_\alpha\Big|_{\substack{u_1=0\\u_2=0}}, \quad \mathbf{Z}_2 = \left(\frac{\partial}{\partial u_2}\right)_\alpha\Big|_{\substack{u_1=0\\u_2=0}}.$$

Обратно, если вдоль $\gamma(t)$ заданы два непрерывных векторных поля $\mathbf{Z}_1$ и $\mathbf{Z}_2$, кусочно класса C^2, можно определить вариацию, для которой они будут векторами вариации:

$$\alpha(u_1, u_2, t) = \exp_r(u_1\mathbf{Z}_1 + u_2\mathbf{Z}_2),$$
$$r = \gamma(t).$$

Лемма 4.5.6

При двухпараметрической вариации геодезической кривой $\gamma(t)$ вторая производная длина равна *)

$$\frac{\partial^2 L}{\partial u_2\,\partial u_1}\Big|_{\substack{u_1=0\\u_2=0}} = \sum_{i=1}^{n-1}\int_{t_i}^{t_{i+1}} g\left(\mathbf{Z}_1, \left\{\frac{\mathrm{D}^2}{\partial s^2}(\mathbf{Z}_2 + g(\mathbf{V}, \mathbf{Z}_2)\mathbf{V}) - \right.\right.$$

$$\left.\left. - \mathbf{R}(\mathbf{V}, \mathbf{Z}_2)\mathbf{V}\right\}\right)\mathrm{d}s + \sum_{i=2}^{n-1} g\left(\mathbf{Z}_1, \left[\frac{\mathrm{D}}{\partial s}(\mathbf{Z}_2 + g(\mathbf{V}, \mathbf{Z}_2)\mathbf{V})\right]\right).$$

Согласно лемме 4.5.4, имеем

$$\frac{\partial L}{\partial u_1}\Big|_{\substack{u_1=0\\u_2=0}} = \sum\int g\left(\frac{\partial}{\partial u_1}, \left\{f^{-1}\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t} - f^2\left(\frac{\partial f}{\partial t}\right)\frac{\partial}{\partial t}\right\}\right)\mathrm{d}t +$$

$$+ \sum g\left(\frac{\partial}{\partial u_1}, \left[f^{-1}\frac{\partial}{\partial t}\right]\right).$$

Отсюда

$$\frac{\partial^2 L}{\partial u_2\,\partial u_1}\Big|_{\substack{u_1=0\\u_2=0}} = \sum\int g\left(\frac{\mathrm{D}}{\partial u_2}\frac{\partial}{\partial u_1}, \left\{f^{-1}\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t} - f^{-2}\left(\frac{\partial f}{\partial t}\right)\frac{\partial}{\partial t}\right\}\right)\mathrm{d}t -$$

$$- \sum\int g\left(\frac{\partial}{\partial u_1}, \left\{f^{-2}\left(\frac{\partial f}{\partial u_2}\right)\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t} - f^{-1}\frac{\mathrm{D}}{\partial u_2}\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t} - \right.\right.$$

$$\left.\left. - 2f^{-3}\left(\frac{\partial f}{\partial u_2}\right)\left(\frac{\partial f}{\partial t}\right)\frac{\partial}{\partial t} + f^{-2}\left(\frac{\partial^2 f}{\partial u_2\,\partial t}\right)\frac{\partial}{\partial t} + f^{-2}\left(\frac{\partial f}{\partial t}\right)\frac{\mathrm{D}}{\partial u_2}\frac{\partial}{\partial t}\right\}\right)\mathrm{d}t +$$

$$+ \sum g\left(\frac{\mathrm{D}}{\partial u_2}\frac{\partial}{\partial u_1}, \left[f^{-1}\frac{\partial}{\partial t}\right]\right) + \sum g\left(\frac{\partial}{\partial u_1}, \frac{\mathrm{D}}{\partial u_2}\left[f^{-1}\frac{\partial}{\partial t}\right]\right).$$

Первый и третий члены равны нулю, так как $\gamma(t)$ — геодезическая кривая без изломов. Во втором члене можно воспольз-

*) Напомним, что векторное поле $\mathbf{R}(\mathbf{X}, \mathbf{Y})\mathbf{Z}$ определяется формулой (2.18). — *Прим. перев.*

зоваться тем, что

$$\frac{D}{\partial u_2}\frac{D}{\partial t}\frac{\partial}{\partial t} = -\mathbf{R}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial u_2}\right)\frac{\partial}{\partial t} + \frac{D}{\partial t}\frac{D}{\partial u_2}\frac{\partial}{\partial t} =$$

$$= -\mathbf{R}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial u_2}\right)\frac{\partial}{\partial t} + \frac{D^2}{\partial t^2}\frac{\partial}{\partial u_2},$$

$$\frac{\partial^2 f}{\partial u_2\,\partial t} = -\frac{\partial}{\partial t}\left(f^{-1}g\left(\frac{D}{\partial u_2}\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right)\right) =$$

$$= -\frac{\partial}{\partial t}\left\{f^{-1}\frac{\partial}{\partial t}\left(g\left(\frac{\partial}{\partial u_2}, \frac{\partial}{\partial t}\right)\right) - f^{-1}g\left(\frac{\partial}{\partial u_2}, \frac{D}{\partial t}\frac{\partial}{\partial t}\right)\right\}.$$

В четвертом члене

$$\frac{D}{\partial u_2}\left[f^{-1}\frac{\partial}{\partial t}\right] = \left[f^{-1}\frac{D}{\partial t}\frac{\partial}{\partial u_2} + f^{-3}g\left(\frac{D}{\partial t}\frac{\partial}{\partial u_2}, \frac{\partial}{\partial t}\right)\frac{\partial}{\partial t}\right].$$

Затем, выбирая в качестве t длину дуги s, получим требуемый результат. □

Из определения второй производной следует, что она симметрична по двум векторам вариации $\mathbf{Z}_1$ и $\mathbf{Z}_2$, хотя это и не видно сразу из полученного выражения. Мы видим лишь, что оно зависит от проекций $\mathbf{Z}_1$ и $\mathbf{Z}_2$ в пространство, ортогональное к $\mathbf{V}$. Следовательно, мы можем ограничиться рассмотрением вариаций, для которых $\mathbf{Z}_1$ и $\mathbf{Z}_2$ ортогональны к $\mathbf{V}$. Пусть T_γ — (бесконечномерное) векторное пространство, состоящее из всех непрерывных, векторных полей кусочно класса C^2 вдоль $\gamma(t)$, которые ортогональны к $\mathbf{V}$ и равны нулю в q и p. Тогда $\partial^2 L/\partial u_2 \partial u_1$ будет симметричным отображением $T_\gamma \times T_\gamma$ на R^1. Его можно рассматривать как симметричный тензор в T_γ и записать в виде

$$L(\mathbf{Z}_1, \mathbf{Z}_2) = \frac{\partial^2 L}{\partial u_2\,\partial u_1}\Big|_{\substack{u_1=0\\u_2=0}}, \quad \mathbf{Z}_1, \mathbf{Z}_2 \in T_\gamma.$$

Мы можем вычислить также вторую производную длины от $\mathcal{H}$ до p геодезической кривой $\gamma(t)$, нормальной к $\mathcal{H}$. Воспользуемся прежним методом, только теперь один из концов кривой $\gamma(t)$ не закреплен, а может перемещаться по $\mathcal{H}$.

Лемма 4.5.7

Вторая производная длины $\gamma(t)$ от $\mathcal{H}$ до p равна

$$\frac{\partial^2 L}{\partial u_2\,\partial u_1}\Big|_{\substack{u_1=0\\u_2=0}} = \sum_{i=1}^{n-1}\int_{t_i}^{t_{i+1}} g\left(\mathbf{Z}_1, \left\{\frac{D^2}{ds^2}\mathbf{Z}_2 - \mathbf{R}(\mathbf{V}, \mathbf{Z}_2)\mathbf{V}\right\}\right)ds +$$

$$+ \sum_{i=2}^{n-1} g\left(\mathbf{Z}_1, \left[\frac{D}{\partial s}\mathbf{Z}_2\right]\right) + g\left(\mathbf{Z}_1, \frac{D}{\partial s}\mathbf{Z}_2\right)\Big|_{\mathcal{H}} - \chi(\mathbf{Z}_1, \mathbf{Z}_2)|_{\mathcal{H}},$$

где $\mathbf{Z}_1$ и $\mathbf{Z}_2$ взяты ортогональными к $\mathbf{V}$, а $\chi(\mathbf{Z}_1, \mathbf{Z}_2)$ есть второй фундаментальный тензор на $\mathscr{H}$.

Первые два члена такие же, как в лемме 4.5.6. Добавочные члены имеют вид

$$\frac{\mathrm{D}}{\partial u_2} g\left(\frac{\partial}{\partial u_1}, f^{-1}\frac{\partial}{\partial t}\right)\Big|_{\mathscr{H}} = f^{-1} g\left(\frac{\mathrm{D}}{\partial u_2}\frac{\partial}{\partial u_1}, \frac{\partial}{\partial t}\right)\Big|_{\mathscr{H}} +$$

$$+ f^{-3} g\left(\frac{\mathrm{D}}{\partial u_2}\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) g\left(\frac{\partial}{\partial u_1}, \frac{\partial}{\partial t}\right)\Big|_{\mathscr{H}} + f^{-1} g\left(\frac{\partial}{\partial u_1}, \frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial u_2}\right)\Big|_{\mathscr{H}}.$$

Второй член обращается в нуль, поскольку $\partial/\partial u_1$ ортогонально к $\partial/\partial t$. Если принять, что t — длина дуги s, то $\partial/\partial t$ будет равно единичной нормали $\mathbf{N}$ к $\mathscr{H}$. Поскольку изменения одной конечной точки $\gamma(t)$ ограничены перемещениями по $\mathscr{H}$, $\partial/\partial u_1$ будет всегда ортогонален к $\mathbf{N}$. Таким образом,

$$g\left(\frac{\mathrm{D}}{\partial u_2}\frac{\partial}{\partial u_1}, \mathbf{N}\right) = \frac{\partial}{\partial u_2} g\left(\frac{\partial}{\partial u_1}, \mathbf{N}\right) - g\left(\frac{\partial}{\partial u_1}, \frac{\mathrm{D}}{\partial u_2}\mathbf{N}\right) =$$

$$= -\chi\left(\frac{\partial}{\partial u_1}, \frac{\partial}{\partial u_2}\right). \qquad \square$$

Будем говорить, что времениподобная геодезическая линия $\gamma(t)$, соединяющая q с p, *максимальна*, если функционал $L(\mathbf{Z}_1, \mathbf{Z}_2)$ отрицательно полуопределен. Иначе говоря, если кривая $\gamma(t)$ не максимальна, то существует малая вариация α, которая дает более длинную кривую от p до q. Аналогично, будем говорить, что времениподобная нормальная к $\mathscr{H}$ геодезическая кривая от $\mathscr{H}$ до p *максимальна*, если функционал $L(\mathbf{Z}_1, \mathbf{Z}_2)$ отрицательно полуопределен; следовательно, если кривая $\gamma(t)$ не максимальна, существует малая вариация, которая дает более длинную кривую от $\mathscr{H}$ до p.

Предложение 4.5.8

Времениподобная геодезическая линия $\gamma(t)$, соединяющая q с p, максимальна тогда и только тогда, когда в (q, p) нет точек, сопряженных q вдоль $\gamma(t)$.

Предположим, что в (q, p) нет сопряженных точек. Введем ортонормированный базис вдоль $\gamma(t)$, полученный переносом Ферми. Якобиевы поля вдоль $\gamma(t)$, равные нулю в q, можно охарактеризовать матрицей $A_{\alpha\beta}(t)$, которая неособенна в (q, p), но особенна в q (и, возможно, в p). Ввиду изолированности сопряженных точек величина $\mathrm{d}(\ln \det \mathbf{A})/\mathrm{d}s$ будет бесконечна там, где матрица $A_{\alpha\beta}$ особенна. Следовательно, векторное поле $\mathbf{Z} \in T_\gamma$ класса C^0 и кусочно класса C^2 можно представить в $[q, p]$ как

$$Z^\alpha = A_{\alpha\beta} W^\beta,$$

где W^β — векторное поле класса C^0 и кусочно класса C^1 в $[q, p]$. Тогда

$$L(\mathbf{Z}, \mathbf{Z}) = \sum \int_0^{s_p} A_{\alpha\beta} W^\beta \left\{ \frac{d^2}{ds^2} (A_{\alpha\delta} W^\delta) + R_{\alpha 4 \gamma 4} A_{\gamma\delta} W^\delta \right\} ds =$$

$$+ \sum A_{\alpha\beta} W^\beta \left[\frac{d}{ds} (A_{\alpha\delta} W^\delta) \right] =$$

$$= \lim_{\varepsilon \to 0+} \sum \int_\varepsilon^{s_p} A_{\alpha\beta} W^\beta \left\{ 2 \frac{d}{ds} A_{\alpha\delta} \frac{d}{ds} W^\delta + A_{\alpha\delta} \frac{d}{ds^2} W^\delta \right\} ds +$$

$$+ \sum A_{\alpha\beta} W^\beta A_{\alpha\delta} \left[\frac{d}{ds} W^\delta \right] =$$

$$= - \sum \int_0^{s_p} \left\{ A_{\alpha\beta} \frac{d}{ds} W^\beta A_{\alpha\delta} \frac{d}{ds} W^\delta + \right.$$

$$\left. + W^\beta \left(\frac{d}{ds} A_{\alpha\beta} A_{\alpha\delta} - A_{\alpha\beta} \frac{d}{ds} A_{\alpha\delta} \right) \frac{d}{ds} W^\delta \right\} ds.$$

(Мы ввели предельный переход, поскольку вторая производная W^δ в точке q не определена.) Однако

$$\left(\frac{d}{ds} A_{\alpha\beta} A_{\alpha\delta} - A_{\alpha\beta} \frac{d}{ds} A_{\alpha\delta} \right) = - 2 A_{\alpha\beta} \omega_{\alpha\gamma} A_{\gamma\delta} = 0.$$

Отсюда следует, что $L(\mathbf{Z}, \mathbf{Z}) \leqslant 0$.

Обратно, допустим, что существует точка $r \in (q, p)$, сопряженная q вдоль $\gamma(t)$. Пусть $\mathbf{W}$ — якобиево поле вдоль $\gamma(t)$, обращающееся в нуль в q и r, и пусть $\mathbf{K} \in T_\gamma$ — такой вектор, что

$$K^a g_{ab} \frac{D}{\partial s} W^b = -1 \text{ в точке } r.$$

Продолжим $\mathbf{W}$ до p, положив его равным нулю в $[r, p]$. Положим $\mathbf{Z} = \varepsilon \mathbf{K} + \varepsilon^{-1} \mathbf{K}$, где ε — некоторая постоянная. Тогда

$$L(\mathbf{Z}, \mathbf{Z}) = \varepsilon^2 L(\mathbf{K}, \mathbf{K}) + 2L(\mathbf{K}, \mathbf{W}) + 2\varepsilon^{-2} L(\mathbf{W}, \mathbf{W}) = \varepsilon^2 L(\mathbf{K}, \mathbf{K}) + 2.$$

Следовательно, взяв достаточно малое ε, мы можем сделать $L(\mathbf{Z}, \mathbf{Z})$ положительным. □

Аналогичный результат можно получить для времениподобной геодезической линии $\gamma(t)$ от $\mathscr{H}$ до p, ортогональной к $\mathscr{H}$:

Предложение 4.5.9

Времениподобная геодезическая линия $\gamma(t)$ от $\mathscr{H}$ до p максимальна тогда и только тогда, когда в $(\mathscr{H}, q)$ нет точек, сопряженных $\mathscr{H}$ вдоль γ. □

Исследуем теперь вариации непространственноподобных кривых от q до p. Выясним условия, при которых можно найти вариацию α кривой $\gamma(t)$, приводящую к условию $g(\partial/\partial t, \partial/\partial t) < 0$ всюду, иначе говоря, переводящую $\gamma(t)$ во времениподобную кривую от q до p. При вариации α

$$\frac{\partial}{\partial u}\left(g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right)\right) = 2g\left(\frac{\mathrm{D}}{\partial u}\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) = 2g\left(\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial u}, \frac{\partial}{\partial t}\right) =$$
$$= 2\frac{\partial}{\partial t}\left(g\left(\frac{\partial}{\partial u}, \frac{\partial}{\partial t}\right)\right) - 2g\left(\frac{\partial}{\partial u}, \frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t}\right). \quad (4.48)$$

Чтобы получить времениподобную кривую от q до p, требуется, чтобы это выражение было меньше или равно нулю на всей кривой $\gamma(t)$.

Предложение 4.5.10

Если p и q соединены непространственноподобной кривой $\gamma(t)$, которая не есть изотропная геодезическая, то их можно соединить времениподобной кривой.

Если $\gamma(t)$ не является изотропной геодезической от p до q, то должна быть точка, в которой касательный вектор имеет разрыв, или должен существовать открытый интервал, на котором вектор ускорения $(\mathrm{D}/\partial t)(\partial/\partial t)$ отличен от нуля и не параллелен $\partial/\partial t$. Рассмотрим сначала случай, когда нет разрывов. Имеем

$$g\left(\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) = \frac{1}{2}\frac{\partial}{\partial t}\left(g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right)\right) = 0.$$

Отсюда видно, что $(\mathrm{D}/\partial t)(\partial/\partial t)$ является пространственноподобным вектором там, где он отличен от нуля и не параллелен $\partial/\partial t$. Пусть $\mathbf{W}$ — времениподобное векторное C^2-поле вдоль $\gamma(t)$, причем $g(\mathbf{W}, \partial/\partial t) < 0$. Тогда мы получим времениподобную кривую от p до q, варьируя вектор вариации

$$\mathbf{Z} = x\mathbf{W} + y\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t},$$

где

$$x = c^{-1}e^{bt}\int_{t_q}^{t} e^{-bt}\left(1 - \frac{1}{2}ya^2\right)\mathrm{d}t,$$

$$a^2 = g\left(\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t}, \frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t}\right),$$

$$c = -g\left(\mathbf{W}, \frac{\partial}{\partial t}\right),$$

$$b = \max\left(-c^{-1}g\left(\mathbf{W}, \frac{\mathrm{D}}{\partial t}\right)\right) \quad \text{на} \quad [p, q],$$

а y — неотрицательная функция на $[p, q]$, такая, что $y_p = y_q = 0$ и

$$\int_{t_q}^{t_p} e^{-bt}\left(1 - \frac{1}{2}\, ya^2\right) \mathrm{d}t = 0.$$

Предположим теперь, что существует некоторое подразбиение $t_q < t_1 < t_2 < \ldots < t_p$, такое, что касательный вектор $\partial/\partial t$ непрерывен на каждом сегменте $[t_i, t_{i+1}]$. Если сегмент $[t_i, t_{i+1}]$ не является изотропной геодезической кривой, его можно варьировать так, чтобы получить между концами сегмента времениподобную кривую. Поэтому нужно только показать, что мы можем получить времениподобную кривую из непространственноподобной кривой, составленной из сегментов изотропных геодезических, касательные векторы которых не параллельны в точках разрыва $\gamma(t_i)$. В качестве t можно выбрать аффинный параметр на каждом сегменте $[t_i, t_{i+1}]$. Разрыв $[\partial/\partial t]|_{t_i}$ будет пространственноподобным вектором, будучи разностью двух непараллельных изотропных векторов на одной и той же половине изотропного конуса. Следовательно, вдоль $[t_{i-1}, t_{i+1}]$ можно построить такое векторное C^2-поле $\mathbf{W}$, для которого

$$g(\mathbf{W}, \partial/\partial t) < 0 \quad \text{на} \quad [t_{i-1}, t_i]$$

и

$$g(\mathbf{W}, \partial/\partial t) > 0 \quad \text{на} \quad [t_i, t_{i+1}].$$

Тогда времениподобную кривую между $\gamma(t_{i-1})$ и $\gamma(t_{i+1})$ получим, варьируя с вектором вариации $\mathbf{Z} = x\mathbf{W}$, где

$$x = c^{-1}(t_{i+1} - t_i)(t - t_{i-1})$$

при $t_{i-1} \leqslant t \leqslant t_i$ и

$$x = c^{-1}(t_i - t_{i-1})(t_{i+1} - t)$$

при $t_i \leqslant t \leqslant t_{i+1}$, где $c = -g(\mathbf{W}, \partial/\partial t)$. □

Итак, если $\gamma(t)$ не является геодезической кривой, то из нее вариацией можно получить времениподобную кривую. Если $\gamma(t)$ является геодезической кривой, то в качестве t можно взять аффинный параметр. Тогда видно, что необходимым (но не достаточным) условием получения в результате вариации времениподобной кривой является ортогональность вектора вариации $\partial/\partial u$ к касательному вектору $\partial/\partial t$ на всей $\gamma(t)$, ибо в противном случае производная $(\partial/\partial t)\, g(\partial/\partial u, \partial/\partial t)$ была бы положительна где-нибудь на $\gamma(t)$. Для такой вариации $(\partial/\partial u)\, g(\partial/\partial t, \partial/\partial t) = 0$ и надо исследовать вторую производную.

Рассмотрим поэтому двухпараметрическую вариацию α изотропной геодезической $\gamma(t)$, соединяющей q и p. Ее определение будет прежним, за тем исключением, что мы ограничимся (по

причинам, приведенным выше) варьированием с векторами вариации

$$\left.\frac{\partial}{\partial u_1}\right|_{\substack{u_1=0\\u_2=0}} \quad \text{и} \quad \left.\frac{\partial}{\partial u_2}\right|_{\substack{u_1=0\\u_2=0}},$$

ортогональными к касательному вектору $(\partial/\partial t)$ на $\gamma(t)$.

Исследовать поведение L при таких вариациях затруднительно ввиду того, что $(-g(\partial/\partial t;\, \partial/\partial t))^{1/2}$ — недифференцируемая функция, если $g(\partial/\partial t, \partial/\partial t) = 0$. Вместо этого мы рассмотрим вариацию

$$\Lambda \equiv -\sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} g\left(\frac{\partial}{\partial t},\, \frac{\partial}{\partial t}\right) \mathrm{d}t.$$

Ясно, что необходимым (но не достаточным) условием получения времениподобной кривой от q до p при вариации α является положительность Λ.

Имеем

$$\frac{1}{2}\frac{\partial^2}{\partial u_2\, \partial u_1}\left(g\left(\frac{\partial}{\partial t},\, \frac{\partial}{\partial t}\right)\right) =$$

$$= \frac{\partial^2}{\partial u_2\, \partial t}\left(g\left(\frac{\partial}{\partial u_1},\, \frac{\partial}{\partial t}\right)\right) - \frac{\partial}{\partial u_2}\left(g\left(\frac{\partial}{\partial u_1},\, \frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t}\right)\right) =$$

$$= \frac{\partial^2}{\partial u_2\, \partial t}\left(g\left(\frac{\partial}{\partial u_1},\, \frac{\partial}{\partial t}\right)\right) - g\left(\frac{\partial}{\partial u_1},\, \left\{\frac{\mathrm{D}^2}{\partial t^2}\frac{\partial}{\partial u_2} - \mathbf{R}\left(\frac{\partial}{\partial t},\, \frac{\partial}{\partial u_2}\right)\frac{\partial}{\partial t}\right\}\right),$$

откуда

$$\frac{1}{2}\left.\frac{\partial^2 \Lambda}{\partial u_2\, \partial u_1}\right|_{\substack{u_1=0\\u_2=0}} = \sum \int g\left(\frac{\partial}{\partial u_1},\, \left\{\frac{\mathrm{D}^2}{\partial t^2}\frac{\partial}{\partial u_2} - \mathbf{R}\left(\frac{\partial}{\partial t},\, \frac{\partial}{\partial u_2}\right)\frac{\partial}{\partial t}\right\}\right) \mathrm{d}t +$$

$$+ \sum g\left(\frac{\partial}{\partial u_1},\, \left[\frac{\mathrm{D}}{\partial t},\, \frac{\partial}{\partial u_2}\right]\right). \quad (4.49)$$

Эта формула очень похожа на формулу для вариации длины времениподобной кривой. Можно показать, что вариация Λ равна нулю, если вектор вариации пропорционален касательному вектору $\partial/\partial t$, поскольку он изотропен, а $\mathbf{R}(\partial/\partial t,\, d/\partial t)\,(\partial/\partial t) = 0$ в силу антисимметричности тензора Римана. Такая вариация эквивалентна перепараметризации $\gamma(t)$. Итак, если мы хотим получить вариацию, приводящую к времениподобной кривой, нужно рассматривать только проекции вектора вариации в пространство S_q в каждой точке q кривой $\gamma(t)$. Другими словами, при введении вдоль $\gamma(t)$ псевдоортонормированного базиса $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$, $\mathbf{E}_4$ с $\mathbf{E}_4 = \partial/\partial t$ вариация Λ будет зависеть только от компонент Z^m вектора вариации $(m = 1, 2)$.

Предложение 4.5.11

Если в (q, p) нет точек, сопряженных q вдоль $\gamma(t)$, то $d^2\Lambda/du^2|_{u=0} < 0$ для любой вариации α кривой $\gamma(t)$, вектор вариации которой $\partial/\partial u|_{u=0}$ ортогонален к касательному вектору $\partial/\partial t$ к $\gamma(t)$, не равен всюду нулю и не пропорционален всюду $\partial/\partial t$. Иначе говоря, если в (q, p) нет точек, сопряженных q вдоль $\gamma(t)$, то не существует малых вариаций, переводящих $\gamma(t)$ во времениподобную кривую от q до p.

Доказательство аналогично доказательству предложения 4.5.8 с использованием вместо $A_{\alpha\beta}$ матрицы 2×2 $\tilde{A}_{mn}$ из разд. 4.2. □

Предложение 4.5.12

Если в (q, p) есть точка r, сопряженная q вдоль $\gamma(t)$, то найдется вариация $\gamma(t)$, которая дает времениподобную кривую от q до p.

Доказательство предложения будет несколько громоздким из-за того, что необходимо показать времениподобность касательного вектора всюду. Пусть W^m — компоненты в пространстве S (см. разд. 5.2) якобиева поля, равного нулю в q и r. Оно удовлетворяет уравнению

$$\frac{d^2}{dt^2} W^m = - R_{m4n4} W^n,$$

где для удобства в качестве t взят аффинный параметр. Поскольку W является по крайней мере векторным полем класса C^3 и $(dW^m/dt) \neq 0$ в q и r, можно написать $W^m = f\hat{W}^m$, где $\hat{W}^m$ — единичный вектор, причем f и $\hat{W}$ — поля класса C^2. Тогда

$$\frac{d^2}{dt^2} f + hf = 0,$$

где

$$h = \hat{W}^m \frac{d^2}{dt^2} \hat{W}^m + R_{m4n4} \hat{W}^m \hat{W}^n.$$

Пусть точка $x \in [r, p]$ такова, что $W^m \neq 0$ в $[r, x]$ и h_1 есть минимальное значение h в $[r, x]$. Тогда, если $a > 0$ и $a^2 + h_1 > 0$ и если $b = \{-f(e^{at} - 1)^{-1}\}|_x$, поле

$$Z^m = \{b(e^{at} - 1) + f\} \hat{W}^m$$

равно нулю в q и x и удовлетворяет в (q, x) неравенству

$$Z^m \left(\frac{d^2}{dt^2} Z^m + R_{m4n4} Z^n \right) > 0.$$

Выберем вариацию $\alpha(u, t)$ кривой $\gamma(t)$ от q до x так, что компоненты ее вектора вариации $(\partial/\partial u)|_{u=0}$ в S равны Z^m и выпол-

няется условие:

$$g\left(\frac{\mathrm{D}}{\partial u}\frac{\partial}{\partial u}, \frac{\partial}{\partial t}\right)\Big|_{u=0} + g\left(\frac{\partial}{\partial u}, \frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial u}\right)\Big|_{u=0} =$$

$$= \begin{cases} -\varepsilon t & \text{при } 0 \leqslant t \leqslant \frac{1}{4}t_x, \\ \varepsilon\left(t - \frac{1}{2}t_x\right) & \text{при } \frac{1}{4}t_x \leqslant t \leqslant \frac{3}{4}t_x, \\ \varepsilon(t_x - t) & \text{при } \frac{3}{4}t_x \leqslant t \leqslant t_x, \end{cases}$$

где t_x — значение t в точке x и $\varepsilon > 0$, но меньше наименьшего значения величины $Z^m(\mathrm{d}^2Z^m/dt^2 + R_{m4n4}Z^n)$ в области $\frac{1}{4}t_x \leqslant t \leqslant \frac{3}{4}t_x$. Тогда, в силу (4.49), $(\partial^2/\partial u^2)g(\partial/\partial t, \partial/\partial t) < 0$ всюду в (q, x) и в этом интеравле достаточно малые u, α дают времениподобную геодезическую. Если мы присоединим эту кривую к куску $\gamma(t)$ от x до p, то получим непространственноподобную кривую от q до p, которая не будет изотропной геодезической кривой. Следовательно, существует вариация, которая приводит к времениподобной кривой, соединяющей q и p.

Подобным же методом можно доказать

Предложение 4.5.13

Если $\gamma(t)$ — изотропная геодезическая кривая 2-поверхности $\mathscr{S}$ до p, ортогональная к $\mathscr{S}$, и если в $(\mathscr{S}, p)$ нет точек, сопряженных $\mathscr{S}$ вдоль γ, то не существует малых вариаций γ, дающих времениподобную кривую от $\mathscr{S}$ до p.

Предложение 4.5.14

Если в $(\mathscr{S}, p)$ имеется точка, сопряженная $\mathscr{S}$ вдоль γ, то существует вариация γ, дающая времениподобную кривую от $\mathscr{S}$ до p.

Результаты, касающиеся вариаций времениподобных и непространственноподобных кривых, будут использованы в гл. 8 для доказательства отсутствия геодезических наибольшей длины.

Глава 5

Точные решения

Любое пространство-время можно в определенном смысле считать решением уравнений поля Эйнштейна

$$R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} = 8\pi T_{ab} \tag{5.1}$$

(мы пользуемся единицами, принятыми в гл. 3), поскольку, вычислив по метрическому тензору пространства-времени левую часть (5.1), мы можем *ввести* тензор энергии-импульса T_{ab} как правую часть (5.1). Определенный таким образом тензор материи будет, вообще говоря, обладать неприемлемыми физическими свойствами, тогда как решение уравнений (5.1) будет иметь смысл лишь при разумном характере распределения материи.

Под *точным решением* уравнений Эйнштейна мы будем подразумевать пространство-время $(\mathscr{M}, \mathbf{g})$, в котором уравнения поля удовлетворяются тензором энергии-импульса T_{ab} какой-то конкретной формы материи, причем выполняются постулат (а) («локальной причинности») гл. 3 и одно из энергетических условий (разд. 4.3). В частности, можно искать точные решения для пустого пространства ($T_{ab} = 0$), для электромагнитного поля [T_{ab} вида (3.7)], для идеальной жидкости [T_{ab} вида (3.8)] или для пространства, содержащего электромагнитное поле и идеальную жидкость. Ввиду сложности уравнений поля получить точные решения можно только в пространствах достаточно высокой симметрии. Кроме того, точные решения являются идеализацией в том смысле, что любая область пространства-времени, по-видимому, содержит разные формы материи, в то время как точные решения можно получить лишь при довольно простом материальном содержимом. Тем не менее точные решения служат источником идей о качественно новых явлениях, которые могут возникнуть в общей теории относительно и, таким образом, указывают на возможные свойства реалистических решений уравнений поля. Примеры, которые мы приведем, демонстрируют многие интересные для последующих глав типы поведения решений. Мы рассмотрим эти решения, обращая особое внимание на их глобальные свойства. Многие из этих свойств были обнаружены только недавно, хотя сами решения в локальной форме уже известны давно.

В разд. 5.1 и 5.2 мы рассматриваем простейшие лоренцевы метрики — те, для которых кривизна постоянна. Пространственно-однородные и изотропные модели описаны в разд. 5.3, а их простейшие анизотропные обобщения — в разд. 5.4. Во всех этих простых моделях мы видим наличие сингулярного начала, если только Λ не принимает больших положительных значений. Сферически-симметричные метрики, описывающие поле массивного заряженного или нейтрального тела, и аксиально-симметричные метрики, описывающие поле тяготения вне массивных вращающихся тел конкретного типа, рассмотрены соответственно в разд. 5.5 и 5.6. В них показано, что появление некоторых сингулярностей в явном выражении для метрики вызвано лишь неудачным выбором координаты. В разд. 5.7 дано описание модели Вселенной Геделя, а в разд. 5.8 — решение Тауба — Ньютона — Тамбурино — Унти (Тауба — НУТ). Последние модели, вероятно, не описывают реальную Вселенную, но они интересны своими патологическими глобальными свойствами. Наконец, ряд других интересных точных решений приведен в разд. 5.9.

5.1. Пространство-время Минковского

Пространство-время Минковского $(\mathscr{M}, \boldsymbol{\eta})$ является простейшим пустым пространством-временем в общей теории относительности и в действительности представляет собой пространство-время специальной теории относительности. Математически оно есть многообразие R^4 с плоской лоренцевой метрикой $\boldsymbol{\eta}$. В естественных координатах (x^1, x^2, x^3, x^4) на R^4 метрику $\boldsymbol{\eta}$ можно записать в виде

$$ds^2 = -(dx^4)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2. \tag{5.2}$$

Если используются сферические полярные координаты (t, r, θ, ϕ), где $x^4 = t$, $x^3 = r\cos\theta$, $x^2 = r\sin\theta\cos\phi$, $x^1 = r\sin\theta\sin\phi$, то метрика принимает вид

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{5.3}$$

Очевидно, эта метрика, сингулярна при $r = 0$ и $\sin\theta = 0$, однако это происходит из-за того, что используемые координаты не являются допустимыми в этих точках. Чтобы получить регулярные координатные окрестности, необходимо на координаты наложить ограничения, например $0 < r < \infty$, $0 < \theta < \pi$, $0 < \phi < 2\pi$. Для покрытия всего пространства Минковского нужны две такие координатные окрестности.

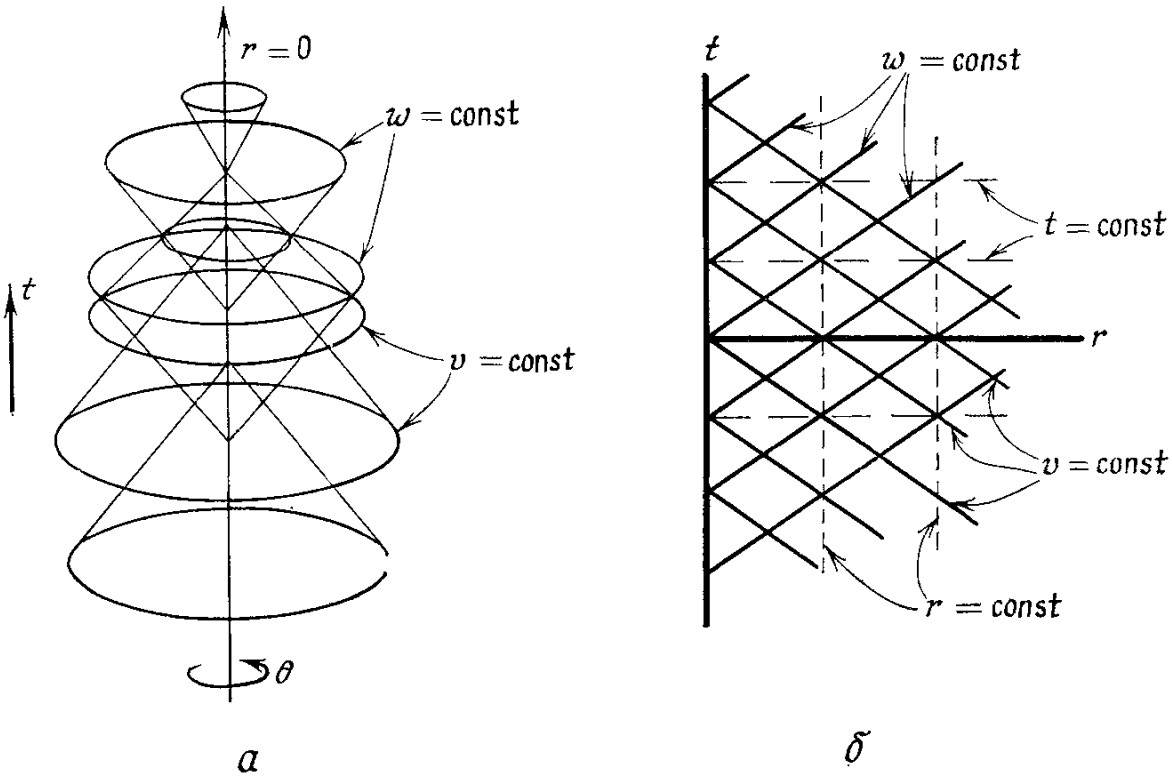
Другая координатная система получается при выборе опережающей и запаздывающей изотропных координат v, w:

$$v = t + r, \quad w = t - r \quad (\Rightarrow v \geqslant w).$$

Тогда метрика имеет вид

$$ds^2 = -dv\,dw + \frac{1}{4}(v-w)^2(d\theta^2 + \sin^2\theta\,d\phi^2), \qquad (5.4)$$

где $-\infty < v < \infty$, $-\infty < w < \infty$. Отсутствие в метрике членов dv^2, dw^2 отражает тот факт, что поверхности $\{w = \text{const}\}$,



а *б*

Рис. 12. Пространство Минковского. Изотропные координаты v (w) можно представить себе в виде сходящихся (расходящихся) сферических волн, распространяющихся со скоростью света; они являются опережающими (запаздывающими) временными координатами. Пересечение поверхности $\{v = \text{const}\}$ с поверхностью $\{w = \text{const}\}$ является 2-сферой. *а* — координатные поверхности v, w (одна координата опущена). *б* — плоскость (t, r); каждая точка изображает 2-сферу радиуса r

$\{v = \text{const}\}$ изотропны (т. е. $w_{;a}w_{;b}g^{ab} = 0 = v_{;a}v_{;b}g^{ab}$), см. рис. 12.

В системе координат, в которой метрика выражается в виде (5.2), геодезические имеют вид $x^a(v) = b^a v + c^a$, где b^a и c^a — постоянные. Таким образом, экспоненциальное отображение $\exp_p: T_p \to \mathscr{M}$ определяется формулой

$$x^a(\exp_p \mathbf{X}) = X^a + x^a(p),$$

где X^a — компоненты $\mathbf{X}$ относительно координатного базиса $\{\partial/\partial x^a\}$ пространства T_p. Поскольку exp является взаимно однозначным отображением на все многообразие $\mathscr{M}$, оно представляет собой диффеоморфизм между T_p и $\mathscr{M}$. Следовательно, лю-

бые две точки $\mathscr{M}$ можно соединить единственной геодезической линией. Отображение exp определено всюду на T_p для всех p, и в силу этого пространство-время $(\mathscr{M}, \eta)$ геодезически полно.

Область Коши (Cauchy development) будущего, прошлого) $D^+(\mathscr{S})$ (соответственно $D^-(\mathscr{S})$) пространственноподобной гиперповерхности $\mathscr{S}$ определяется как множество всех точек в $\mathscr{M}$, ко-

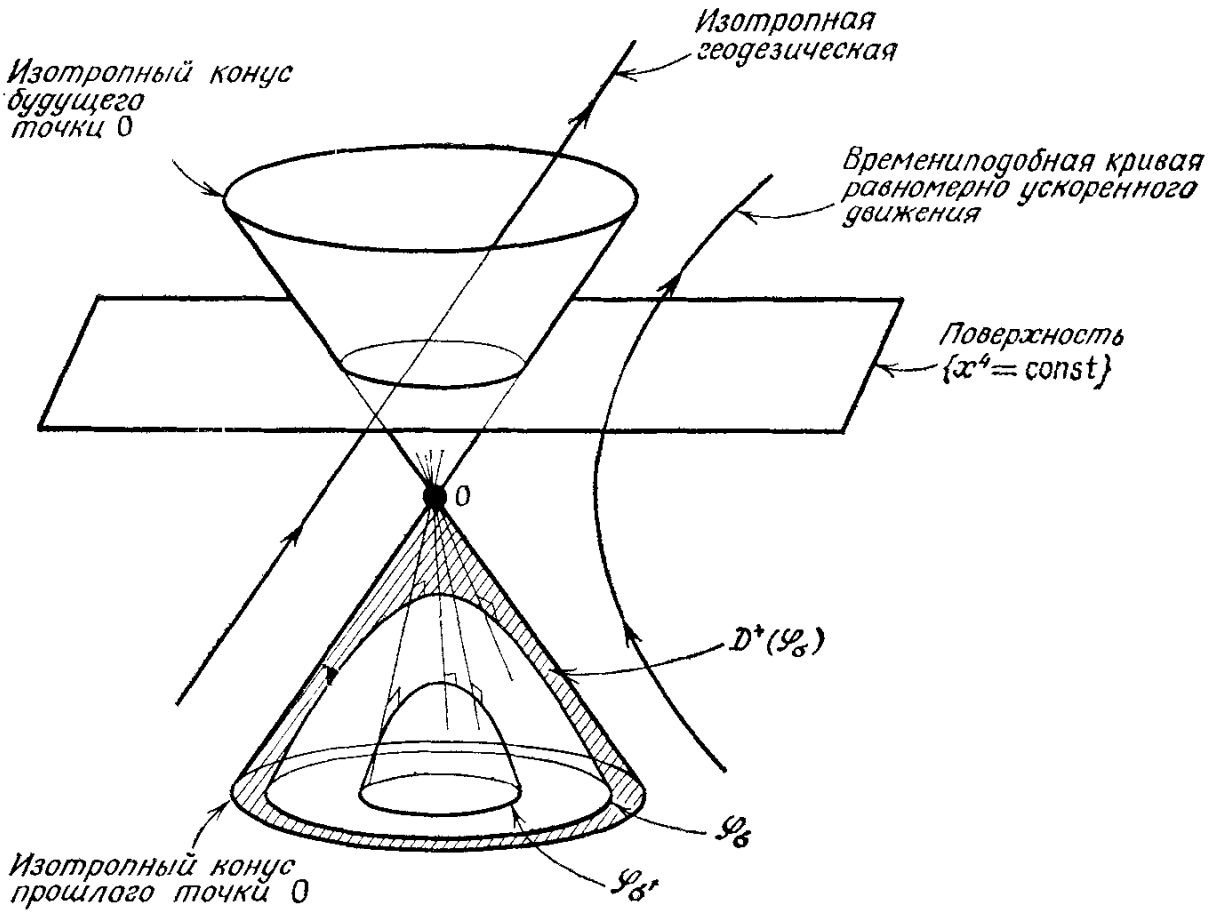


Рис. 13. Поверхность Коши $\{x^4 = \text{const}\}$ в пространстве-времени Минковского и пространственноподобные поверхности $\mathscr{S}_\sigma$, $\mathscr{S}_{\sigma'}$, которые не являются поверхностями Коши. Все нормальные к поверхностям $\mathscr{S}_\sigma$, $\mathscr{S}_{\sigma'}$ геодезические пересекаются в точке O.

торые обладают тем свойством, что все непродолжимые в прошлое (будущее) непространственноподобные кривые, проведенные из этих точек, пересекают $\mathscr{S}$ (ср. с разд. 6.5). Если $D^+(\mathscr{S}) \cup D^-(\mathscr{S}) = \mathscr{M}$, т. е. если каждая непродолжимая непространственноподобная кривая пересекает $\mathscr{S}$, то такую $\mathscr{S}$ будем называть *поверхностью Коши*. В пространстве-времени Минковского поверхности $\{x^4 = \text{const}\}$ образуют семейство поверхностей Коши, покрывающих $\mathscr{M}$ полностью. Можно, однако, построить нерасширяемые пространственноподобные поверхности, которые

не будут поверхностями Коши; например, поверхности

$$\mathscr{S}_\sigma\colon \{-(x^4)^2+(x^1)^2+(x^2)^2+(x^3)^2=\sigma=\text{const}\},$$

где $\sigma<0$, $x^4<0$, пространственноподобны, но лежат целиком внутри изотропного конуса прошлого начала O и потому не являются поверхностями Коши (см. рис. 13). В самом деле, область Коши будущего поверхности $\mathscr{S}_\sigma$ — это область, ограниченная $\mathscr{S}_\sigma$ и световым конусом прошлого точки O. В силу леммы 4.5.2, времениподобные геодезические ортогональны к поверхностям $\mathscr{S}_\sigma$. Если $r\in D^+(\mathscr{S}_\sigma)\cup D^-(\mathscr{S}_\sigma)$, то времениподобная геодезическая, проходящая через r и O, является времениподобной геодезической наибольшей длины между r и $\mathscr{S}_\sigma$. Если же r не лежит в $D^+(\mathscr{S}_\sigma)\cup D^-(\mathscr{S}_\sigma)$, то времениподобной кривой наибольшей длины между r и $\mathscr{S}_\sigma$ не существует, поскольку или r лежит в области $\sigma\geqslant 0$, и тогда вообще нет времениподобных геодезических, проходящих через r и ортогональных σ, или r лежит в области $\sigma<0$, $x^4\geqslant 0$, и тогда времениподобная геодезическая, проходящая через r и ортогональная к $\mathscr{S}_\sigma$, существует, но она не будет кривой наибольшей длины между r и $\mathscr{S}_\sigma$, ибо на ней имеется точка O, сопряженная к $\mathscr{S}_\sigma$ (ср. с рис. 13).

Для исследования структуры бесконечности в пространстве-времени Минковского воспользуемся интересным представлением этого пространства, принадлежащим Пенроузу. От изотропных координат v, w перейдем к новым изотропным координатам, в которых бесконечности v и w преобразованы к конечным значениям. Для этого введем соотношениями $\operatorname{tg} p=v$, $\operatorname{tg} q=w$ координаты p и q, причем $-\pi/2<p<\pi/2$, $-\pi/2<q<\pi/2$ (и $p\geqslant q$). Тогда метрика пространства $(\mathscr{M},\eta)$ принимает вид

$$ds^2=\sec^2 p\sec^2 q\left(-dp\,dq+\frac{1}{4}\sin^2(p-q)(d\theta^2+\sin^2\theta\,d\phi^2)\right),$$

т. е. физическая метрика конформна метрике $\bar{g}$, для которой

$$d\bar{s}^2=-4dp\,dq+\sin^2(p-q)(d\theta^2+\sin^2\theta\,d\phi^2). \tag{5.5}$$

Эту метрику можно привести к более привычному виду, если ввести

$$t'=p+q,\quad r'=p-q,$$

где

$$-\pi<t'+r'<\pi,\quad -\pi<t'-r'<\pi,\quad r'\geqslant 0; \tag{5.6}$$

тогда вместо (5.5) имеем

$$d\bar{s}^2=-(dt')^2+(dr')^2+\sin^2 r'(d\theta^2+\sin^2\theta\,d\phi^2). \tag{5.7}$$

Таким образом, все пространство-время Минковского изображается областью (5.6) метрики

$$ds^2=\frac{1}{4}\sec^2\left(\frac{1}{2}(t'+r')\right)\sec^2\left(\frac{1}{2}(t'-r')\right)d\bar{s}^2,$$

где $d\bar{s}$ определяется из (5.7); координаты (t, r) из (5.3) связаны с t', r' соотношениями

$$2t = \operatorname{tg}\left(\tfrac{1}{2}(t'+r')\right) + \operatorname{tg}\left(\tfrac{1}{2}(t'-r')\right),$$

$$2r = \operatorname{tg}\left(\tfrac{1}{2}(t'+r')\right) - \operatorname{tg}\left(\tfrac{1}{2}(t'-r')\right).$$

Метрика (5.7) локально совпадает с метрикой статической модели Эйнштейна (см. разд. 5.3), пространство-время которой полностью однородно *). Можно аналитически продолжить (5.7) на всю статическую модель Эйнштейна, т. е. можно продолжить координаты t', r' так, чтобы они покрывали все многообразие $R^1 \times S^3$, где $-\infty < t' < \infty$ и r', θ, ϕ рассматриваются как координаты на S^3 (с координатными сингулярностями при $r' = 0$, $r' = \pi$ и $\theta = 0$ и $\theta = \pi$, подобными тем, что были в (5.3); эти сингулярности устраняются переходом к другим локальным координатам в окрестности точек, где метрика (5.7) особенна). Опуская два измерения, можно представить статическую вселенную Эйнштейна в виде цилиндра $x^2 + y^2 = 1$, вложенного в трехмерное пространство Минковского с метрикой $ds^2 = -dt^2 + dx^2 + dy^2$ (полностью статическую модель Эйнштейна можно получить вложением цилиндра $x^2 + y^2 + z^2 + w^2 = 1$ в пятимерное евклидово пространство с метрикой $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 + dw^2$, ср. с [141]).

Итак, мы имеем следующую картину: все пространство-время Минковского конформно области (5.6) статической модели Эйнштейна, т. е. заштрихованной области на рис. 14. Можно считать, что граница этой области отражает конформную структуру бесконечности пространства-времени Минковского. Она состоит из изотропных поверхностей $p = \pi/2$ (обозначенной $\mathscr{I}^+$) и $q = -\pi/2$ (обозначенной $\mathscr{I}^-$) и включает в себя точки $p = \pi/2$, $q = \pi/2$ (обозначенную i^+), $p = \pi/2$, $q = -\pi/2$ (обозначенную i^0) и $p = -\pi/2$, $q = -\pi/2$ (обозначенную i^-). Любая направленная в будущее времениподобная геодезическая в пространстве Минковского достигает i^+ (i^-) при бесконечно большом положительном (отрицательном) значении своего аффинного параметра, и можно считать, что любая времениподобная геодезическая начинается в i^- и кончается в i^+ (ср. с рис. 15, *а*). Аналогично можно считать, что изотропные геодезические начинаются на $\mathscr{I}^-$ и кончаются на $\mathscr{I}^+$, в то время как пространственноподобные геодезические начинаются и кончаются в i^0. Таким образом, i^+ и i^- можно

*) Под однородностью пространства-времени здесь имеется в виду существование группы изометрий, преобразованиями которой каждую точку пространства-времени можно перевести в любую другую точку; обычно такая группа преобразований называется транзитивной (см. стр. 188). — *Прим. перев.*

рассматривать как изображения времениподобной бесконечности будущего и прошлого $\mathscr{I}^+$ и $\mathscr{I}^-$ — изотропной бесконечности будущего и прошлого и i^0 — пространственноподобной бесконечности. (Однако эти утверждения не относятся к негеодезическим

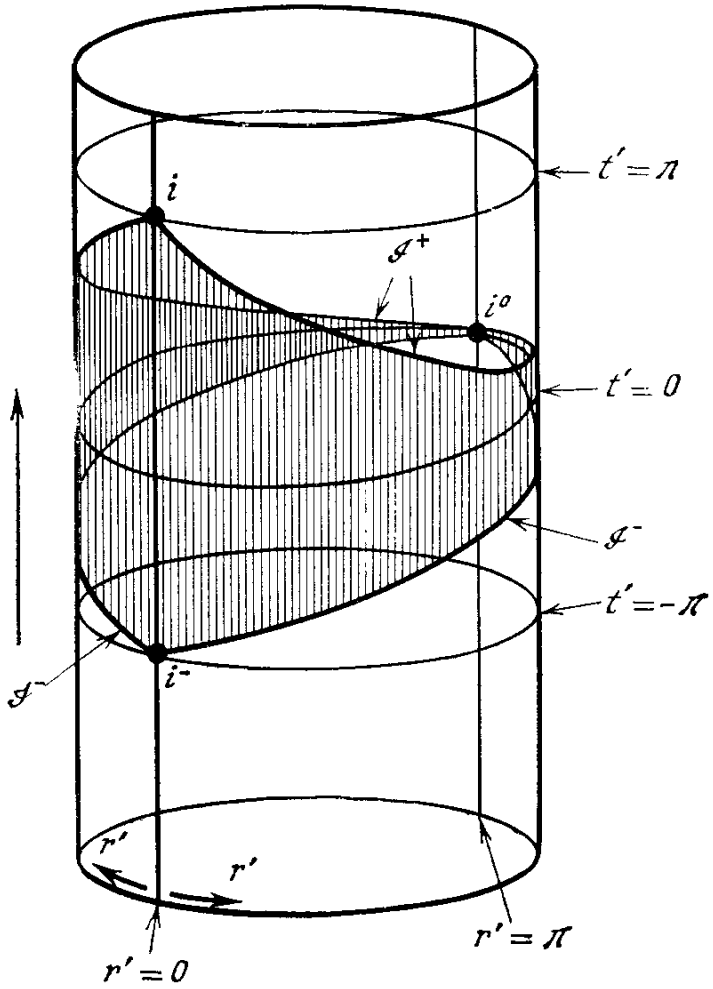


Рис. 14. Статическая вселенная Эйнштейна, представленная как вложенный цилиндр; координаты θ, ϕ опущены. Каждая точка изображает половину 2-сферы площадью $4\pi \sin^2 r'$. Заштрихованная область конформна всему пространству-времени Минковского; ее границу (часть изотропных конусов точек i^+, i^0, i^-) можно рассматривать как конформную бесконечность пространства-времени Минковского.

кривым: негеодезическая времениподобная кривая может начинаться, например, на $\mathscr{I}^-$ и кончаться на $\mathscr{I}^+$.) Поскольку любая поверхность Коши пересекает все времениподобные и изотропные геодезические, то она должна выглядеть как сечение всего пространства-времени, достигающее границы i^0. Конформную структуру бесконечности можно представить также, начертив диаграмму плоскости (t', r') (см. рис. 15, *б*). Как и на рис. 12, *а*, каждая точка этой диаграммы изображает сферу S^2, а радиальные изотропные геодезические изображаются прямыми, идущими под углом $\pm 45°$. Такого рода диаграммой может быть представлена

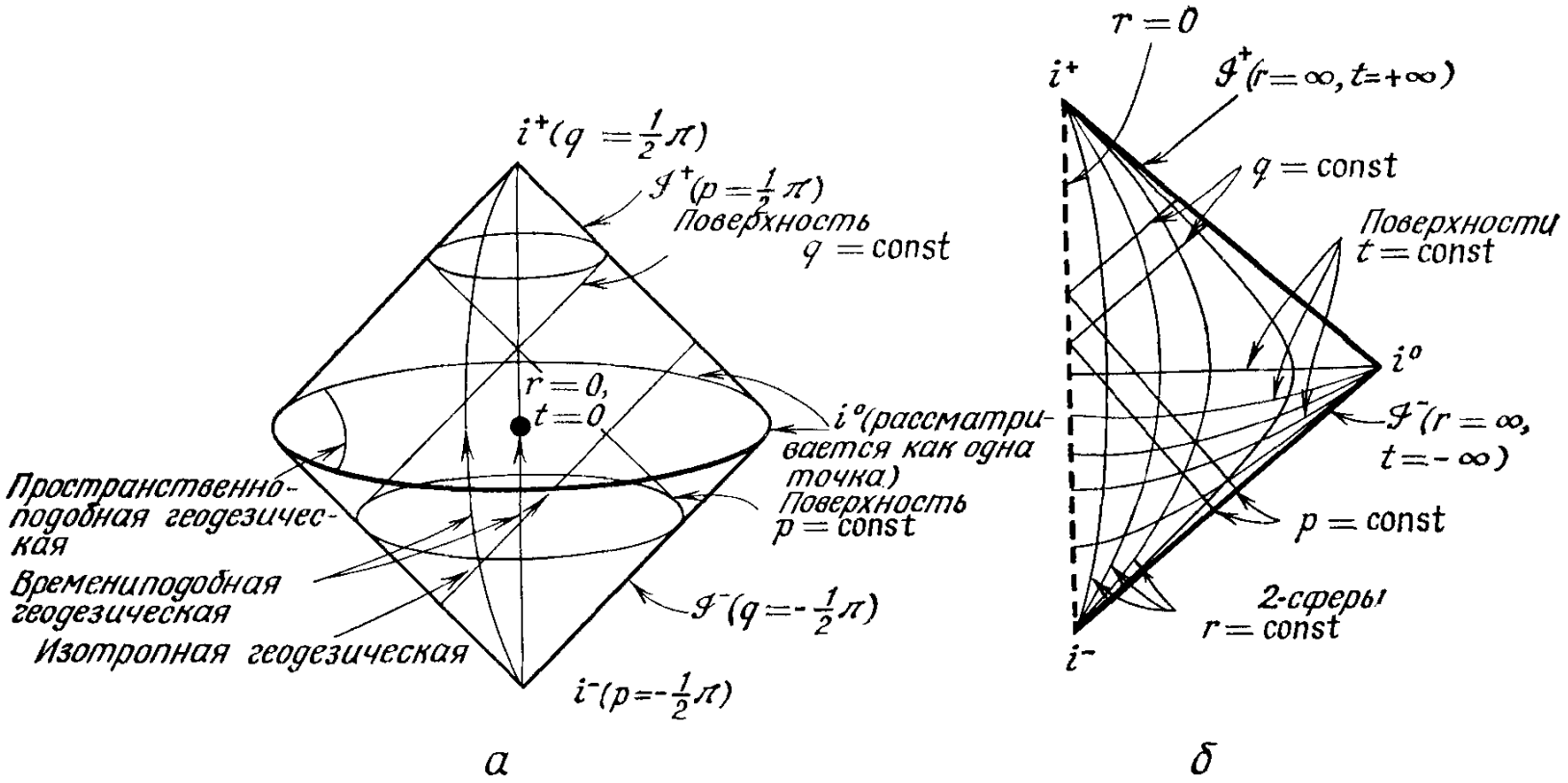


Рис. 15. *а* — заштрихованная область на рис. 14 (теперь опущена только одна координата), изображающая пространство-время Минковского и его конформную бесконечность; *б* — диаграмма Пенроуза пространства-времени Минковского. Каждая точка изображает 2-сферу, за исключением точек i^+, i^0, i^-, каждая из которых является единственной точкой, и точек на линии $r=0$ (где полярные координаты сингулярны).

структура бесконечности любого сферически симметричного пространства. На таких диаграммах бесконечность будет изображаться одной линией, начало полярных координат — пунктиром и неустранимые особенности метрики — двойной линией. Конформная структура пространства Минковского, описанная здесь, представляет собой то, что можно было бы считать «нормальным» поведением пространства-времени на бесконечности; другие типы поведения встретятся нам в следующих разделах.

Упомянем, наконец, что можно построить пространства, локально совпадающие с $(\mathscr{M}, \eta)$, но с иными (крупномасштабными) топологическими свойствами; это достигается отождествлением в $\mathscr{M}$ точек, которые эквивалентны относительно дискретной изометрии без фиксированной точки (например, отождествление точки (x^1, x^2, x^3, x^4) с точкой $(x^1, x^2, x^3, x^4 + c)$, где c — постоянная, меняет топологическую структуру R^4 на $R^3 \times S^1$ и приводит к появлению в пространстве-времени замкнутых времениподобных линий). Очевидно, $(\mathscr{M}, \eta)$ представляет собой универсальное накрывающее пространство для всех таких производных пространств; последние подробно рассмотрены в [3].

5.2. Пространство-время де Ситтера 1-го и 2-го рода

Пространственно-временные метрики постоянной кривизны локально характеризуются условием $R_{abcd} = (1/12)R(g_{ac}g_{bd} - g_{ad}g_{bc})$. Ему эквивалентно равенство $C_{abcd} = 0 = R_{ab} - (1/4)Rg_{ab}$, и, следовательно, тензор Римана определяется одним лишь скаляром Риччи R. Из свернутых тождеств Бианки сразу следует, что $R = \text{const}$ во всем пространстве-времени; действительно, эти пространства однородны. Тензор Эйнштейна имеет вид

$$R_{ab} - \frac{1}{2}Rg_{ab} = -\frac{1}{4}Rg_{ab};$$

поэтому эти пространства можно считать решениями уравнений поля в пустом пространстве с $\Lambda = R/4$ или решениями для идеальной жидкости с постоянной плотностью $R/32\pi$ и с постоянным давлением $-R/32\pi$. Однако последняя трактовка, по-видимому, неприемлема, поскольку тогда или давление, или плотность должны быть отрицательными; кроме того, уравнение движения (3.10) для такой жидкости становится неопределенным.

Пространство-время Минковского является пространством постоянной кривизны с $R = 0$. Пространство постоянной кривизны с $R > 0$ есть *пространство-время де Ситтера 1-го рода*, которое обладает топологией $R^1 \times S^3$ (см. интересное описание свойств этого пространства в [152], принадлежащее Шредингеру). Оно наиболее просто реализуется в виде гиперболоида

$$-v + w^2 + x^2 + y^2 + z^2 = \alpha^2$$

в плоском пятимерном пространстве R^5 с метрикой

$$-\mathrm{d}v^2+\mathrm{d}w^2+\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2=\mathrm{d}s^2$$

(рис. 16). На этом гиперболоиде можно ввести координаты (t, χ, θ, ϕ), определяемые соотношениями

$$\alpha\,\mathrm{sh}\,(\alpha^{-1}t)=v, \quad \alpha\,\mathrm{ch}\,(\alpha^{-1}t)\cos\chi=w,$$
$$\alpha\,\mathrm{ch}\,(\alpha^{-1}t)\sin\chi\cos\theta=x, \quad \alpha\,\mathrm{ch}\,(\alpha^{-1}t)\sin\chi\sin\theta\cos\phi=y,$$
$$\alpha\,\mathrm{ch}\,(\alpha^{-1}t)\sin\chi\sin\theta\sin\phi=z.$$

В этих координатах метрика имеет вид

$$\mathrm{d}s^2=-\mathrm{d}t^2+\alpha^2\,\mathrm{ch}^2\,(\alpha^{-1}t)\cdot\{\mathrm{d}\chi^2+\sin^2\chi\,(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\phi^2)\}.$$

Особенности метрики при $\chi=0$, $\chi=\pi$ и $\theta=0$, $\theta=\pi$ обусловлены лишь полярными координатами; за исключением этих тривиальных сингулярностей, координаты t, χ, θ, ϕ покрывают все пространство при $-\infty<t<\infty$, $0\leqslant\chi\leqslant\pi$, $0\leqslant\theta\leqslant\pi$, $0\leqslant\phi\leqslant2\pi$. Пространственными сечениями постоянного t являются сферы S^3 постоянной положительной кривизны, и они представляют собой поверхности Коши. Их геодезические нормали представляют собой линии, монотонно сходящиеся до некоторого минимального пространственного расстояния и затем снова расходящиеся до бесконечности (см. рис. 16, *а*).

На гиперболоиде можно ввести также координаты

$$\hat{t}=\alpha\ln\frac{w+v}{\alpha}, \quad \hat{x}=\frac{\alpha x}{w+v}, \quad \hat{y}=\frac{\alpha y}{w+v}, \quad \hat{z}=\frac{\alpha z}{w+v}.$$

В этих координатах метрика принимает вид

$$\mathrm{d}s^2=-\mathrm{d}\hat{t}^2+\exp\,(2\alpha^{-1}\hat{t})\,(\mathrm{d}\hat{x}^2+\mathrm{d}\hat{y}^2+\mathrm{d}\hat{z}^2),$$

однако они покрывают лишь половину гиперболоида, так как $\hat{t}$ не определены при $w+v\leqslant0$ (см. рис. 16, *б*). Область пространства де Ситтера 1-го рода, для которой $v+w>0$, образует пространство-время *стационарной модели* Вселенной, предложенной Бонди и Голдом [12] и Хойлом [80]. В этой модели предполагается, что материя движется по геодезическим нормалям к поверхностям $\{\hat{t}=\mathrm{const}\}$. Поскольку при этом происходит «разрежение» материи, то для поддержания постоянного значения плотности предполагается непрерывное рождение новой материи. Бонди и Голд не пытались вводить уравнения поля для этой модели, но Пирани [135] и Хойл и Нарликар [82] заметили, что метрику стационарной модели можно рассматривать как решение уравнений Эйнштейна (с $\Lambda=0$), если в дополнение к обычной материи ввести скалярное поле с отрицательной плотностью энергии. Это «*C*-поле» вызывало бы также и непрерывное рождение материи.

Стационарная модель обладает тем достоинством, что ее предсказания просты и определенны. Однако с нашей точки зрения у нее имеются два неудовлетворительных свойства. Первое — существование отрицательной энергии, о чем говорилось в разд. 4.3. Второе — возможность расширения пространства-времени, связанная с тем, что оно является только половиной

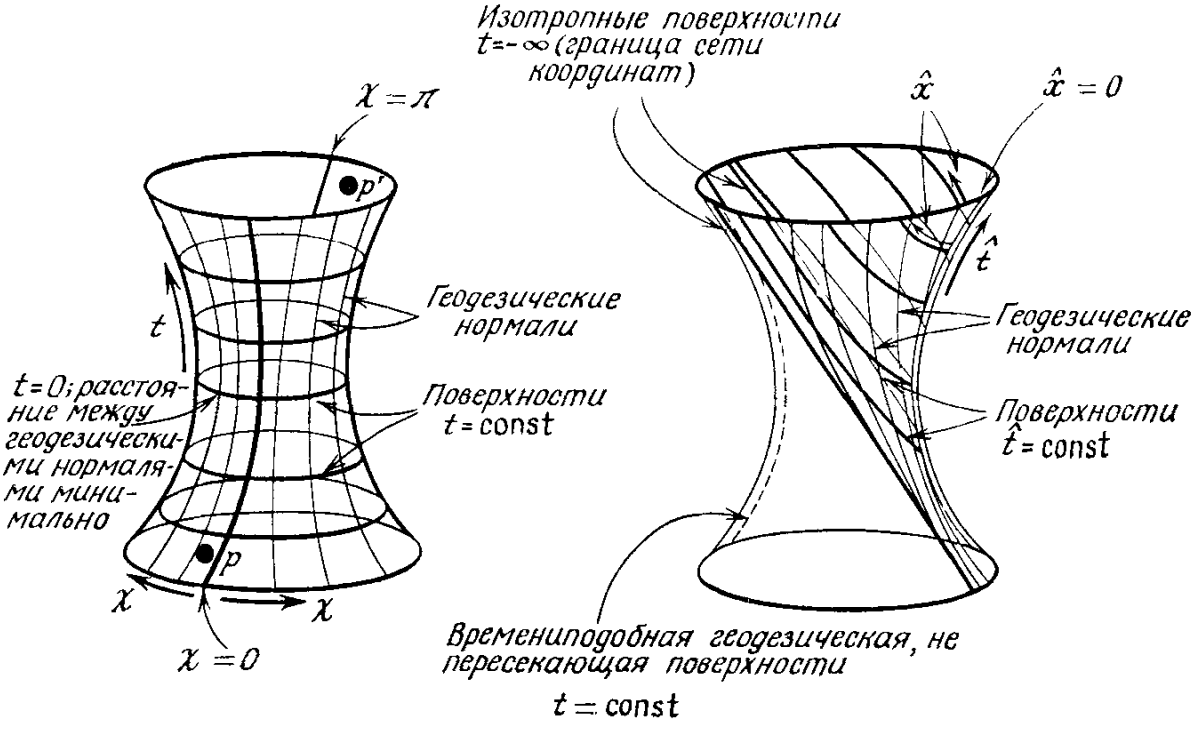


Рис. 16. Пространство-время де Ситтера 1-го рода, представленное в виде гиперболоида, вложенного в пятимерное плоское пространство (на рисунке два измерения опущены); *а* — координаты (t, χ, θ, ϕ) покрывают весь гиперболоид; сечения $\{t = \text{const}\}$ являются поверхностями кривизны $k = +1$. *б* — координаты $(\hat{t}, \hat{x}, \hat{y}, \hat{z})$ покрывают половину гиберболоида; поверхности $\{\hat{t} = \text{const}\}$ являются плоскими 3-пространствами; их геодезические нормали расходятся из точки в бесконечно удаленном прошлом.

пространства де Ситтера 1-го рода. Несмотря на эти возражения, по-настоящему справедливость стационарной модели определяется тем, насколько ее предсказания согласуются с наблюдениями. Согласия в настоящее время, по-видимому, нет, но наблюдения еще не дают достаточно оснований для окончательного вывода.

Пространство де Ситтера 1-го рода геодезически полно, но в нем существуют точки, которые нельзя соединить никакой геодезической. Этим оно сильно отличается от пространств с положительно определенной метрикой, где геодезическая полнота гарантирует возможность соединения любых двух точек по крайней мере одной геодезической. Половина пространства де Ситтера,

соответствующая стационарной модели, не полна в прошлом (существуют геодезические, которые полны во всем пространстве и пересекают границу области стационарной модели, поэтому они неполны в этой области).

Чтобы исследовать бесконечность в пространстве-времени де Ситтера 1-го рода, определим временную координату t' следующим образом:

$$t' = 2 \operatorname{arctg} (\exp \alpha^{-1} t) - \frac{1}{2} \pi, \tag{5.8}$$

где

$$-\frac{1}{2}\pi < t' < \frac{1}{2}\pi.$$

Тогда

$$ds^2 = \alpha^2 \operatorname{ch} (\alpha^{-1} t') \cdot d\tilde{s}^2,$$

где ds^2 дается формулой (5.7) после замены r' на χ. Таким образом, пространство де Ситтера 1-го рода конформно части статической вселенной Эйнштейна, определенной неравенством (5.8) (рис. 17, *а*). Диаграмма Пенроуза для пространства де Ситтера 1-го рода изображена на рис. 17, *б*. Половина этого рисунка, представленная на рис. 17, *в*, изображает ту часть пространства-времени де Ситтера, которая соответствует стационарной модели.

Мы видим, что в отличие от пространства Минковского в пространстве де Ситтера 1-го рода для времениподобных и изотропных кривых существует как в будущем, так и в прошлом пространственноподобная бесконечность. Это различие соответствует существованию в пространстве-времени де Ситтера горизонтов частиц и событий для геодезических семейств наблюдателей.

Рассмотрим в пространстве де Ситтера 1-го рода семейство частиц, историями которых являются времениподобные геодезические. Эти линии начинаются на пространственноподобной бесконечности $\mathscr{I}^-$ и кончаются на пространственноподобной бесконечности $\mathscr{I}^+$. Пусть p — некоторое событие на мировой линии частицы O этого семейства, т. е. некоторый момент ее истории (собственное время, измеренное вдоль мировой линии O). Световой конус прошлого события p — это множество событий в пространстве-времени, которые может увидеть наблюдатель, находящийся на частице в это время (наблюдатель O). Мировые линии некоторых других частиц могут пересекать этот конус, и эти частицы наблюдатель O может увидеть. Но могут существовать частицы, мировые линии которых не пересекают этот изотропный конус и, таким образом, еще не видны наблюдателю O. В более поздние моменты времени наблюдатель O может увидеть большее число частиц, но и в эти моменты будут существовать частицы, невидимые O.

Мы будем говорить, что частицы, которые наблюдатель O в событии p может увидеть, отделяются от частиц, не видимых O в это время, *горизонтом частиц наблюдателя O в событии p*;

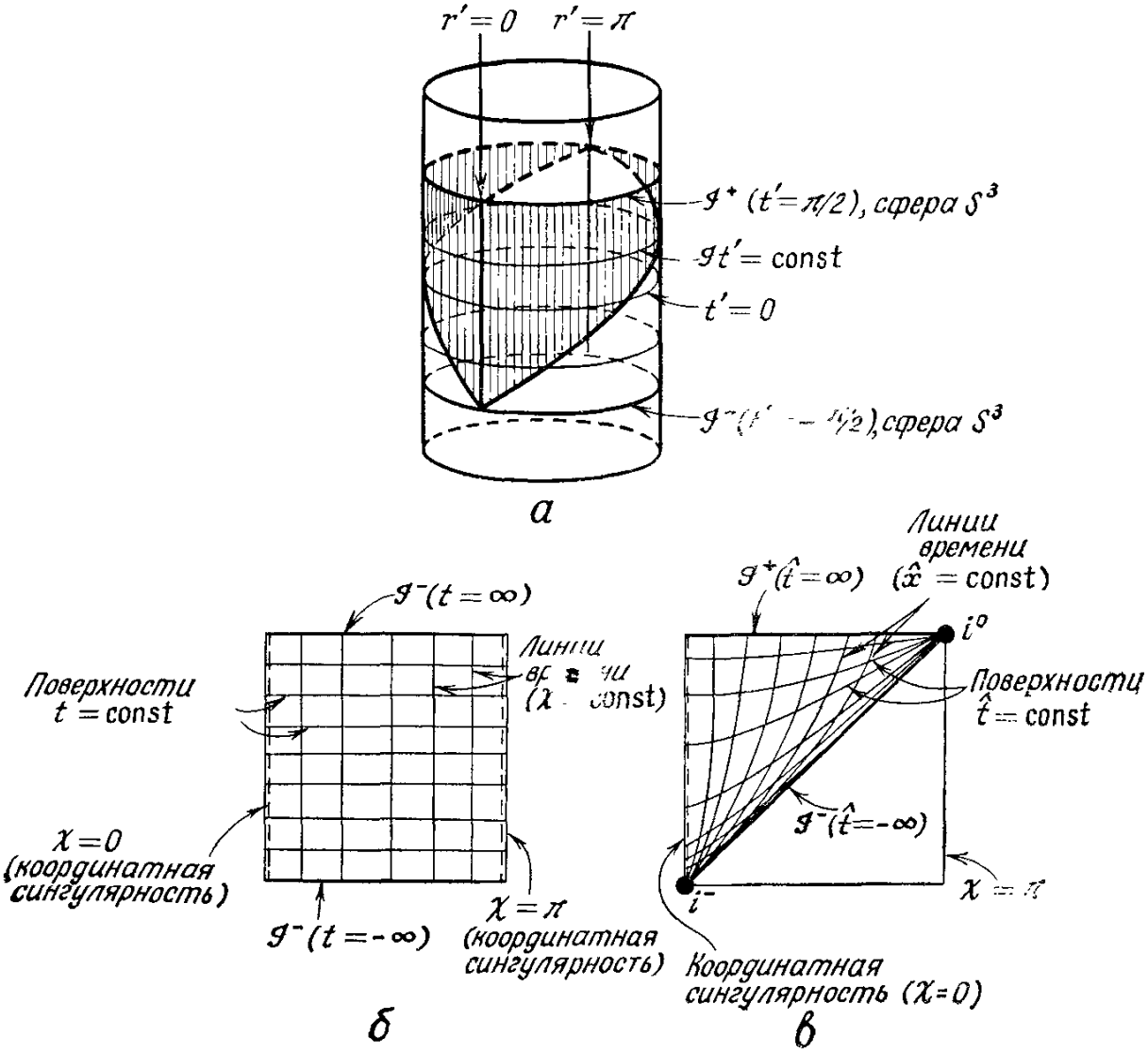


Рис. 17. *а* — пространство-время де Ситтера 1-го рода конформно области $-\pi/2 < t' < \pi/2$ статической вселенной Эйнштейна; *б* — диаграмма Пенроуза пространства-времени де Ситтера 1-го рода; *в* — диаграмма Пенроуза стационарной модели Вселенной. На диаграммах *а* и *б* каждая точка изображает 2-сферу площадью $2\pi \sin^2\chi$; изотропные линии проходят под углом 45°; $\chi = 0$ и $\chi = \pi$ отождествляются.

этот горизонт определяется мировыми линиями частиц, лежащих на пределе видимости наблюдателя O. Отметим, что горизонт частиц определен, только если мировые линии всех частиц семейства известны. Если некоторая частица лежит на этом горизонте, то p — это событие, в котором изотропный конус момента «рождения» этой частицы пересекает мировую линию O. С другой стороны, в пространстве Минковского в любом событии p

на мировой линии O можно увидеть все другие частицы, если они движутся по времениподобным геодезическим. Пока мы рассматриваем только семейства геодезических наблюдателей,

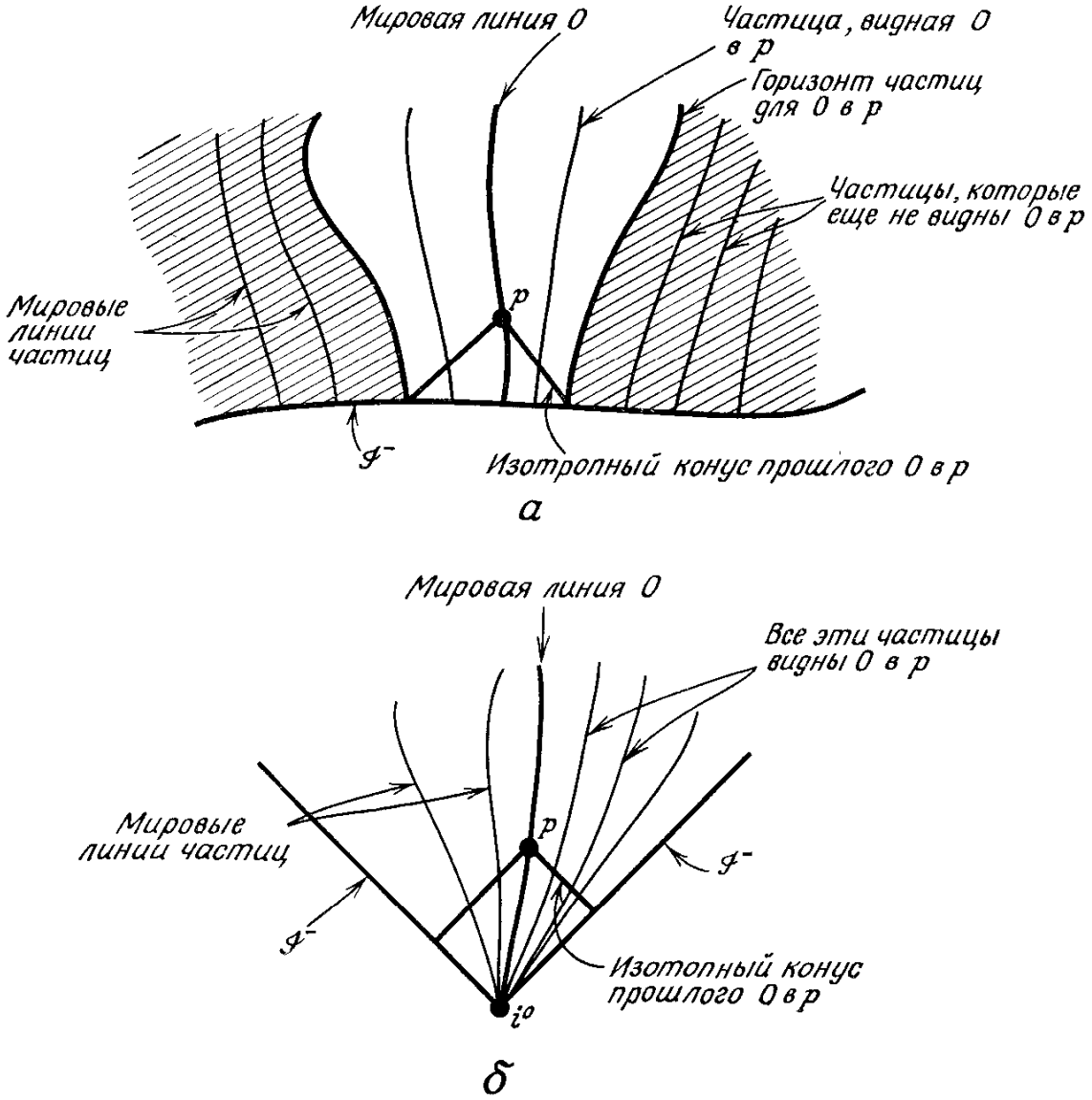
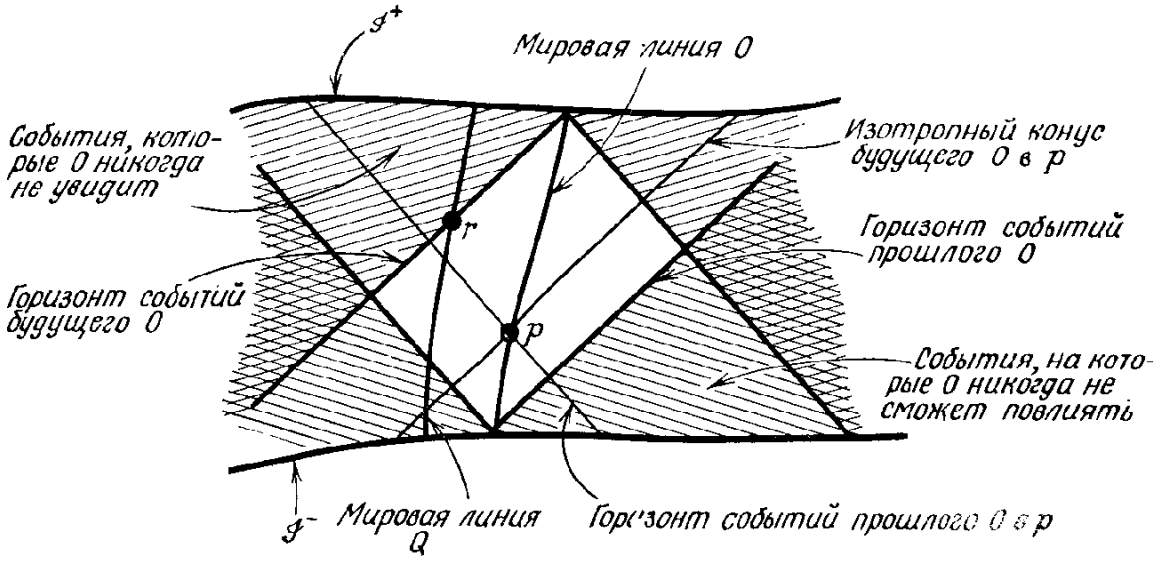


Рис. 18. *а* — горизонт частиц, определяемый конгруэнцией геодезических кривых при пространственноподобной бесконечности прошлого $\mathscr{I}$ изотропных геодезических; *б* — отсутствие такого горизонта при изотропной $\mathscr{I}^-$.

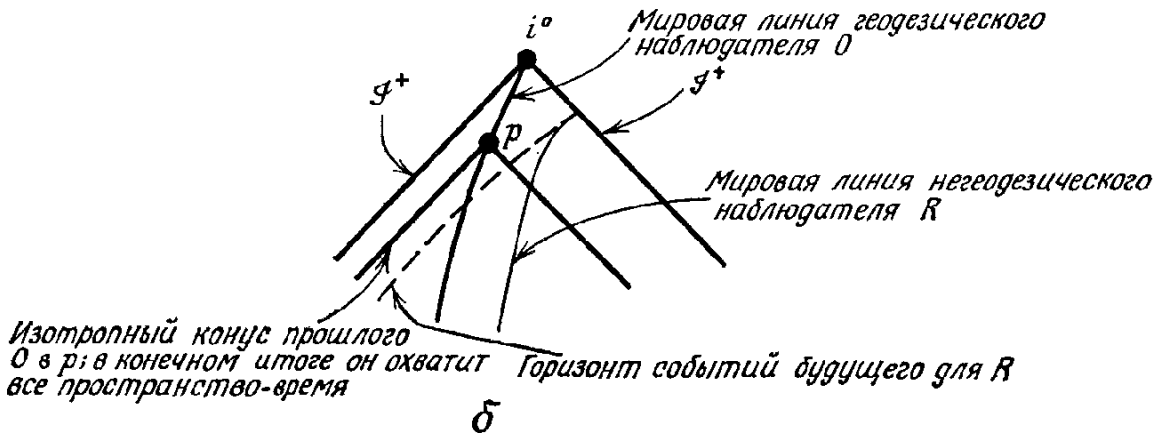
существование горизонта частиц можно считать следствием пространственноподобности бесконечности прошлого (рис. 18).

Все события вне изотропного конуса прошлого невидимы из p и до события p никогда не были видимы. Мировая линия O имеет предельную точку на $\mathscr{I}^+$. В пространстве-времени де Ситтера 1-го рода изотропный конус прошлого этой точки (получаемый предельным переходом в реальном пространстве или непосредственно из конформного пространства-времени) является границей между событиями, которые хотя бы в какой-то момент

времени станут видимы наблюдателем O, и событиями, которые O никогда не увидит. Такую поверхность мы называем *горизонтом событий будущего* данной мировой линии. Это граница



а



б

Рис. 19. *а* — горизонт событий будущего для частицы O, который существует, когда бесконечность будущего $\mathscr{I}^+$ пространственноподобна; аналогично горизонт событий прошлого, который существует, когда пространственноподобна бесконечность прошлого $\mathscr{I}^-$; *б* — если бесконечность будущего состоит из изотропной поверхности $\mathscr{I}^+$ и из i^0, для геодезического наблюдателя O не существует горизонта событий будущего. Однако у ускоряющегося наблюдателя может быть горизонт событий будущего.

мировой линии прошлого. Напротив, в пространстве-времени Минковского предельный изотропный конус любого геодезического наблюдателя включает все пространство-время и, следовательно, в нем нет событий которых геодезический наблюдатель никогда не сможет увидеть. Однако, если наблюдатель движется равноускоренно, его мировая линия может иметь горизонт событий будущего. Существование горизонта событий будущего для

геодезического наблюдателя можно рассматривать как следствие пространственноподобности $\mathscr{I}^+$ (рис. 19).

Рассмотрим горизонт событий для наблюдателя O в пространстве-времени де Ситтера и предположим, что в некоторый момент собственного времени (событие p на его мировой линии) его изотропный конус пересекает мировую линию частицы Q. После этого наблюдатель O всегда может видеть частицу Q. Но на мировой линии Q имеется событие r, которое лежит на горизонте событий будущего частицы O; O никогда не увидит тех событий на мировой линии Q, которые расположены позднее r. Более того, любое данное событие на мировой линии O отделено бесконечным промежутком собственного времени от события, когда он сможет увидеть r, но вместе с тем между любым событием вдоль мировой линии Q и событием r, представляющим собой обычное событие на этой линии, проходит конечное время. Таким образом, наблюдатель O за бесконечное время видит конечную часть истории Q; выражаясь более физично, при наблюдении частицы Q наблюдателем O происходит красное смещение, которое стремится к бесконечности по мере приближения наблюдаемых на мировой линии Q событий к точке r. Соответственно с частицы Q никогда нельзя будет увидеть событий, следующих за некоторой точкой на мировой линии O, а близлежащие к ней точки наблюдаются с очень большим красным смещением.

Изотропный конус будущего любой точки мировой линии является границей множества таких событий в пространстве-времени, на которые O может влиять начиная с данного момента. Чтобы получить максимальное множество тех событий в пространстве-времени, на которые O может влиять хоть в какой-то момент, возьмем изотропный конус будущего предельной точки мировой линии O, расположенной в бесконечности прошлого $\mathscr{I}^-$; иначе говоря, возьмем границу будущего этой мировой линии (ее можно рассматривать как изотропный конус момента рождения частицы O). Такая нетривиальная область существует для геодезического наблюдателя только тогда, когда бесконечность прошлого $\mathscr{I}^-$ пространственноподобна (и тогда она фактически является горизонтом событий прошлого частицы O). Из приведенных рассуждений ясно, что в стационарной модели, где бесконечность прошлого для изотропных и времениподобных геодезических изотропна, а бесконечность будущего пространственноподобна, любой фундаментальный наблюдатель имеет горизонт событий будущего, но не имеет горизонта событий прошлого.

Отождествлением точек пространства де Ситтера 1-го рода мы можем получить другие локально-эквивалентные ему пространства. Простейший случай — отождествление антиподальных точек p, p' (см. рис. 16) на гиперболоиде. Полученное пространство не ориентируемо во времени: если в p время течет

в направлении стрелки, то антиподальное отождествление требует, чтобы оно в p' текло в направлении против стрелки, но такое отождествление будущей прошлой половины изотропного конуса нельзя распространить на весь гиперболоид. В работе [19] подробно изучены пространства, получающиеся при таких отождествлениях; в частности, в ней показано, что произвольная точка в полученном пространстве может быть соединена геодезической с любой другой точкой, если и только если пространство не ориентируемо во времени.

Пространство постоянной кривизны с $R < 0$ называется *пространством де Ситтера 2-го рода*. Оно обладает топологией $S^1 \times R^3$ и может быть представлено в виде гиперболоида

$$-u^2 - v^2 + x^2 + y^2 + z^2 = -1$$

в плоском пятимерном пространстве R^5 с метрикой

$$ds^2 = -(du)^2 - (dv)^2 + (dx)^2 + (dy)^2 + (dz)^2.$$

В этом пространстве существуют замкнутые времениподобные линии, однако оно не является односвязным, и если мы развернем круг S^1 (чтобы получить его накрывающее пространство R^1), то будем иметь универсальное накрывающее пространство пространства де Ситтера 2-го рода, в котором нет замкнутых времениподобных линий. В дальнейшем под пространством де Ситтера 2-го рода мы будем подразумевать это универсальное накрывающее пространство; оно имеет топологию R^4.

Метрика этого пространства может быть записана в виде

$$ds^2 = -dt^2 + \cos^2 t\{d\chi^2 + \operatorname{sh}^2 \chi (d\theta^2 + \sin^2 \theta\, d\phi^2)\}. \qquad (5.9)$$

Эта система координат покрывает лишь часть пространства и имеет кажущиеся сингулярности при $t = \pm\pi/2$. Пространство целиком может быть покрыто координатами (t', r, θ, ϕ), в которых метрика принимает статическую форму:

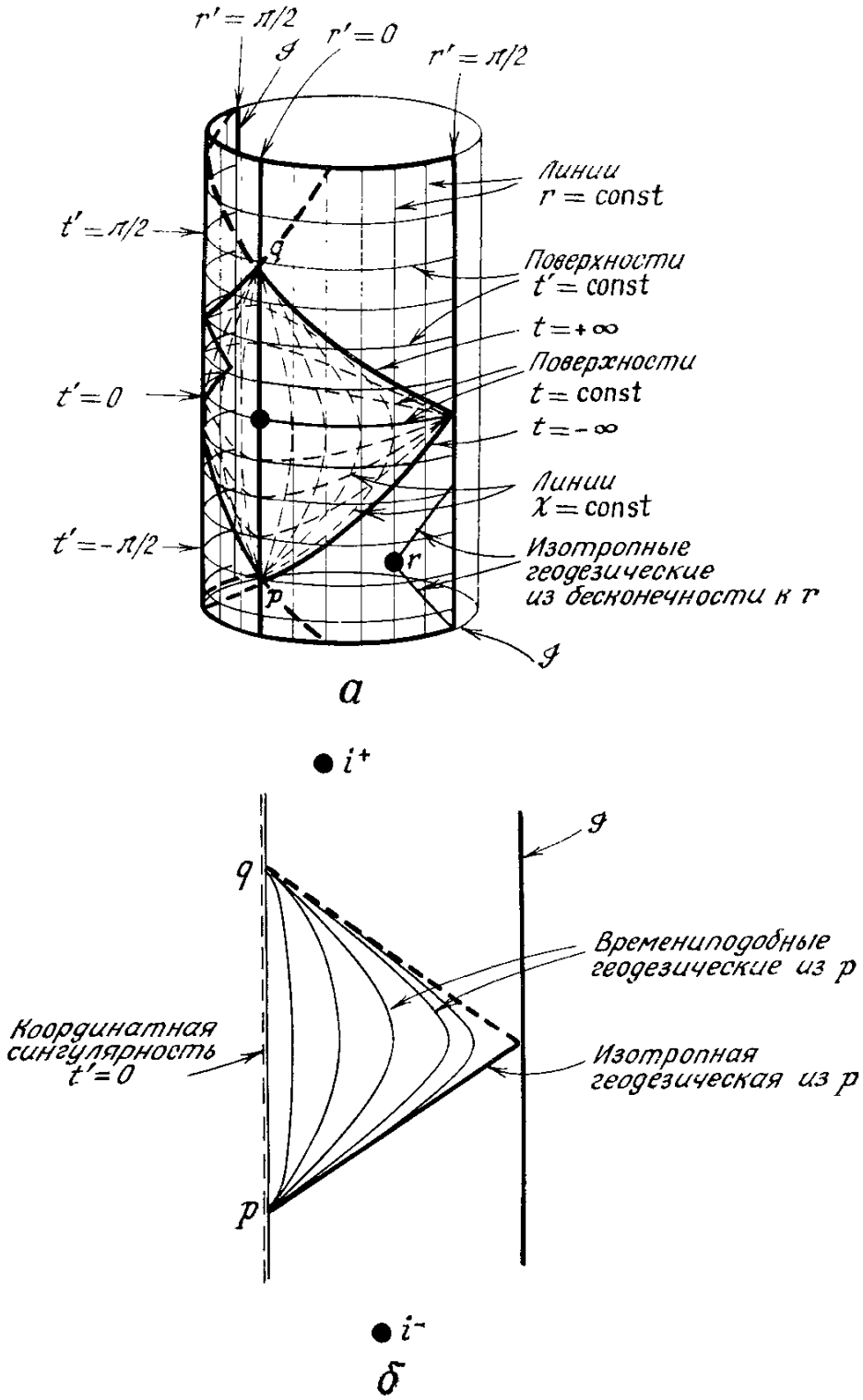
$$ds^2 = -\operatorname{ch}^2 r\, dt'^2 + dr^2 + \operatorname{sh}^2 r (d\theta^2 + \sin^2 \theta\, d\phi^2).$$

В таком представлении пространство покрывается поверхностями $\{t' = \text{const}\}$ с негеодезическими нормалями.

Для исследования структуры бесконечности введем координату r'.

$$r' = 2 \operatorname{arctg}(\exp r) - \pi/2, \quad 0 \leqslant r' < \pi/2.$$

Тогда получаем $ds^2 = \operatorname{ch}^2 r\, d\bar{s}^2$, где $d\bar{s}^2$ определяется формулой (5.7); иными словами, все пространство де Ситтера 2-го рода конформно области $0 \leqslant r' \leqslant \pi/2$ эйнштейновского статического цилиндра. Соответствующая диаграмма Пенроуза приведена на рис. 20. Изотропную и пространственную бесконечности в этом случае можно представлять себе как времениподобную поверхность. Топология этой поверхности эквивалентна $R^1 \times S^2$.



Р и с. 20. *а* — универсальное пространство де Ситтера 2-го рода конформно половине статической вселенной Эйнштейна. Координаты (t, χ, θ, ϕ) покрывают лишь изображенную на рисунке ромбовидную область тогда как координаты (t', r, θ, ϕ) покрывают все пространство. Все геодезические, ортогональные к поверхностям $\{t = \text{const}\}$, сходятся в точках p и q и затем расходятся в новую такую же ромбовидную область и т. д.; *б* — диаграмма Пенроуза пространства де Ситтера 2-го рода. Бесконечность состоит из времениподобной поверхности $\mathscr{I}$ и отдельных точек i^+, i^-. Приведены проекции некоторых времениподобных изотропных геодезических.

В пространстве де Ситтера 2-го рода нельзя построить конформное преобразование, переводящее временную бесконечность в конечную область, если не сжимать статическую вселенную Эйнштейна в точку (если при конформном преобразовании временная координата бесконечности становится конечной, то масштаб пространственного сечения умножается на бесконечный множитель), и поэтому мы изображаем времениподобную бесконечность отдельно стоящими точками i^+, i^-.

Линии $\{\chi, \theta, \phi = \text{const}\}$ являются геодезическими, ортогональными к поверхностям $\{t = \text{const}\}$; они все сходятся в будущем (соответственно в прошлом) в точках q (соответственно p) пространства-времени, и этим схождением объясняется наличие кажущихся (координатных) особенностей в первоначальном виде метрики. Этими координатами покрывается область между поверхностью $t = 0$ и изотропными поверхностями, на которых геодезические нормали становятся вырожденными.

Это пространство обладает двумя другими интересными чертами. Во-первых, вследствие времениподобности бесконечности в нем нигде нет поверхности Коши. Хотя и можно построить семейства пространственноподобных поверхностей (например, поверхности $\{t' = \text{const}\}$), которые покрывают пространство полностью, все же для любой данной поверхности такого семейства существуют изотропные геодезические, которые ее нигде не пересекают. Если задать начальные данные на какой-либо из этих поверхностей, за пределами ее области Коши нельзя что-либо предсказать; таким образом, с поверхности $\{t=0\}$ можно делать предсказания лишь в области, покрытой координатами t, χ, θ, ϕ. Любой попытке предсказаний вне этой области препятствует поступление «свежей» информации из времениподобной бесконечности.

Во-вторых, в соответствии с тем, что все геодезические нормали из $t = 0$ сходятся в p и q, все геодезические из p в прошлое расходятся (нормально к поверхности $\{t = \text{const}\}$) и снова сходятся в q. В действительности все времениподобные геодезические из любой точки этого пространства (и в прошлое, и в будущее) фокусируются в точку-изображение, от нее снова расходятся и снова фокусируются во вторую точку-изображение и т. д. Поэтому времениподобная геодезическая из p в будущее никогда не может достичь $\mathscr{I}$ в отличие от изотропной геодезической в будущее, которая доходит до $\mathscr{I}$ и образует границу будущего точки p. Как следствие такого различия между времениподобными и изотропными геодезическими в будущем точки p (т. е. там, куда могут быть проведены из p направленные в будущее времениподобные линии) существуют области, которых нельзя достичь из p по какой-либо геодезической. Множеством точек, которых можно достичь по направленным в будущее времени-

подобным линиям из p, будет множество точек, лежащих выше изотропного конуса будущего точки p; множество точек, которых можно достичь из p по направленным в будущее времениподобным геодезическим, образует внутренность бесконечной последовательности ромбообразных областей, подобных той, которая покрывается координатами (t, χ, θ, ϕ). Заметим, что любую точку в области Коши поверхности $t = 0$ можно достичь с этой поверхности по единственной геодезической нормали, но, вообще говоря, до точек вне области Коши нельзя провести геодезическую нормаль к $t = 0$.

5.3. Пространство Робертсона — Уокера

До сих пор мы не интересовались тем, как связаны с реальной физической Вселенной точные решения. Следуя Эйнштейну, мы можем спросить: найдется ли пространство-время, которое является точным решением для какой-либо подходящей формы материи и при этом хорошо описывает крупномасштабные свойства наблюдаемой Вселенной? Если таковое найдется, то мы можем утверждать, что имеем разумную «космологическую модель», т. е. модель физической Вселенной.

Однако мы не можем построить космологическую модель, которая была бы свободна от некоторой примеси философии. В самых ранних космологиях человек приписывал себе господствующее положение в центре Вселенной. Начиная с времен Коперника, мы постепенно спустились на среднего размера планету, обращающуюся вокруг среднего размера звезды на окраине довольно обычной галактики, которая сама является всего лишь членом Местной группы галактик. Мы стали теперь столь демократичны, что не возьмемся утверждать, будто наше положение в пространстве хоть как-то выделено. Вслед за Бонди [10] будем называть предположение о невыделенности нашего положения *принципом Коперника*.

Этот несколько туманный принцип можно интерпретировать следующим образом: при выборе подходящего масштаба Вселенная приблизительно пространственно однородна.

Под пространственной однородностью мы подразумеваем существование действующей на $\mathscr{M}$ группы изометрий, поверхности транзитивности *) которой являются пространственноподобными 3-поверхностями; другими словами, любая точка на одной из этих поверхностей эквивалентна любой другой точке той же поверхности. Конечно, Вселенная не в точности пространственно однородна; в ней существуют локальные неоднородности, такие,

*) Поверхностью транзитивности группы преобразований называется поверхность, каждая точка которой может быть переведена преобразованием группы в любую другую точку этой поверхности. — *Прим. ред.*

как звезды и галактики. Тем не менее представляется разумным предположить, что Вселенная пространственно однородна в достаточно больших масштабах.

Хотя мы и можем построить математические модели, удовлетворяющие этому требованию однородности (см. следующий раздел), но непосредственно проверить однородность с помощью наблюдений трудно вследствие того, что нет простого способа измерения расстояний до удаленных объектов. Эта трудность смягчается тем обстоятельством, что в принципе мы довольно легко можем обнаружить путем наблюдений внегалактических объектов изотропию (т. е. проверить, совпадают ли результаты наблюдений в различных направлениях), а изотропия тесно связана с однородностью. Выполненные до сих пор наблюдения по проверке изотропии подтверждают вывод о приближенной сферической симметрии окружающей нас Вселенной.

В частности, было показано, что внегалактические радиоисточники распределены приблизительно изотропно, а обнаруженный недавно фон микроволнового излучения в исследованной области спектра в высшей степени изотропен.

Записать и исследовать метрики для всех случаев сферически-симметричного пространства-времени вполне возможно. Конкретными примерами пространств с такой симметрией являются решения Шварцшильда и Райснера—Нордстрема (см. разд. 5.5), которые, однако, являются асимптотически плоскими пространствами. Вообще говоря, в сферически-симметричных пространствах есть самое большее две точки, из которых пространство представляется сферически-симметричным. Такие пространства могут служить моделями пространства-времени вблизи массивного тела, но если принять их в качестве моделей Вселенной, то это согласуется с изотропией по данным наших наблюдений лишь при условии, что мы расположены вблизи какой-то весьма специфической точки. Исключением являются те модели, в которых Вселенная изотропна относительно каждой точки в пространстве-времени; поэтому мы будем трактовать принцип Коперника как утверждение о приближенной сферической симметрии Вселенной относительно каждой точки (поскольку она приближенно сферически-симметрична вокруг нас).

Как было показано Уокером [171], точная сферическая симметрия относительно каждой точки означает, что Вселенная пространственно однородна и допускает 6-параметрическую группу изометрий с поверхностями транзитивности в виде пространственноподобных 3-поверхностей постоянной кривизны. Такого рода пространства называются пространствами *Робертсона — Уокера* (или *Фридмана*); пространство Минковского, пространства де Ситтера обоих родов являются частными случаями пространства Робертсона — Уокера общего вида. Отсюда мы заклю-

чаем, что пространства Робертсона — Уокера дают хорошее приближение к крупномасштабной геометрии пространства-времени в области, доступной нашим наблюдениям.

В пространстве Робертсона — Уокера координаты можно выбрать так, чтобы метрика имела вид

$$ds^2 = -dt^2 + S^2(t)\,d\sigma^2,$$

где $d\sigma^2$ — метрика трехмерного пространства постоянной кривизны, причем она не зависит от времени. Геометрия этого трехмерного пространства существенно зависит от того, положительна, отрицательна или равна нулю его кривизна K; подбором функции S можно нормировать K так, чтобы в первых двух случаях она была равна $+1$ или -1. Тогда метрику $d\sigma^2$ можно записать в виде

$$d\sigma^2 = d\chi^2 + f^2(\chi)(d\theta^2 + \sin^2\theta\, d\phi^2),$$

где

$$f(\chi) = \begin{cases} \sin\chi & \text{при } K = +1, \\ \chi & \text{при } K = 0, \\ \operatorname{sh}\chi & \text{при } K = -1. \end{cases}$$

Если $K = 0$ или $K = -1$, координата χ пробегает значения от 0 до ∞, а при $K = +1$ имеем $0 \leqslant \chi \leqslant 2\pi$. При $K=0$ или $K=-1$ трехмерное пространство диффеоморфно R^3 и, таким образом, «бесконечно»; когда $K = +1$, оно диффеоморфно трехмерной сфере S^3 и, следовательно, компактно («закрыто» или «конечно»). Можно было бы отождествить подходящие точки в этих трехмерных пространствах и получить иные глобальные топологии; в случае отрицательной и нулевой кривизны это можно сделать даже так, чтобы в результате получилось компактное трехмерное пространство [102]. Однако такая компактная поверхность постоянной кривизны не допускает никакой непрерывной группы изометрий [175]: хотя в каждой точке существуют векторы Киллинга, они не задают какое-либо глобальное векторное поле Киллинга, и локальные группы изометрий, определяемые ими, нельзя объединить в одну глобальную группу. В случае нулевой кривизны компактное пространство может обладать только 3-параметрической группой изометрий. Ни в одном случае полученное пространство-время не будет изотропным. Мы не будем делать подобных отождествлений, поскольку поводом для изучения пространств Робертсона — Уокера была их изотропия (и, следовательно, наличие 6-параметрической группы изометрий). В действительности единственным отождествлением, которое не приводит к анизотропному пространству, является отождествление антиподальных точек на S^3 в случае постоянной положительной кривизны.

Симметрия решений Робертсона — Уокера требует, чтобы тензор энергии-импульса имел вид, соответствующий идеальной жидкости, плотность μ и давление p которой являются функциями только временной координаты t, а линии тока описываются кривыми χ, θ, $\phi = \text{const}$ (так что χ, θ, ϕ — сопутствующие координаты). Можно считать, что эта жидкость описывает материю во Вселенной в приближении равномерного «размазывания»; тогда функция $S(t)$ характеризует девиацию соседних линий тока, т. е. «близких» галактик.

Уравнение сохранения энергии (3.9) теперь принимает вид

$$\dot{\mu} = -3(\mu + p)\dot{S}/S. \tag{5.10}$$

Уравнение Райчаудхури (4.26) выглядит следующим образом:

$$4\pi(\mu + 3p) - \Lambda = -3\ddot{S}/S, \tag{5.11}$$

уравнения поля сводятся к одному уравнению [по существу, это соотношение (2.35)]:

$$3\dot{S}^2 = 8\pi(\mu S^3)/S + \Lambda S^2 - 3K. \tag{5.12}$$

Когда $\dot{S} \neq 0$, соотношение (5.12) можно получить (с произвольным значением K) как первый интеграл уравнений (5.10), (5.11); таким образом, оно фактически служит для отождествления постоянной K с кривизной трехмерного пространства $\{t = \text{const}\}$ с метрикой $d\sigma^2$.

Разумно предположить (см. энергетические условия, разд. 4.3), что $\mu > 0$ и $p \geqslant 0$. (В действительности по современным оценкам 10^{-29} г/см$^3 \geqslant \mu_0 \geqslant 10^{-31}$ г/см^3, $\mu_0 \gg p_0 \geqslant 0$.) Тогда, если $\Lambda = 0$, то из (5.11) видно, что S не может быть постоянной; иначе говоря, из уравнений поля следует, что Вселенная должна или расширяться, или сжиматься. Наблюдения показывают (это впервые обнаружили Слайфер и Хаббл), что галактики удаляются от нас, т. е. что в настоящее время вещество во Вселенной расширяется. Последние наблюдения дают для современного значения $\dot{S}/S$ значение

$$H \equiv (\dot{S}/S)|_0 \approx 10^{-10} \text{ лет}^{-1}$$

(по-видимому, с точностью до множителя ~ 2). Отсюда, согласно (5.11), следует, что при $\Lambda = 0$ S должна обратиться в нуль за некоторый конечный интервал времени t_0 в прошлом (время t_0 измеряется вдоль мировой линии нашей Галактики), причем

$$t_0 < H^{-1} \approx 10^{10} \text{ лет}.$$

Из (5.10) следует, что по мере расширения Вселенной плотность уменьшается; в прошлом, наоборот, она была выше и принимала сколь угодно большие значения при $S \to 0$. Поэтому точка, где

$S = 0$, не является только координатной особенностью, как при введении координат (5.9) во вселенной де Ситтера 2-го рода; тот факт, что плотность обращается при $S = 0$ в бесконечность, означает, что некоторый скаляр, определяемый тензором кривизны, тоже бесконечен. Это делает подобную сингулярность даже серьезней, чем та, которая возникает в соответствующей ньютоновской ситуации; в обоих случаях мировые линии всех частиц пересекаются в одной точке и плотность становится бесконечной, но здесь становится сингулярным в точке $S = 0$ само пространство-время. Мы должны исключить эту точку из пространственно-временного многообразия, поскольку в ней не могут выполняться никакие известные физические законы.

Эта сингулярность — наиболее удивительное свойство решений Робертсона — Уокера. Она возникает во всех моделях, в которых $\mu + 3p > 0$ и Λ отрицательна, равна нулю или имеет не слишком большое положительное значение. Сингулярность означает, что у Вселенной (или по крайней мере у той ее части, о физике которой мы имеем некоторое представление) имеется начало, отдаленное от нас конечным промежутком времени. Однако этот результат основан на предположении о строгой пространственной однородности и сферической симметрии. Хотя это предположение может оказаться разумным приближением для достаточно больших масштабов в настоящее время, оно, конечно, не выполняется локально. Можно допустить, что при прослеживании эволюции Вселенной назад во времени местные неоднородности настолько возрастут, что они смогут воспрепятствовать появлению сингулярности, приводя к «упругому отражению» частей Вселенной. Может ли подобное произойти и содержатся ли сингулярности в реалистических решениях с неоднородностями — это центральные вопросы космологии; их обсуждение составляет основной предмет этой книги. Как мы увидим, имеются веские основания считать, что физическая Вселенная действительно была сингулярной в прошлом.

Если задана какая-либо разумная связь между p и μ, уравнение (5.10) можно проинтегрировать и получить μ в функции S. Давление в настоящую эпоху очень мало. Если p и Λ положить равными нулю, то из (5.10) находим

$$\frac{4\pi}{3}\mu = \frac{M}{S^3},$$

где M — постоянная; при этом (5.12) имеет вид

$$3\dot{S}^2 - 6\frac{M}{S} = -3K \equiv \frac{E}{M}. \tag{5.13}$$

Первое уравнение выражает сохранение массы при нулевом давлении, а второе (*уравнение Фридмана*) представляет собой уравнение сохранения энергии для сопутствующего объема материи:

постоянная E — это сумма кинетической и потенциальной энергий. Если $E < 0$ (т. е. $K > 0$), то S будет возрастать до некоторого максимального значения и затем спадет до нуля; если же $E \geqslant 0$ (т. е. $K \leqslant 0$), то S возрастет неограниченно.

Явные решения уравнения (5.13) примут простой вид, если их записать через новый временной параметр $\tau(t)$, определяемый равенством

$$\frac{d\tau}{dt} = S^{-1}(t); \tag{5.14}$$

тогда имеем

$$S = \left(\frac{E}{3}\right)(\operatorname{ch}\tau - 1), \quad t = \frac{E}{3}(\operatorname{sh}\tau - \tau) \quad \text{при} \quad K = -1;$$

$$S = \tau^2, \quad t = \frac{1}{3}\tau^3 \quad \text{при} \quad K = 0;$$

$$S = -\frac{E}{3}(1 - \cos\tau), \quad t = -\frac{E}{3}(\tau - \sin\tau) \quad \text{при} \quad K = 1.$$

(Случаю $K = 0$ соответствует вселенная Эйнштейна — де Ситтера; очевидно, тогда $S \sim t^{\frac{2}{3}}$.)

При $p > 0$ поведение качественно такое же. В частности, для $p = (\gamma - 1)\mu$, где $\gamma = \text{const}$, $1 \leqslant \gamma \leqslant 2$, получаем $4\pi\mu/3 = M/S^{3\gamma}$, и решение (5.12) вблизи сингулярности имеет вид

$$S \sim t^{\frac{2}{3}\gamma}.$$

Если $\Lambda < 0$, то решение свидетельствует о расширении от начальной сингулярности к некоторому максимуму с последующим коллапсом во вторую сингулярность. Если $\Lambda > 0$, то для $K = 0$ или -1 расширение продолжается вечно, и асимптотически решение приближается к стационарной модели. При $K = 1$ существует несколько возможностей. Если Λ больше некоторого значения $\Lambda_{\text{крит}}$ (при $p = 0$ $\Lambda_{\text{крит}} = (-E/3M)^3/(3M)^2$), решение начинается с сингулярности и описывает вечное расширение с асимптотическим приближением к стационарной модели. При $\Lambda = \Lambda_{\text{крит}}$ имеем статическое решение — *статическую вселенную Эйнштейна*. [Метрическая форма (5.7) — это частный случай статического решения Эйнштейна с $\mu + p = (4\pi)^{-1}$, $\Lambda = 1 + 8\pi p$.] Существует еще решение, начинающееся с сингулярности и асимптотически приближающееся к вселенной Эйнштейна, и решение, начинающееся со вселенной Эйнштейна в бесконечном прошлом и все время расширяющееся. Если $\Lambda < \Lambda_{\text{крит}}$, имеется два решения: одно описывает расширение от начальной сингулярности и затем коллапс во вторую сингулярность; второе — сжатие, начинающееся при бесконечном радиусе в бесконечном прошлом, до некоторого минимального радиуса с последующим

расширением. Только это последнее решение и решение, которое в бесконечном прошлом асимптотически приближается к статической вселенной, могли бы соответствовать наблюдаемой Вселенной и не иметь при этом сингулярностей. Однако в этих моделях всегда $S > 0$, что, по-видимому, противоречит результатам наблюдений красных смещений далеких галактик [145, 146]. Кроме того, максимальная плотность в этих моделях, оказывается, не слишком сильно отличается от современной плотности, что приводит к трудностям в объяснении таких явлений, как микроволновой фон излучения и распространенность гелия в космосе: перечисленные явления скорее всего свидетельствуют об очень горячей плотной фазе в истории Вселенной.

Так же как и в предыдущих случаях, мы можем построить конформные отображения пространств Робертсона — Уокера в статическое пространство Эйнштейна. В качестве временной координаты будем использовать параметр τ, определяемый соотношением (5.14); тогда метрика принимает вид

$$ds^2 = S^2(\tau)\{-d\tau^2 + d\chi^2 + f^2(\chi)(d\theta^2 + \sin^2\theta\, d\phi^2)\}. \quad (5.15)$$

В случае $K = +1$ она уже конформна статическому пространству Эйнштейна [для согласования с обозначениями формулы (5.7) нужно положить $\tau = t'$, $\chi = r'$]. Таким образом, это пространство отображается точно на ту часть статического пространства Эйнштейна, которая определяется областью значений τ. Когда $p = \Lambda = 0$, τ меняется в пределах $0 < \tau < \pi$ и все пространство отображается на эту область статического пространства, а его граница — на трехмерные сферы $\tau = 0$, $\tau = \pi$. (Если $p > 0$, оно отображается на область, в которой $0 < \tau < a < \pi$, где a — некоторое число). В случае $K = 0$ пространство в тех же координатах оказывается конформно-плоским [см. (5.15)], так что после конформных преобразований (см. разд. 5.1) мы получим отображение рассматриваемого пространства на область, представляющую пространство-время Минковского в статической вселенной Эйнштейна (см. рис. 14); конкретный вид области отображения снова определяется значениями, которые принимает τ. При $\Lambda = 0$ имеем $0 < \tau < \infty$, и, следовательно, это пространство (которое при $p = 0$ представляет собой пространство Эйнштейна — де Ситтера) конформно половине области, изображающей пространство-время Минковского. В случае $K = -1$ получаем метрику, конформную части той области статического пространства Эйнштейна, для которой $\pi/2 \geqslant t' + r' \geqslant -\pi/2$, $\pi/2 \geqslant t' - r' \geqslant -\pi/2$, причем

$$t' = \operatorname{arctg}\left(\operatorname{th}\tfrac{1}{2}(\tau + \chi)\right) + \operatorname{arctg}\left(\operatorname{th}\tfrac{1}{2}(\tau - \chi)\right),$$

$$r' = \operatorname{arctg}\left(\operatorname{th}\tfrac{1}{2}(\tau + \chi)\right) - \operatorname{arctg}\left(\operatorname{th}\tfrac{1}{2}(\tau - \chi)\right).$$

Какая именно часть этой области покрывается, зависит от области изменения τ; при $\Lambda = 0$ рассматриваемое пространство отображается на верхнюю половину.

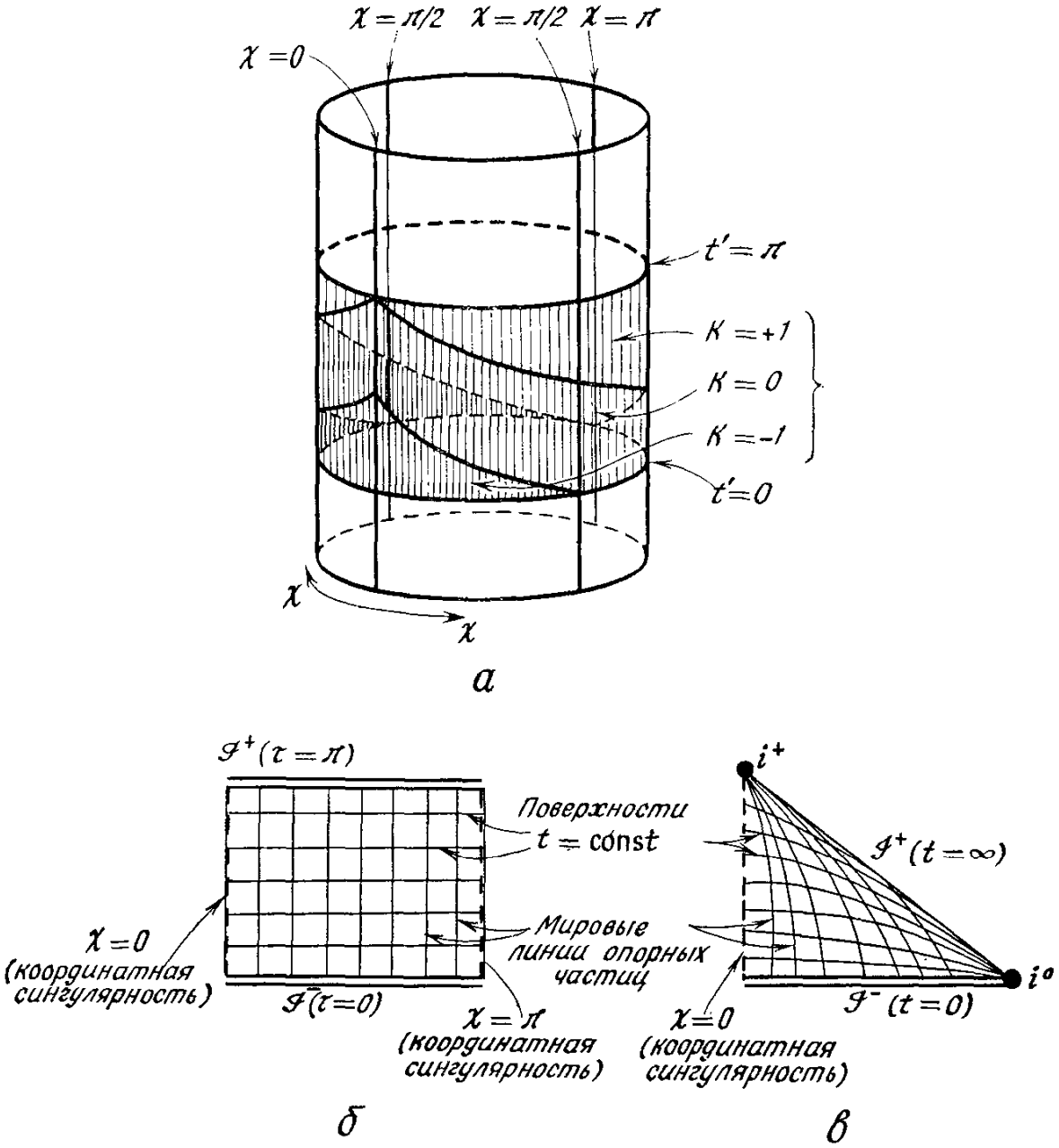


Рис. 21. *а* — пространства Робертсона — Уокера ($p = \Lambda = 0$) конформны изображенным на рисунке областям, которые соответствуют трем случаям: $K = +1, 0, -1$; *б* — диаграмма Пенроуза пространства Робертсона — Уокера при $K = +1$ и $p = \Lambda = 0$; *в* — диаграмма Пенроуза пространства Робертсона — Уокера при $K = 0$ или $K = -1$ и $p = \Lambda = 0$.

Итак, мы установили, что эти пространства и их границы конформны некоторой (вообще говоря, конечной) области статического пространства Эйнштейна (рис. 21, *а*). Однако имеется важное отличие от предыдущих случаев: часть границы теперь не

является «бесконечностью» в прежнем смысле, а представляет собой сингулярность, в которой $S = 0$. (Можно считать, что конформный множитель переводит бесконечную область в конечную путем бесконечно сильного сжатия, а сингулярную точку $S = 0$ превращает в конечную область бесконечно сильным растяжением.) Вносимые при этом в конформные диаграммы отличия по существу невелики: мы по-прежнему можем построить диаграммы Пенроуза (рис. 21, *б* и *в*). В каждом случае, когда $p \geqslant 0$, сингулярность при $t = 0$ изображается пространственноподобной поверхностью. Это соответствует существованию горизонтов частиц (определяемых точно так же, как в разд. 5.2) в пространствах Робертсона — Уокера. Когда $K = +1$, граница в будущем также пространственноподобна, из чего следует существование горизонтов событий для фундаментальных наблюдателей. Если $\Lambda = 0$, а $K = 0$ или $K = -1$, то бесконечность будущего изотропна, и в этих случаях нет никаких горизонтов событий будущего для фундаментальных наблюдений.

На этом этапе нам следует разобраться в таком вопросе: пространство де Ситтера 2-го рода можно представить в робертсон-уокеровском виде (5.9) и затем отобразить конформно на часть статической вселенной Эйнштейна. Если это проделать, то мы обнаружим, что координаты Робертсона — Уокера покрывают только малую часть всего пространства-времени. Иначе говоря, пространство-время, описываемое координатами Робертсона — Уокера, можно расширить. Поэтому нам нужно показать, что вселенные Робертсона — Уокера, в которых имеется материя, на самом деле нерасширяемы.

Это действительно так, поскольку при $\mu > 0$ и $p \geqslant 0$ для любого вектора $\mathbf{X}$ в произвольной точке q можно показать, что геодезическая $\gamma(v)$, проходящая через $q = \gamma(0)$ в направлении $\mathbf{X}$, удовлетворяет одному из двух условий:

1) $\gamma(v)$ можно продолжить на произвольные положительные значения v или

2) существует некоторое значение $v = v_0 > 0$, такое, что скалярный инвариант

$$\left(R_{ij} - \tfrac{1}{2} R g_{ij}\right)\left(R^{ij} - \tfrac{1}{2} R g^{ij}\right) = (\mu + \Lambda)^2 + 3(p - \Lambda)^2$$

не ограничен на $\gamma([0, v_0])$.

Ясно, что поверхности $\{t = \text{const}\}$ являются поверхностями Коши в этих пространствах. Далее, мы видим, что сингулярность в них универсальна в следующем смысле: все времениподобные и изотропные геодезические, проходящие через какую-либо точку пространства, приближаются к сингулярности при некотором конечном значении своего аффинного параметра. (Поскольку

изотропные геодезические образуют границу прошлого каждой точки, все непродолжимые времениподобные кривые также стремятся к сингулярности.)

5.4. Пространственно-однородные космологические модели

Как мы видим, в любом пространстве-времени Робертсона — Уокера, в котором $\mu > 0$, $p \geqslant 0$ и постоянная Λ не слишком велика, имеются сингулярности. Однако из этого нельзя делать вывода о существовании сингулярностей в более реалистических моделях мира, учитывающих тот факт, что Вселенная не однородна и не изотропна. Конечно, никто не надеется получить точное решение, которое бы описывало Вселенную во всех деталях. Однако можно найти точные решения с меньшими ограничениями, чем в решениях Робертсона — Уокера, которые могли бы служить приемлемой моделью Вселенной, и посмотреть, имеются в них сингулярности или нет; тот факт, что эти модели действительно содержат сингулярности, указывает на то, что существование сингулярностей, возможно, является общим свойством всех пространств, которые можно принять в качестве разумных моделей Вселенной.

Мы получим простой класс таких решений, если, сохраняя требование *пространственной однородности* (строгий принцип Коперника), откажемся от требования изотропии (в настоящее время Вселенная кажется приблизительно изотропной, но могла быть значительная анизотропия в более ранние эпохи). Таким образом, в этих моделях предполагается существование группы изометрий G_r, орбитами которой в некоторой части модели являются пространственноподобные гиперповерхности. (Орбита точки p под действием группы G_r есть множество точек, в которые переводится точка p под действием всех элементов группы.) Такие модели локально можно построить известными методами; случай $r = 3$ рассмотрен в [76], случай $r = 4$ — в [87] (если $r > 4$, пространство-время по необходимости является пространством Робертсона — Уокера).

Наиболее простой случай пространственной однородности — это когда соответствующая группа изометрий абелева; тогда пространство относится к типу I по классификации Бианки [7]; так что пространство-время с такой группой изометрий будем называть пространством *типа I Бианки*. Мы рассмотрим некоторые свойства этих пространств и затем сформулируем теорему, утверждающую, что сингулярности имеют место во всех непустых пространственно-однородных моделях, в которых удовлетворяется условие времениподобного схождения (разд. 4.3).

Допустим, что пространственно-однородное пространство-время обладает абелевой группой изометрий; для простоты

предположим, что $\Lambda = 0$ и что материальное содержимое Вселенной представляет собой идеальную жидкость без давления («пыль»). Тогда существуют сопутствующие координаты (t, x, y, z), в которых метрика имеет вид

$$ds^2 = -dt^2 + X^2(t)\,dx^2 + Y^2(t)\,dy^2 + Z^2(t)\,dz^2. \qquad (5.16)$$

Если соотношением $S^2 = XYZ$ ввести функцию $S(t)$, то из законов сохранения видно, что плотность материи определяется равенством $4\pi\mu/3 = M/S^3$, где M — подходящим образом выбранная постоянная. Общее решение полевых уравнений можно записать в виде

$$X = S\left(\frac{t^{2/3}}{S}\right)^{2\sin\alpha}, \quad Y = S\left(\frac{t^{2/3}}{S}\right)^{2\sin(\alpha+2\pi/3)},$$

$$Z = S\left(\frac{t^{2/3}}{S}\right)^{2\sin(\alpha+4\pi/3)},$$

причем

$$S^3 = \frac{9}{2} Mt(t+\Sigma);$$

$\Sigma\,(>0)$ — постоянная, характеризующая степень анизотропии (мы исключаем изотропный случай $\Sigma = 0$, совпадающий со вселенной Эйнштейна — де Ситтера (разд. 5.3)), а постоянная α $(-\pi/6 < \alpha \leqslant \pi/2)$ задает направление наиболее быстрого расширения. Средняя скорость расширения определяется равенством

$$\frac{\dot{S}}{S} = \frac{2}{3t}\,\frac{t+\frac{\Sigma}{2}}{t+\Sigma},$$

скорость расширения в направлении x равна

$$\frac{\dot{X}}{X} = \frac{2}{3t}\,\frac{t+\frac{1}{2}\Sigma(1+2\sin\alpha)}{t+\Sigma},$$

а скорости расширения $\dot{Y}/Y$, $\dot{Z}/Z$ в направлениях y и z даются выражениями, в которых α заменена соответственно на $\alpha+2\pi/3$, $\alpha + 4\pi/3$. Такое решение описывает расширение из сильно анизотропного сингулярного состояния при $t = 0$, приводящее при больших t к почти изотропной фазе, в которой пространство-время такое же, как во вселенной Эйнштейна — де Ситтера. Средняя длина S монотонно возрастает по мере возрастания t; первоначально высокая скорость ее изменения $(S \sim t^{1/3}$ при малых $t)$ постоянно убывает $(S \sim t^{2/3}$ при больших $t)$. Следовательно, в ранние периоды анизотропная Вселенная эволюционирует быстрее, чем ее изотропный эквивалент.

Допустим, что мы рассматриваем эту модель назад во времени и приближаемся к сингулярности. Первоначальное почти

изотропное сжатие станет сильно анизотропным в ранние периоды. При общем значении α, т. е. при $\alpha \neq \pi/2$, член $1 + 2\sin(\alpha + 4\pi/3)$ будет отрицательным. Следовательно, коллапс в направлении оси z должен прекратиться и смениться расширением, причем скорость расширения в достаточно поздние моменты времени стремится к бесконечности. Напротив, в направлении осей x и y коллапс будет монотонно стремиться к сингулярности. Таким образом, при развитии рассматриваемой модели в обычном направлении времени мы имеем «сигарообразную» сингулярность: вдоль оси z материя коллапсирует из бесконечности, затем коллапс прекращается и начинается расширение, в то время как в направлениях осей x и y материя монотонно расширяется во все моменты времени. Если в такой Вселенной можно было бы принять какой-либо сигнал, испущенный в достаточно ранние моменты времени, то в направлении оси z наблюдалось бы максимальное красное смещение; материю в более ранние моменты времени мы наблюдали бы со все уменьшающимся красным смещением, которое переходило бы по мере удаления в прошлое в фиолетовое смещение.

Несколько иначе происходит развитие в специальном случае $\alpha = \pi/2$. Тогда оба множителя $1 + 2\sin(\alpha + 2\pi/3)$ и $1 + 2\sin(\alpha + 4\pi/3)$ обращаются в нуль и скорости расширения имеют вид

$$\frac{\dot{X}}{X} = \frac{2}{3t}\,\frac{t + \frac{3\Sigma}{2}}{t + \Sigma}, \qquad \frac{\dot{Y}}{Y} = \frac{\dot{Z}}{Z} = \frac{2}{3}\,\frac{1}{t + \Sigma}.$$

Если рассматривать обращенную во времени модель, то скорость коллапса в направлениях y и z асимптотически замедляется до нуля, а в направлении x бесконечно возрастает. При обычном направлении времени мы имеем «блинообразную» сингулярность: материя монотонно расширяется во всех направлениях с бесконечно большой начальной скоростью расширения в направлении оси x и без начальной скорости в направлениях осей y и z. Красное смещение в направлении оси x будет сколь угодно большим, а в направлении y и z — конечным.

Дальнейшее исследование показывает, что в общем («сигарообразном») случае, несмотря на анизотропию расширения, существуют горизонты частиц в каждом направлении. В специальном («блинообразном») случае в направлении оси x нет никаких горизонтов; действительно, координаты частиц, которые видны в момент t_0 из начала координат (x, y, z), лежат внутри бесконечного цилиндра

$$x^2 + y^2 < \rho^2,$$

где

$$\rho = \frac{2}{3M}\left\{\left(\frac{9M}{2}(t_0 + \Sigma)\right)^{1/3} - \left(\frac{9M}{2}\Sigma\right)^{1/3}\right\}.$$

Хотя мы рассмотрели пространственно-однородные модели только при нулевом давлении и $\Lambda = 0$, легко можно получить свойства этих пространств при более реалистическом материальном содержимом; если, например, мы возьмем идеальную жидкость с $p = (\gamma - 1)\mu$, $\gamma = \text{const}\,(1 < \gamma < 2)$ или смесь фотонного газа и вещества с $p \leqslant \mu/3$, поведение вблизи сингулярности будет таким же, как и в случае пыли.

Интересным следствием отсутствия горизонта частиц в направлении оси x в специальном («блинообразном») случае является возможность непрерывного продолжения решения за сингулярность. Покажем это в случае решения для пыли.

Метрика имеет вид (5.16), причем теперь

$$X(t) = t\left(\frac{9}{2}Mt(t+\Sigma)\right)^{-1/3}, \quad Y(t) = Z(t) = \left(\frac{9}{2}Mt(t+\Sigma)\right)^{2/3}. \tag{5.17}$$

Введем новые координаты τ, η, удовлетворяющие уравнениям

$$\text{th}\,\frac{2x}{9M\Sigma} = \frac{\eta}{\tau}, \quad \exp\left(\frac{4}{9M}\int_0^t \frac{dt}{X(t)}\right) = \tau^2 - \eta^2.$$

Тогда получаем, что метрика (5.16), (5.17) в новых координатах имеет вид

$$ds^2 = A^2(t)(-d\tau^2 + d\eta^2) + B^2(t)(dy^2 + dz^2), \tag{5.18}$$

где

$$A(t) = \exp\left(-\frac{t+\Sigma}{\Sigma}\right)\left(\frac{9}{2}M(t+\Sigma)\right)^{-1/3}, \quad B(t) = \left(\frac{9}{2}M(t+\Sigma)\right)^{2/3}; \tag{5.19}$$

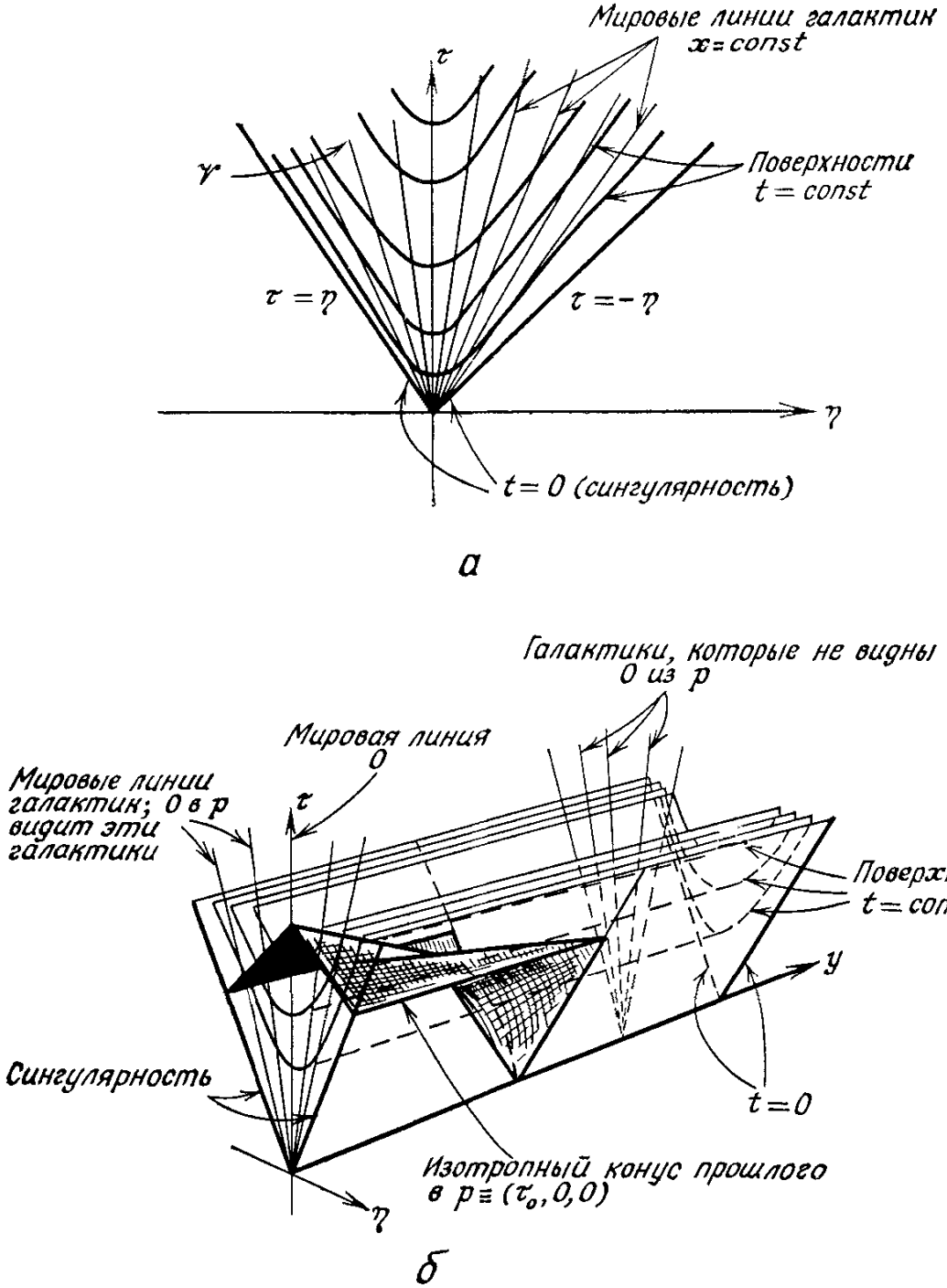
при этом все пространство (при $t > 0$) отображается на область $\mathscr{V}$, определяемую неравенствами $\tau > 0$, $\tau^2 - \eta^2 > 0$. Функция $t(\tau, \eta)$ теперь задана как решение уравнения

$$\tau^2 - \eta^2 = \frac{9}{2}Mt^2 \exp\frac{2(t+\Sigma)}{\Sigma}, \tag{5.20}$$

где $t > 0$. На плоскости (τ, η) введены конформно-плоские координаты. Проекция области $\mathscr{V}$, ограниченной поверхностью $t > 0$, на эту плоскость изображена на рис. 22. На этой диаграмме мировые линии частиц изображаются прямыми линиями, расходящимися от начала.

Функции $A(t)$, $B(t)$ непрерывны при $t \to 0$ сверху. Поэтому решения можно продолжить непрерывно из верхней полуплоскости на всю плоскость (τ, η), положив, что (5.19) выполняется всюду (5.20) — внутри $\mathscr{V}$ и что вне области $\mathscr{V}$ имеет место равенство

$$t(\tau, \eta) = 0.$$



Р и с. 22. Пространство типа I Бианки с пылевидной материей, обладающее «блинообразной» сингулярностью. *а* — плоскость (τ, η); изотропные линии проходят под углом ±45°; *б* — поперечный разрез пространства в координатах (τ, η, *y*) (координата *z* опущена), на котором виден световой конус прошлого точки $p = (\tau_0, 0, 0)$. В направлении оси *y* имеется горизонт частиц, но его нет в направлении оси *x* (т. е. η).

Тогда формула (5.18) дает C^0-метрику, которая внутри $\mathscr{V}$ эквивалентна (5.16), (5.17), а вне $\mathscr{V}$ является метрикой плоского пространства-времени. Однако это решение не принадлежит классу C^1 при переходе через границу $\mathscr{V}$, и плотность материи становится на этой границе бесконечной (поскольку на ней $S \to 0$). Из-за того, что первые производные квадратично-неинтегрируемы, уравнения Эйнштейна на этой границе невозможно трактовать даже в смысле обобщенных функций (см. разд. 8.4). Хотя продолжение решения на границу единственно, нет способа сделать его однородным вне области $\mathscr{V}$. Итак, мы получили продолжение для случая пыли; подобным же образом можно продолжить решение в случае смеси вещества и излучения.

Вернемся теперь к рассмотрению непустых пространственно-однородных моделей общего вида. Существование сингулярности в таких моделях прямо следует из уравнения Райчаудхури, если материя движется по геодезическим без вращения (как это будет, например, в случае ортогональности геодезических к поверхностям однородности и при выполнении условия времениподобного сближения). Однако существуют пространства, в которых вещество обладает ускорением и турбулентностью; возможно, какой-либо из этих двух факторов может воспрепятствовать появлению сингулярности. Следующий результат, являющийся усовершенствованным вариантом теоремы Хокинга и Эллиса [74], показывает, что в действительности ни ускорение, ни вращение не могут помешать существованию сингулярностей в этих моделях.

Теорема

Модель $(\mathscr{M}, g)$ не может быть времениподобно геодезически полной, если;

1) $R_{ab}K^aK^b > 0$ для всех времениподобных и изотропных векторов K (это справедливо, если тензор энергии-импульса типа I (разд. 4.3) и $\mu + p_i > 0$, $\mu + \sum_i p_i - 4\pi\Lambda > 0$);

2) существуют уравнения движения материальных полей, для которых задача Коши имеет единственное решение (см. гл. 7);

3) данные Коши на некоторой пространственноподобной 3-поверхности $\mathscr{H}$ инвариантны относительно группы диффеоморфизмов $\mathscr{H}$, которая транзитивна на $\mathscr{H}$.

Поскольку внутренняя геометрия $\mathscr{H}$ инвариантна относительно транзитивной группы диффеоморфизмов, последние являются изометриями и многообразие $\mathscr{H}$ полно, т. е. не может иметь края. Можно показать (см. разд. 6.5), что из существования непространственноподобной кривой, пересекающей $\mathscr{H}$ более одного раза, следует существование для $\mathscr{M}$ накрывающего

многообразия $\hat{\mathscr{M}}$, в котором каждая связная компонента образа поверхности $\mathscr{H}$ будет пересекаться не более одного раза с любой непространственноподобной кривой. Допустим, что $\hat{\mathscr{M}}$ — времениподобное геодезически полное многообразие, и покажем, что это противоречит условиям (1)—(3).

Пусть $\hat{\mathscr{H}}$ — связная компонента образа $\mathscr{H}$ в $\hat{\mathscr{M}}$. Согласно условию (3), данные Коши на $\hat{\mathscr{H}}$ однородны. Поэтому по условию (2) область Коши какой-либо области $\hat{\mathscr{H}}$ изометрична области Коши любой другой подобной области $\hat{\mathscr{H}}$. Вследствие этого однородны и поверхности $\{s = \text{const}\}$, где s есть расстояние от $\hat{\mathscr{H}}$, измеренное вдоль геодезических нормалей к $\hat{\mathscr{H}}$, если они лежат внутри области Коши поверхности $\hat{\mathscr{H}}$. Эти поверхности должны лежать или целиком внутри или целиком вне области Коши $\hat{\mathscr{H}}$, ибо в противном случае на $\hat{\mathscr{H}}$ существовали бы эквивалентные области, которые имели бы неэквивалентные области Коши. Поверхности $\{s = \text{const}\}$ будут лежать в области Коши постольку, поскольку они остаются пространственноподобными, ибо граница (если она существует) области Коши поверхности $\hat{\mathscr{H}}$ должна быть изотропной (разд. 6.5).

Геодезическая, ортогональная к $\hat{\mathscr{H}}$, будет ортогональной и к поверхностям $\{s = \text{const}\}$, поскольку вектор девиации точек, отстоящих на равном расстоянии от $\hat{\mathscr{H}}$ вдоль соседних геодезических, сохраняет ортогональность к этим геодезическим, если он был ортогонален к ним первоначально. Как и в разд. 4.1, пространственное разделение соседних геодезических, ортогональных к $\hat{\mathscr{H}}$, можно охарактеризовать матрицей $\mathbf{A}$, которая единична на $\hat{\mathscr{H}}$. Пока $\mathbf{A}$ не вырождена, отображение из $\hat{\mathscr{H}}$ на поверхности $\{s = \text{const}\}$, задаваемое нормальными геодезическими, будет ранга 3, так что эти поверхности будут пространственноподобными 3-поверхностями, находящимися в пределах области Коши $\hat{\mathscr{H}}$. Расхождение

$$\theta = (\det \mathbf{A})^{-1} \frac{\mathrm{d}(\det \mathbf{A})}{\mathrm{d}s}$$

этих геодезических подчиняется уравнению Райчаудхури (4.26) с нулевыми вращением и ускорением. По условию (1) $R_{ab}V^aV^b \geqslant 0$ для всех времениподобных векторов V^a. Следовательно, θ становится бесконечным, а матрица $\mathbf{A}$ — вырожденной при некотором конечном положительном или отрицательном значении s_0 параметра s. Отображение из $\hat{\mathscr{H}}$ на поверхность $s = s_0$ может иметь самое большее ранг 2, поэтому на $\hat{\mathscr{H}}$ найдется по крайней мере одно векторное поле $\mathbf{Z}$, для которого $\mathbf{AZ} = 0$. Интегральные кривые этого векторного поля являются теми кривыми на $\hat{\mathscr{H}}$,

которые отображаются геодезическими нормалями в одну точку на поверхности $s = s_0$. Таким образом, эта поверхность будет самое большее двумерной. Поскольку при $|s| < |s_0|$ геодезические лежат в области Коши поверхности $\hat{\mathscr{H}}$, поверхность $s = s_0$ будет лежать или внутри или на границе этой области. Согласно условию (1), тензор энергии-импульса имеет единственный времениподобный собственный вектор в каждой точке. Эти собственные векторы образуют времениподобное векторное C^1-поле, интегральные кривые которого можно рассматривать как линии тока вещества. Все линии тока, проходящие через $s = s_0$, должны пересекать $\hat{\mathscr{H}}$. Но тогда в силу однородности $\hat{\mathscr{H}}$ все линии тока, проходящие через $\hat{\mathscr{H}}$, должны проходить и через $s = s_0$. Следовательно, линии тока задают диффеоморфизм между $\hat{\mathscr{H}}$ и поверхностью $s = s_0$. Это невозможно, так как $\hat{\mathscr{H}}$ — трехмерная поверхность, а $s = s_0$ — двумерная. □

В самом деле, если все линии тока должны проходить через двумерную поверхность, то следует ожидать, что плотность вещества станет бесконечной. Таким образом, мы убедились, что крупномасштабные вращение или ускорение не могут сами по себе воспрепятствовать появлению сингулярностей во вселенной, подчиняющейся строгому принципу Коперника. Из последующих теорем мы увидим, что неоднородности также не могут запретить существование сингулярностей в моделях мира.

5.5. Решения Шварцшильда и Райсснера — Нордстрема

Хотя пространственно-однородные решения могут служить хорошими моделями для крупномасштабного распределения материи во Вселенной, они непригодны, например, для описания локальной геометрии пространства-времени в Солнечной системе. Эту геометрию в хорошем приближении можно описать решением Шварцшильда, которое представляет собой сферически-симметричное пустое пространство-время вне сферически-симметричного массивного тела. Все эксперименты по проверке различия между ньютоновской теорией и общей теорией относительности, выполненные до сих пор, фактически основаны на предсказаниях, связанных с этим решением.

Метрику пространства Шварцшильда можно представить в виде

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2\theta)\, d\phi^2, \tag{5.21}$$

где $r > 2m$. Очевидно, что это пространство-время — статическое (т. е. $\partial/\partial t$ представляет собой градиентный времениподобный вектор Киллинга) и сферически-симметричное, т. е.

инвариантно относительно группы изометрий $SO(3)$, действующей на 2-сферах $\{t = \text{const}, r = \text{const}\}$ (ср. с приложением Б). Координата r задана в этой метрической форме внутренним образом требованием, чтобы площадь этих поверхностей транзитивности была равна $4\pi r^2$. Решение Шварцшильда — асимптотически плоское, поскольку метрика при больших r имеет вид $g_{ab} = \eta_{ab} + O(1/r)$. Сравнение с ньютоновской теорией (ср. с разд. 3.4) показывает, что m имеет смысл гравитационной массы, измеренной на бесконечном удалении от тела, создающего данное поле. Следует подчеркнуть, что это решение единственно: если какое-либо решение уравнений поля в пустом пространстве сферически-симметрично, то локально оно изометрично решению Шварцшильда (хотя в какой-то другой координатной системе оно, конечно, может выглядеть совсем иначе; см. приложение Б, а также [6]).

Обычно метрику Шварцильда при r, превышающих некоторое значение $r_0 > 2m$, можно рассматривать как решение вне некоторого сферически-симметричного тела, внутри которого ($r < r_0$) метрика имеет иной вид, определяемый тензором энергии-импульса вещества этого тела. Однако интересно посмотреть, что получится, если рассматривать эту метрику как решение в пустом пространстве при всех значениях r.

Тогда метрика особенна при $r = 0$ и $r = 2m$ (имеются также тривиальные сингулярности полярных координат при $\theta = 0$ и $\theta = \pi$). Следовательно, мы должны вырезать точки, в которых $r = 0$ и $r = 2m$, из многообразия, задаваемого координатами (t, r, θ, ϕ), поскольку в разд. 3.1 мы приняли, что пространство-время должно представлять собой многообразие с лоренцевой метрикой. Вырезав поверхность $r = 2m$, мы разделим многообразие на две непересекающиеся компоненты, в которых $0 < r < 2m$ и $2m < r < \infty$. Поскольку мы приняли, что пространственно-временное многообразие должно быть связным, мы должны брать в рассмотрение только одну из компонент и, очевидно, выбрать при этом ту, в которой $r > 2m$, т. е. соответствующую внешнему полю. После этого мы должны задаться вопросом, расширяемо ли многообразие $\mathscr{M}$ со шварцшильдовой метрикой $\mathbf{g}$, т. е. существуют ли большее многообразие $\mathscr{M}'$, в которое можно было бы вложить $\mathscr{M}$, и нужное число раз дифференцируемая лоренцева метрика $\mathbf{g}'$ на $\mathscr{M}'$, совпадающая с $\mathbf{g}$ на образе $\mathscr{M}$. Ясно, что расширение можно осуществить в области, где $r \to 2m$. Вычисление показывает, что, несмотря на особенность метрики в шварцшильдовых координатах t, r, θ, ϕ, при $r = 2m$ ни один скалярный полином тензора кривизны и метрики не расходится при $r \to 2m$. Это наводит на мысль, что особенность при $r = 2m$ не является истинной физической сингулярностью, а скорее всего оказывается следствием неудачного выбора координат.

Чтобы подтвердить это и показать, что $(\mathscr{M}, \mathbf{g})$ можно расширить, введем

$$r^* \equiv \int \frac{dr}{1-2m/r} = r + 2m \ln (r - 2m).$$

Тогда

$$v \equiv t + r^*$$

— опережающая изотропная координата и

$$w \equiv t - r^*$$

— запаздывающая изотропная координата. Используя координаты (v, r, θ, ϕ), получим метрику $\mathbf{g}'$ в форме Эддингтона — Финкельстейна:

$$ds^2 = -(1 - 2m/r)\, dv^2 + 2\, dv\, dr + r^2 (d\theta^2 + \sin^2\theta\, d\phi^2). \quad (5.22)$$

Многообразием $\mathscr{M}$ является область $2m < r < \infty$, но метрика (5.22) не сингулярна и даже аналитична на большем многообразии $\mathscr{M}'$, для которого $0 < r < \infty$. Та область модели $(\mathscr{M}', \mathbf{g}')$, для которой $0 < r < 2m$, фактически изометрична области метрики Шварцшильда, для которой $0 < r < 2m$. Итак, используя другие координаты, т. е. вводя иное многообразие, мы расширили метрику Шварцшильда так, что она перестала быть особенной при $r = 2m$. Как видно из диаграмм Финкельстейна, в многообразии $\mathscr{M}'$ поверхность $r = 2m$ изотропна (рис. 23). Она является сечением $\theta = \text{const}$, $\phi = \text{const}$ пространства-времени, каждая ее точка изображает 2-сферу площади $4\pi r^2$. На этой диаграмме изображены некоторые изотропные конусы и радиальные изотропные геодезические; изображены также поверхности $\{t = \text{const}\}$, и видно, что t становится бесконечным на поверхности $r = 2m$.

Подобное представление решения Шварцшильда обладает тем необычным свойством, что оно не симметрично во времени. Этого можно было бы ожидать из-за наличия перекрестного члена $(dv\, dr)$ в (5.22); качественно асимметрия видна на диаграмме Финкельстейна. Ярче всего она проявляется в том, что поверхность $r = 2m$ ведет себя как полупроницаемая мембрана, пропуская направленные в будущее времениподобные или изотропные кривые только извне $(r > 2m)$ внутрь $(r < 2m)$. Ни одна направленная в прошлое времениподобная или изотропная кривая изнутри $r = 2m$ не может приблизиться к $r = 0$. Однако любая направленная в будущее времениподобная или изотропная кривая, пересекшая $r = 2m$, достигает $r = 0$ в пределах конечного аффинного расстояния. При $r \to 0$ скаляр $R^{abcd}R_{abcd}$ расходится как m^2/r^6. Поэтому $r = 0$ является истинной сингулярностью; пространство $(\mathscr{M}', \mathbf{g}')$ не имеет C^2-расширения через $r = 0$ (на самом деле оно не имеет даже C^0-расширения).

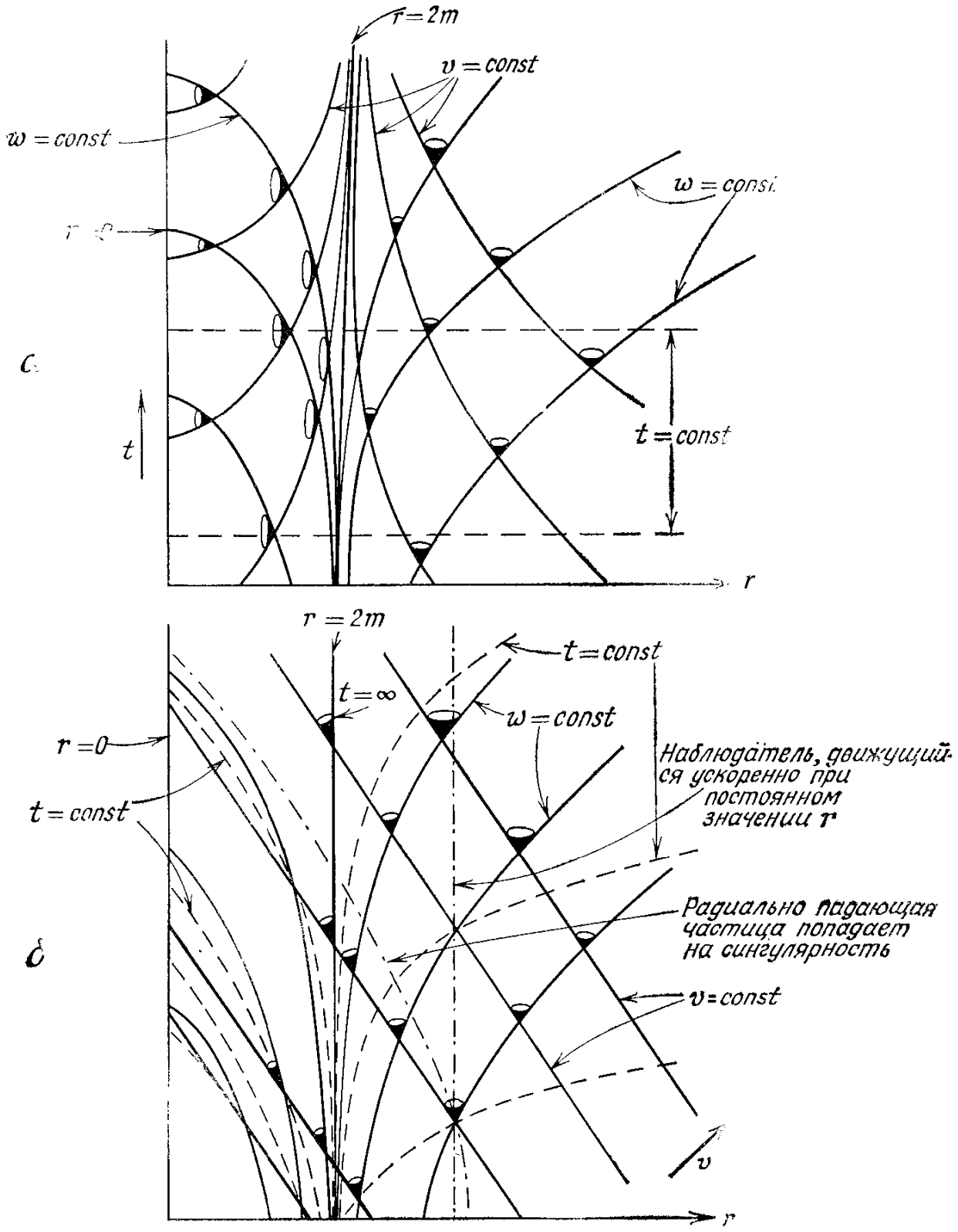


Рис. 23. Сечение постоянных θ и ϕ в решении Шварцшильда. *а* — кажущаяся сингулярность для $r = 2m$ при использовании координат (t, r); *б* — диаграмма Финкельстейна, полученная переходом к координатам (v, r) (линии под углом 45° являются линиями v). Поверхность $r = 2m$ — изотропная поверхность, на которой $t = \infty$.

Если вместо v использовать координаты w, то метрика примет вид g''

$$ds^2 = -(1-2m/r)\,dw^2 - 2\,dw\,dr + r^2(d\theta^2 + \sin^2\theta\,d\phi^2).$$

Она аналитична на многообразии $\mathcal{M}''$, задаваемом координатами (w, r, θ, ϕ) при $0 < r < \infty$. Снова многообразием $\mathcal{M}$ является область $2m < r < \infty$, и новая область $0 < r < 2m$ изометрична области $0 < r < 2m$ метрики Шварцшильда, но эта изометрия обращает направление времени. В многообразии $\mathcal{M}''$ поверхность $r = 2m$ снова является изотропной поверхностью, действующей как полунепроницаемая мембрана. Однако на этот раз она проницаема в противоположном направлении времени; пропускает извне ($r > 2m$) внутрь ($r < 2m$) только направленные в прошлое времениподобные или изотропные кривые.

На самом деле можно осуществить оба расширения $(\mathcal{M}', g')$ и $(\mathcal{M}'', g'')$ одновременно, т. е. существует еще многообразие $\mathcal{M}^*$ с метрикой g^*, в которое и $(\mathcal{M}', g')$, и $(\mathcal{M}'', g'')$ могут быть изометрично погружены так, что они совпадут в области $r > 2m$, которая изометрична пространству $(\mathcal{M}, g)$. Такое расширенное многообразие было построено Крускалом [95]. Чтобы получить его, рассмотрим $(\mathcal{M}, g)$ в координатах (v, w, θ, ϕ); тогда метрика имеет вид

$$ds^2 = -(1-2m/r)\,dv\,dw + r^2(d\theta^2 + \sin^2\theta\,d\phi^2),$$

где r определяется равенством

$$\frac{1}{2}(v-w) = r + 2m\ln(r-2m).$$

Таким образом, 2-пространство $\theta = \text{const}$, $\phi = \text{const}$ представлено в изотропных конформно-плоских координатах, поскольку пространство с метрикой $ds^2 = dv\,dw$ плоское. Самое общее координатное преобразование, оставляющее метрику этого 2-пространства конформно-плоской, а обе координаты изотропными, имеет вид $v' = v'(v)$, $w' = w'(w)$, где v' и w' — произвольные C^1-функции. Получающаяся при этом метрика имеет вид

$$ds^2 = -\left(1-\frac{2m}{r}\right)\frac{dv}{dv'}\frac{dw}{dw'}\,dv'\,dw' + r^2(d\theta^2 + \sin^2\theta\,d\phi^2).$$

Чтобы привести ее к виду, который соответствует тому, что было получено ранее для пространства-времени Минковского, положим

$$x' = \frac{1}{2}(v'-w'), \quad t' = \frac{1}{2}(v'+w')$$

и придем к следующему окончательному выражению:

$$ds^2 = F^2(t', x')(-dt'^2 + dx'^2) + r^2(t', x')(d\theta^2 + \sin^2\theta\,d\phi^2). \qquad (5.23)$$

Конкретный вид этой метрики определяется выбором функций v', w'. Крускал выбрал

$$v' = \exp(v/4m), \quad w' = \exp -(-w/4m).$$

Тогда для r имеем уравнение

$$(t')^2 - (x')^2 = -(r - 2m)\exp(r/2m), \tag{5.24}$$

а для F — формулу

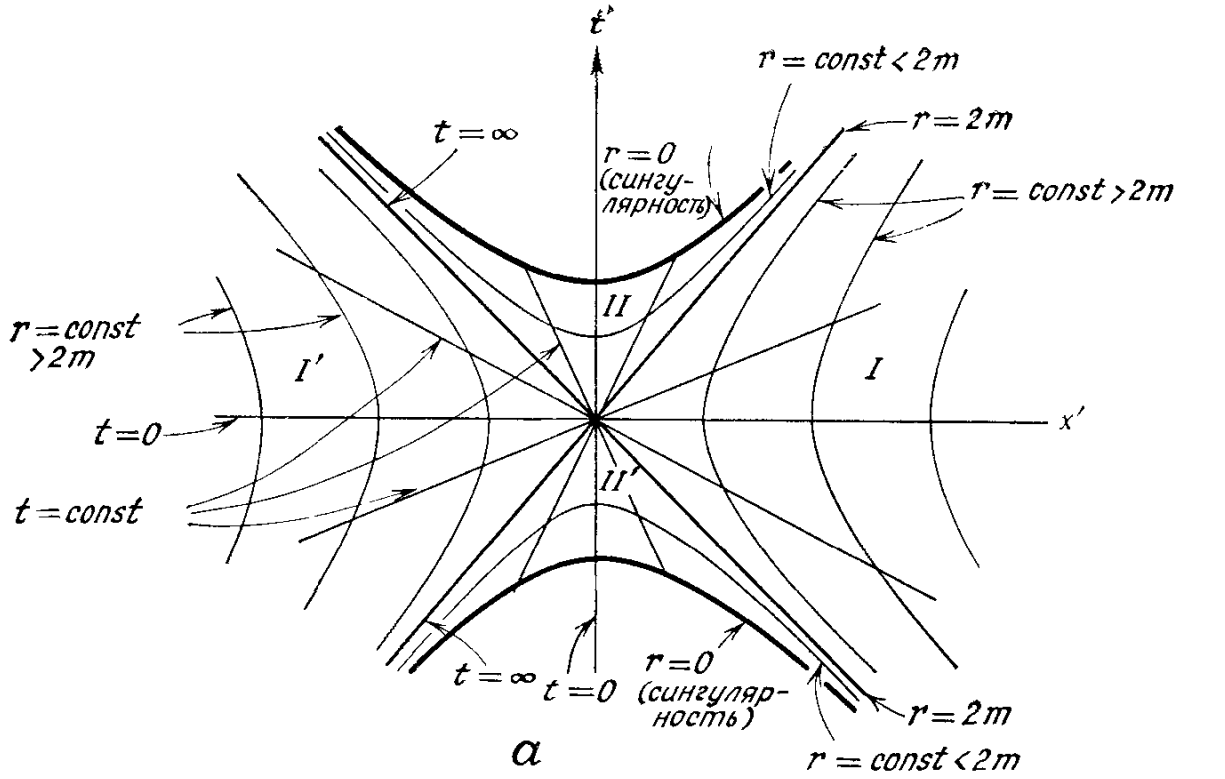
$$F^2 = \exp\left(-\frac{r}{2m}\right) \cdot 16\frac{m^2}{r}. \tag{5.25}$$

На многообразии $\mathscr{M}^*$, определяемом координатами (t', x', θ, ϕ) при $(t')^2 - (x')^2 < 2m$, функции r и F [задаваемые (5.24), (5.25)] положительны и аналитичны. Пусть метрика $\mathbf{g}^*$ дана формулой (5.23); область I пространства $(\mathscr{M}^*, \mathbf{g}^*)$, определенная неравенством $x' > |t'|$, изометрична $(\mathscr{M}, \mathbf{g})$, т. е. области решения Шварцшильда, для которой $r > 2m$. Область $x' > -t'$ (области I и II на рис. 24) изометрична опережающему расширению Финкельстейна $(\mathscr{M}', \mathbf{g}')$. Аналогично, область $x' > t'$ (области I и II' на рис. 24) изометрична запаздывающему расширению Финкельстейна $(\mathscr{M}'', \mathbf{g}'')$. Имеется также область I', определяемая неравенством $x' < -|t'|$, которая тоже оказывается изометричной внешнему решению Шварцшильда. Ее можно рассматривать как другую асимптотически-плоскую вселенную по ту сторону шварцшильдовской «горловины». (Рассмотрим сечение $t = 0$; 2-сферы $\{r = \text{const}\}$ выглядят при больших r так же, как в евклидовом пространстве; однако при малых r их площадь убывает к минимальному значению $16\pi m^2$ и затем вновь возрастает, словно эти 2-сферы расширяются в другом асимптотически-плоском 3-пространстве). Из рис. 24 видно, что области I' и II изометричны опережающему, а области I' и II' — запаздывающему расширению Финкельстейна области I'. Из области I в область I' не проходит ни одна времениподобная или изотропная кривая. Все направленные в будущее времениподобные и изотропные кривые, которые пересекают часть поверхности $r = 2m$, описываемую теперь уравнением $t' = |x'|$, стремятся к сингулярности при $t' = (2m + (x')^2)^{1/2}$, для которой $r = 0$. Подобно этому направленные в прошлое времениподобные и изотропные кривые, пересекающие $t' = -|x'|$, стремятся к другой сингулярности при $t' = -(2m + (x')^2)^{1/2}$, где снова $r = 0$.

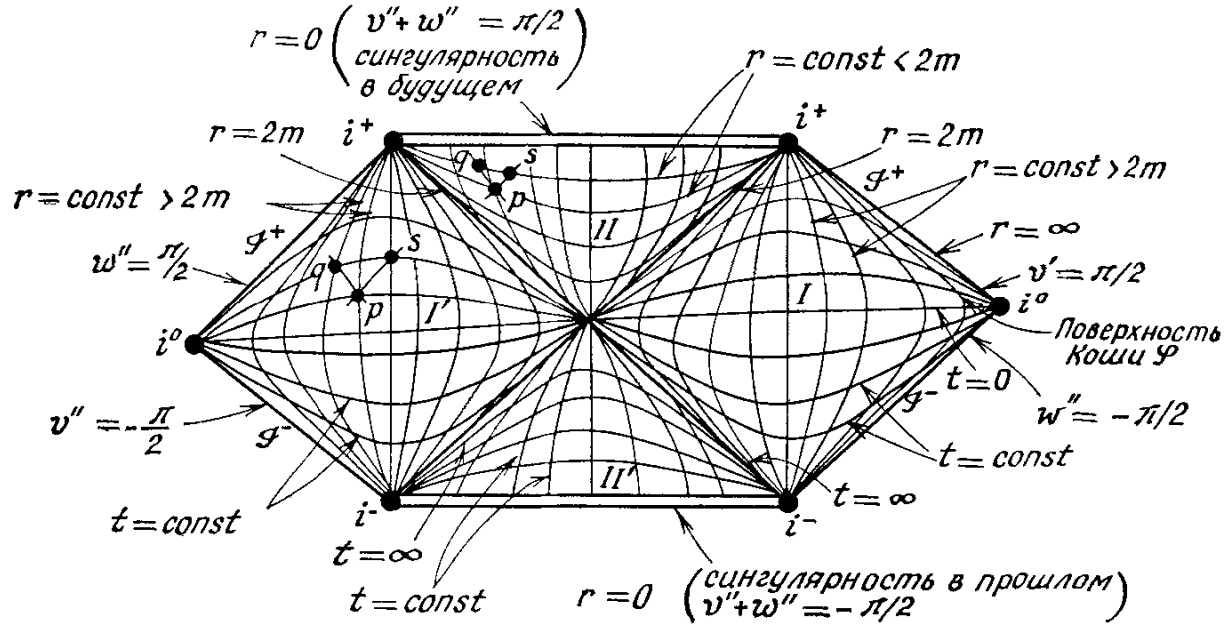
Расширение Крускала $(\mathscr{M}^*, \mathbf{g}^*)$ — единственное аналитическое и локальное нерасширяемое расширение решения Шварцшильда. Вводя новые опережающие и запаздывающие изотропные координаты

$$v'' = \operatorname{arctg}\left(v'(2m)^{-1/2}\right), \quad w'' = \operatorname{arctg}\left(w'(2m)^{-1/2}\right),$$

$$-\pi < v'' + w'' < \pi, \quad -\frac{1}{2}\pi < v'' < \frac{1}{2}\pi, \quad -\frac{1}{2}\pi < w'' < \frac{1}{2}\pi,$$



а



б

Р и с. 24. Максимальное аналитическое расширение решения Шварцшильда. Координаты θ, ϕ опущены; изотропные линии проходят под углом $\pm 45°$. Поверхности $\{r = \text{const}\}$ однородны. *а* — диаграмма Крускала, на которой изображены асимптотически плоские области I и I' и области II, II', для которых $r < 2m$; *б* — диаграмма Пенроуза, изображающая конформную бесконечность и две сингулярности.

можно построить диаграмму Пенроуза для расширения Крускала (рис. 24, *б*). Ее можно сравнить с диаграммой для пространства Минковского (см. рис. 15, *б*). Теперь мы имеем бесконечности будущего, прошлого и изотропную бесконечность для каждой из асимптотически-плоских областей I и I'. В отличие от пространства Минковского конформная метрика теперь непрерывна, но недифференцируема в точках i^0.

Если мы рассмотрим изотропный конус будущего какой-либо точки вне $r = 2m$, то радиальная геодезическая, направленная по радиусу наружу, достигает бесконечности, а направленная внутрь достигает сингулярность в будущем; если точка лежит внутри $r = 2m$, то обе эти геодезические попадают на сингулярность, и все будущее этой точки кончается этой сингулярностью. Таким образом, любая точка вне $r = 2m$ может избежать сингулярности (и, следовательно, сингулярность не «универсальна» в отличие от пространства Робертсона — Уокера), но как только частица оказалась внутри $r = 2m$, она обязательно достигнет сингулярности. Оказывается, этот факт тесно связан с тем обстоятельством, что каждая точка внутри области II изображает 2-сферу, которая является замкнутой ловушечной поверхностью. Это означает следующее: рассмотрим какую-либо 2-сферу p (на рис. 24 она изображается точкой) и две 2-сферы q и s, образованные фотонами, которые испущены с поверхности p в один и тот же момент времени от центра и к центру поверхности p. Площадь q (она равна $4\pi r^2$) будет больше, а площадь s будет меньше площади p, если все три сферы лежат в области $r > 2m$. Если же все они лежат в области II, где $r < 2m$, то площади *обеих* сфер: и q, и s будут меньше площади p (на рис. 24 значение r убывает в области II при движении снизу вверх). В таком случае мы говорим, что p является замкнутой ловушечной поверхностью. Каждая точка внутри области II' является замкнутой ловушечной поверхностью при обращении времени (существование замкнутых ловушечных поверхностей — необходимое следствие того обстоятельства, что поверхности $r = \text{const}$ пространственноподобны), и соответственно все частицы в области II должны быть «пришельцами» из сингулярности в прошлом. В гл. 8 мы увидим, что само существование сингулярностей тесно связано с существованием замкнутых ловушечных поверхностей.

Решение Райсснера — Нордстрема описывает пространство-время вне электрически заряженного сферически-симметричного тела (которое, однако, не обладает ни спином, ни магнитным дипольным моментом, вследствие чего это решение не может служить хорошим описанием поля тяготения вокруг электрона). Следовательно, тензор энергии-импульса такой же, как у электромагнитного поля, созданного зарядом этого тела. Решение

Райсснера — Нордстрема — единственное асимптотически-плоское сферически-симметричное решение уравнений Эйнштейна — Максвелла и локально оно напоминает решение Шварцшильда: существуют координаты, в которых метрика имеет вид

$$ds^2 = -\left(1 - \frac{2m}{r} + \frac{e^2}{r^2}\right) dt^2 + \left(1 - \frac{2m}{r} + \frac{e^2}{r^2}\right)^{-1} dr^2 + \\ + r^2 (d\theta^2 + \sin^2\theta\, d\phi^2), \quad (5.26)$$

где m — гравитационная масса и e — электрический заряд тела. Это асимптотически-плоское решение естественно рассматривать как решение только вне тела, а внутри должна быть какая-то другая метрика; но интересно посмотреть, что получится, если мы будем считать, что решение имеет смысл для всех r.

Если $e^2 > m^2$, метрика неособенна везде, за исключением неустранимой особенности при $r = 0$, которую можно рассматривать как точечный заряд — источник поля. Если $e^2 \leqslant m^2$, метрика имеет особенности также в r_+ и r_-, где $r_\pm = m \pm (m^2 - e^2)^{1/2}$; она регулярна в областях $\infty > r > r_+$, $r_+ > r > r_-$ и $r_- > r > 0$ (если $e^2 = m^2$, существуют только первая и третья области). Как и в случае Шварцшильда, эти сингулярности можно устранить введением подходящих координат и расширением многообразия до получения максимального аналитического многообразия [21, 64]. Главные отличия возникают из-за того, что множитель перед dt^2 имеет два нуля, а не один, как в случае Шварцшильда. В частности, это приводит к тому, что первая и третья области — обе статические, тогда как вторая область (если она существует) — пространственно-однородна, но не статична.

Чтобы получить максимально расширенное многообразие, мы будем действовать аналогично шварцшильдову случаю. Введем координату r^*:

$$r^* = \int \frac{dr}{1 - \frac{2m}{r} + \frac{e^2}{r^2}},$$

тогда для $r > r_+$

$$r^* = r + \frac{r_+^2}{r_+ - r_-} \ln(r - r_+) - \frac{r_-^2}{r_+ - r_-} \ln(r - r_-) \quad \text{при} \quad e^2 < m^2,$$

$$r^* = r + m \ln(r - m)^2 - \frac{2}{r - m} \quad \text{при} \quad e^2 = m^2,$$

$$r^* = r + m \ln(r^2 - 2mr + e^2) + \frac{2}{e^2 - m^2} \operatorname{arctg} \frac{r - m}{e^2 - m^2} \quad \text{при} \quad e^2 > m^2.$$

Вводя опережающие и запаздывающие координаты v, w

$$v = t + r^*, \quad w = t - r^*,$$

приведем метрику (5.26) к изотропному виду

$$ds^2 = -\left(1 - \frac{2m}{r} + \frac{e^2}{r^2}\right) dv\, dw + r^2 (d\theta^2 + \sin^2\theta\, d\phi^2). \quad (5.27)$$

В случае $e^2 < m^2$ определим новые координаты v'', w'' следующим образом:

$$v'' = \operatorname{arctg} \exp\left(\frac{r_+ - r_-}{4r_+^2} v\right), \quad w'' = \operatorname{arctg}\left(-\exp\left(\frac{-r_+ + r_-}{4r_+^2} w\right)\right).$$

Тогда метрика (5.27) принимает вид

$$ds^2 = \left(1 - \frac{2m}{r} + \frac{e^2}{r^2}\right) 64 \frac{r_+^4}{(r_+ - r_-)^2} \operatorname{cosec} 2v'' \operatorname{cosec} 2w''\, dv''\, dw'' + \\ + r^2 (d\theta^2 + \sin^2\theta\, d\phi^2), \quad (5.28)$$

где r определяется неявно уравнением

$$\operatorname{tg} v'' \operatorname{tg} w'' = -\exp\left(\frac{r_+ - r_-}{2r_+^2} r\right)(r - r_+)^{1/2} (r - r_-)^{-\alpha/2}$$

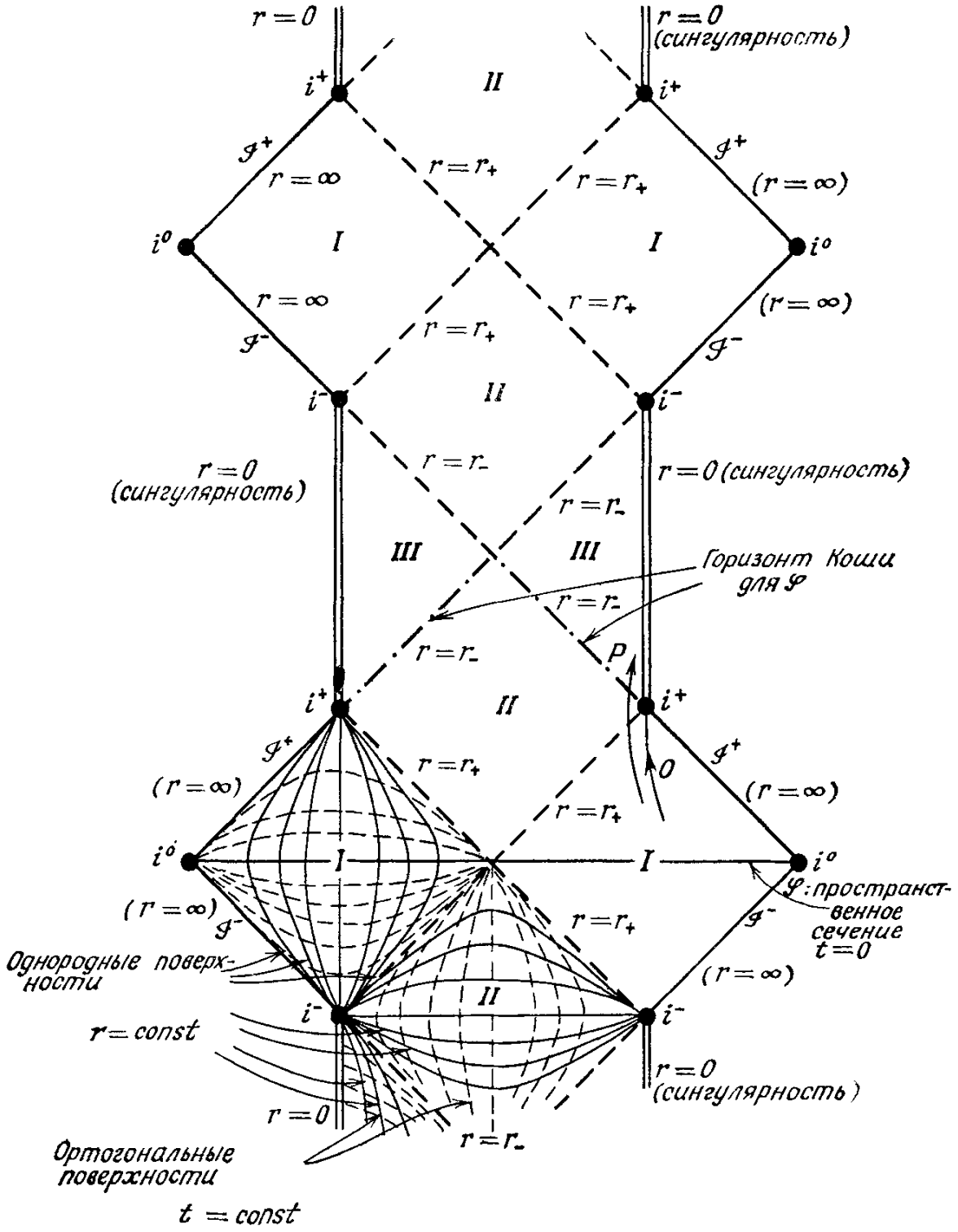
и $\alpha = (r_+)^{-2}(r_-)^2$. Максимальное расширение мы получим, если примем (5.28) в качестве метрики $\mathbf{g}^*$, а в качестве $\mathscr{M}^*$ — максимальное многообразие, на котором $\mathbf{g}^*$ принадлежит классу C^2.

Диаграмма Пенроуза этого максимального расширения приведена на рис. 25. Имеется бесконечное число асимптотически-плоских областей, где $r > r_+$; они обозначены цифрой *I*. Эти области соединены промежуточными областями *II* и *III*, где $r_+ > r > r_-$ и $r_- > r > 0$ соответственно. По-прежнему в каждой области *III* имеется неустранимая сингулярность при $r = 0$, но в отличие от решения Шварцшильда она времениподобна, и направленная в будущее времениподобная кривая из области *I*, пересекающая поверхность $r = r_+$, может избежать этой сингулярности. Такая кривая может пройти через области *II*, *III* и *II* и появиться в другой асимптотически-плоской области *I*. Возникает захватывающая возможность совершить путешествие в другие вселенные через «ходы», проделанные зарядами. К сожалению, скорее всего такой путешественник не сможет вернуться обратно в нашу Вселенную и рассказать о том, что же он видел «по ту сторону».

Метрика (5.28) аналитична всюду, за исключением поверхности $r = r_-$, на которой она по меньшей мере класса C^2. Можно ввести другие координаты v''' и w''':

$$v''' = \operatorname{arctg} \exp\left(\frac{r_+ - r_-}{2nr_-^2} v\right),$$

$$w''' = \operatorname{arctg}\left(\left(-\exp\right)\frac{-r_+ + r_-}{2nr_-^2} w\right),$$



Р и с. 25. Диаграмма Пенроуза для максимально расширенного решения Райснера — Нордстрема ($e^2 < m^2$). Бесконечная цепь асимптотически плоских областей I ($\infty > r > r_+$) соединяется областями II ($r_+ > r > r_-$) и III ($r_- > r > 0$); каждая область III ограничена времениподобной сингулярностью при $r = 0$.

где n — целое число, причем $n \geqslant 2(r_+)^2(r_-)^2$. В этих координатах метрика аналитична всюду, за исключением $r = r_+$, где она по меньшей мере класса C^2. Координаты v''', w''' — аналитические функции координат v'' и w'' всюду вне поверхностей $r = r_+$ и $r = r_-$. Таким образом, многообразие $\mathscr{M}^*$ может быть покрыто аналитическим атласом, состоящим из локальных координатных окрестностей, определяемых координатами v'', w'' при $r \neq r_-$ и координатами v''', w''' при $r \neq r_+$. В этом атласе метрика аналитична.

В случае $e^2 = m^2$ расширение осуществляется подобным же образом; в случае $e^2 > m^2$ модель нерасширяема уже в исходных координатах. Диаграммы Пенроуза для этих двух случаев приведены на рис. 26.

Во всех этих случаях сингулярность времениподобна. Это означает, что в отличие от решения Шварцшильда времениподобные и изотропные кривые всегда могут избежать встречи с сингулярностью. Фактически это выглядит так, словно сингулярности отталкивают частицы: времениподобные геодезические не могут столкнуться с сингулярностью, хотя это может произойти с радиальными изотропными геодезическими и негеодезическими времениподобными кривыми. Следовательно, эти пространства времениподобны (но не изотропны) и геодезически полны. Времениподобный характер сингулярности означает также, что в этих пространствах нет поверхностей Коши: для любой данной пространственноподобной поверхности можно найти времениподобную или изотропную кривую, которая достигает сингулярности, не пересекая эту поверхность. Например, в случае $e^2 < m^2$ можно найти пространственноподобную поверхность $\mathscr{S}$, которая пересекает две асимптотически-плоские области I (см. рис. 25). Она будет поверхностью Коши для этих двух областей I и соседних областей II; но в соседних более поздних областях III существуют направленные в прошлое непродолжимые времениподобные и изотропные кривые, которые подходят к сингулярности, но не пересекают поверхность $r = r_-$. Поэтому можно сказать, что эта поверхность представляет собой горизонт Коши будущего для $\mathscr{S}$. Расширение решения за пределы $r = r_-$ не определяется данными Коши на $\mathscr{S}$. Продолжение, которое мы произвели, является единственным локально нерасширяемым аналитическим продолжением, но существуют другие, неаналитические C^∞-продолжения, удовлетворяющие уравнениям Эйнштейна-Максвелла.

Наблюдатель O, чья мировая линия остается вне $r = r_+$ и продолжается до бесконечности будущего i^+ (см. рис. 25), наблюдал бы частицу P, пересекающую поверхность $r = r_+$, с бесконечным красным смещением. В области II между $r = r_+$ и $r = r_-$ поверхности $r = \text{const}$ пространственноподобны и, следо-

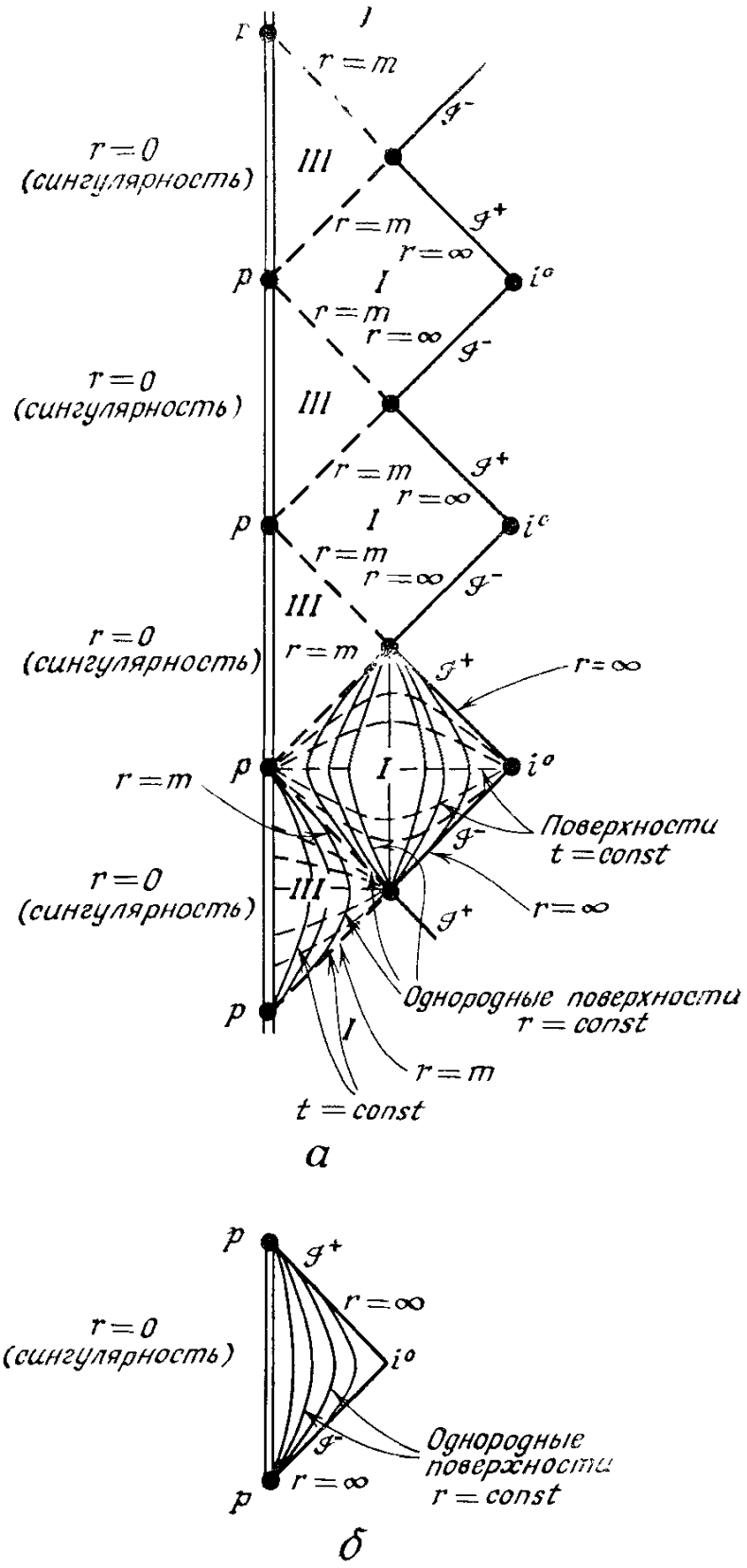


Рис. 26. Диаграмма Пенроуза для максимально расширенных решений Райснера — Нордстрема: $e^2 = m^2$ (*а*), $e^2 > m^2$ (*б*). В первом случае имеется бесконечная цепь областей I ($\infty > r > m$), связанных областями III ($m > r > 0$). Точки p не являются частью сингулярности $r = 0$, а представляют собой исключительные точки на бесконечности.

вательно, каждая точка рисунка изображает 2-сферу, которая является замкнутой ловушечной поверхностью. Наблюдатель P, пересекающий поверхность $r = r_-$, обозрел бы всю историю одной из асимптотически плоских областей за конечное время. Объекты этой области поэтому были бы видны с бесконечным фиолетовым смещением по мере приближения к i^+. Это дает основание думать, что поверхность $r = r_-$, возможно, неустойчива относительно малых возмущений начальных данных на пространственноподобной поверхности $\mathscr{S}$ и что такие возмущения, вообще говоря, приведут к сингулярности на $r = r_-$.

5.6. Решение Керра

Обычно астрономические объекты вращаются, и поэтому нельзя считать, что поле вне их в точности сферически-симметрично. Решение Керра — единственное известное точное решение, которое может служить для описания стационарного осесимметричного асимптотически-плоского поля вне массивного вращающегося объекта. Оно будет внешним решением только для массивных вращающихся полей с конкретным набором мультипольных моментов; для тел с другими мультипольными моментами внешние решения будут иными. Однако решение Керра является, по-видимому, единственно возможным внешним решением для черных дыр (см. разд. 9.2 и 9.3).

Это решение можно записать в координатах Бойера и Линдквиста (r, θ, ϕ, t), в которых метрика имеет вид

$$ds^2 = \rho^2\left(\frac{dr^2}{\Delta} + d\theta^2\right) + (r^2 + a^2)\sin^2\theta\, d\phi^2 - dt^2 + \\ + \frac{2mr}{\rho^2}(a\sin^2\theta\, d\phi - dt^2)^2, \quad (5.29)$$

где

$$\rho^2(r, \theta) \equiv r^2 + a^2\cos^2\theta \quad \text{и} \quad \Delta(r) \equiv r^2 - 2mr + a^2,$$

m и a — постоянные, причем m является массой, а ma — моментом количества движения, которые как бы измерены в бесконечности [16]; при $a = 0$ это решение переходит в шварцшильдово. Ясно, что эта метрика инвариантна относительно одновременного обращения t и ϕ, т. е. относительно преобразования $t \to -t$, $\phi \to -\phi$, хотя она и не инвариантна относительно обращения одного только t (за исключением случая $a = 0$). Этого следовало ожидать, поскольку при обращении времени вращающийся объект меняет направление своего вращения на противоположное.

Если $a^2 > m^2$, то $\Delta > 0$, и приведенная метрика особенна только при $r = 0$. Эта сингулярность в действительности является не точкой, а кольцом; в этом можно убедиться при пере-

ходе к координатам Керра — Шильда $(x, y, z, \bar{t})$, где

$$x + iy = (r + ia)\sin\theta \exp i\int (d\phi + a\Delta^{-1}dr),$$

$$z = r\cos\theta, \quad \bar{t} = \int (dt + (r^2 + a)\Delta^{-1}dr) - r.$$

В этих координатах метрика принимает вид

$$ds^2 = dx^2 + dy^2 + dz^2 - d\bar{t}^2 +$$

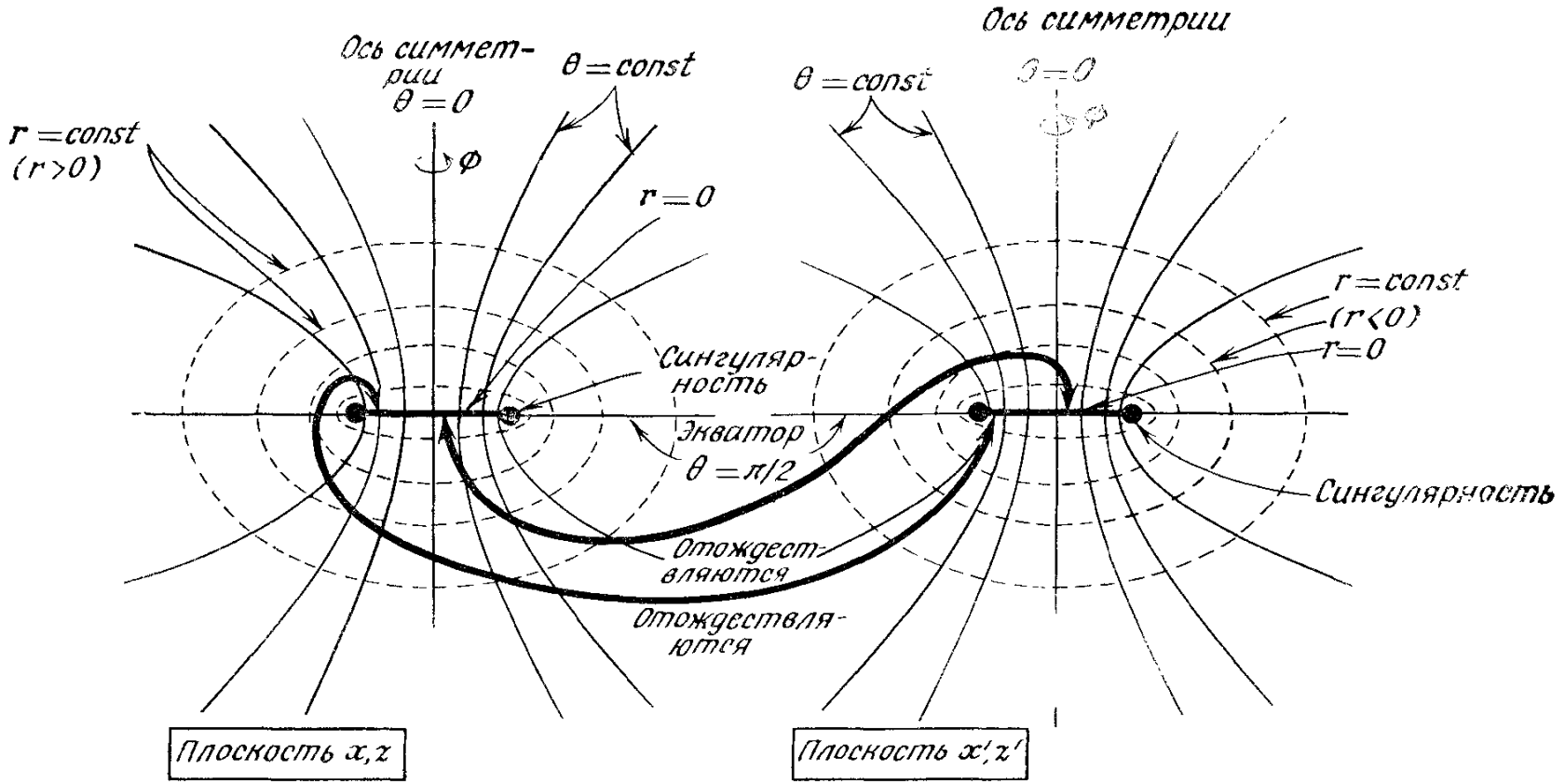
$$+ \frac{2mr^3}{r^4 + a^2z^2}\left[\frac{r(x\,dx + y\,dy) - a(x\,dy - y\,dx)}{r^2 + a^2} + \frac{z\,dz}{r} + d\bar{t}\right]^2, \qquad (5.30)$$

где r как функция x, y, z определяется с точностью до знака уравнением

$$r^4 - (x^2 + y^2 + z^2 - a^2)r^2 - a^2z^2 = 0.$$

При $r \neq 0$ поверхности $\{r = \text{const}\}$ на плоскости x, y, z являются конфокальными эллипсоидами, вырождающимися при $r = 0$ в диск $z^2 + y^2 \leqslant a^2$, $z = 0$. Кольцо $x^2 + y^2 = a^2$, $z = 0$ этого диска является истинной сингулярностью кривизны, поскольку на нем расходится скалярный полином $R_{abcd}R^{abcd}$. Но на самом диске (вне кольца) любой скалярный полином конечен. Чтобы получить максимальное аналитическое расширение этого решения, можно аналитически продолжить функцию r от положительных к отрицательным значениям через открытый диск $x^2 + y^2 < a^2$, $z = 0$.

Для этого присоединим другую плоскость, определяемую координатами x', y', z', так что точки на верхней стороне диска $x^2 + y^2 < a^2$, $z = 0$ в плоскости (x, y, z) отождествляются с точками с теми же координатами x и y на нижней стороне соответствующего диска в плоскости (x', y', z'). Аналогично, точки с нижней стороны диска в плоскости (x, y, z) отождествляются с точками с верхней стороны диска в плоскости (x', y', z') (рис. 27). Метрика (5.30) очевидным образом распространяется на это расширенное многообразие. В области (x', y', z') она попрежнему имеет вид (5.29), но значения r не положительны, а отрицательны. При больших отрицательных r пространство снова асимптотически-плоское, но теперь масса отрицательна. При малых отрицательных r вблизи кольцевой сингулярности вектор $\partial/\partial\phi$ времениподобен, так что окружности $t = \text{const}$, $r = \text{const}$, $\theta = \text{const}$ являются замкнутыми времениподобными кривыми. Эти кривые можно деформировать так, чтобы они проходили через любую точку расширенного пространства [23]. Решение Керра геодезически неполно ввиду кольцевой сингулярности; однако ее достигают только те времениподобные и изотропные геодезические, которые лежат на экваториальной плоскости в той части, где $r > 0$ [23].



Р и с. 27. Максимальное расширение решения Керра при $a^2 > m^2$ получается отождествлением верхней поверхности диска $x^2 + y^2 < a^2$, $z = 0$ в плоскости (x, y, z) с нижней поверхностью соответствующего диска в плоскости (x', y', z'), и наоборот. На рисунке приведены сечения $y = 0$ и $y' = 0$ этих плоскостей. Двойной обход вокруг сингулярности $x^2 + y^2 = a^2$, $z = 0$ переводит из плоскости (x, y, z) в плоскость (x', y', z') (где $r < 0$) и обратно в плоскость (x, y, z), где $r > 0$.

Расширение в случае $a^2 < m^2$ несколько сложнее из-за существования двух значений r: $r_+ = m + (m^2 - a^2)^{1/2}$ и $r_- = m - (m^2 - a^2)^{1/2}$, при которых $\Delta(r)$ обращается в нуль. Эти поверхности подобны поверхностям $r = r_+$ и $r = r_-$ в решении Райсснера — Нордстрема. Для продолжения метрики через эти поверхности перейдем к координатам Керра (r, θ, ϕ_+, u_+), где

$$du_+ = dt + (r^2 + a^2)\Delta^{-1} dr, \quad d\phi_+ = d\phi + a\Delta^{-1} dr.$$

Тогда на многообразии, определяемом этими координатами, метрика имеет вид

$$\begin{aligned} ds^2 = \rho^2 d\theta^2 - 2a \sin^2\theta\, dr\, d\phi_+ + 2\, dr\, du_+ + \\ + \rho^{-2}[(r^2 + a^2)^2 - \Delta a^2 \sin^2\theta] \sin^2\theta\, d\phi_+^2 - \\ - 4a\rho^{-2} mr \sin^2\theta\, d\phi_+\, du_+ - (1 - 2mr\rho^{-2})\, du_+^2, \end{aligned} \quad (5.31)$$

она аналитична при $r = r_+$ и $r = r_-$. Снова мы имеем сингулярность при $r = 0$, которая имеет ту же форму кольца и описанную выше геодезическую структуру. Метрику Керра можно расширить и на другое многообразие, определенное координатами (r, θ, ϕ_-, u_-), где

$$du_- = dt - (r^2 + a^2)\Delta^{-1} dr, \quad d\phi_- = d\phi - a\Delta^{-1} dr,$$

а метрика имеет вид (5.31) с заменой ϕ_+, u_+ на $-\phi_-$, $-u_-$. Максимальное аналитическое расширение получается комбинацией этих двух расширений, как в случае Райсснера — Нордстрема [15, 23]. Его глобальная структура очень похожа на структуру решения Райсснера — Нордстрема, за тем исключением, что теперь возможно продолжение через кольцо к отрицательным значениям r. На рис. 28, *а* показана конформная структура решения вдоль оси симметрии. Области *I* соответствуют асимптотически-плоские области, в которых $r > r_+$. Области *II* $(r_- < r < r_+)$ содержат замкнутые ловушечные поверхности. Области *III* $(-\infty < r < r_-)$ содержат кольцевую сингулярность, в любой области *III* через каждую точку можно провести замкнутую времениподобную кривую, в других двух областях (*I* и *II*) нет никаких нарушений причинности.

В случае $a^2 = m^2$, r_+ и r_- совпадают и области *II* отсутствуют. Максимальное расширение похоже на расширение решения Райсснера — Нордстрема при $e^2 = m^2$. О конформной структуре вдоль оси симметрии для этого случая можно судить по рис. 28, *б*.

Решение Керра, будучи стационарным и осесимметричным, имеет двухпараметрическую группу изометрий; она необходимо абелева [26]. Следовательно, существуют два независимых коммутирующих векторных поля Киллинга. Имеется только одна их

линейная комбинация K^a, которая времениподобна при произвольно больших положительных и отрицательных r. Кроме того, можно составить линейную комбинацию $\tilde{K}^a$ этих полей, равную

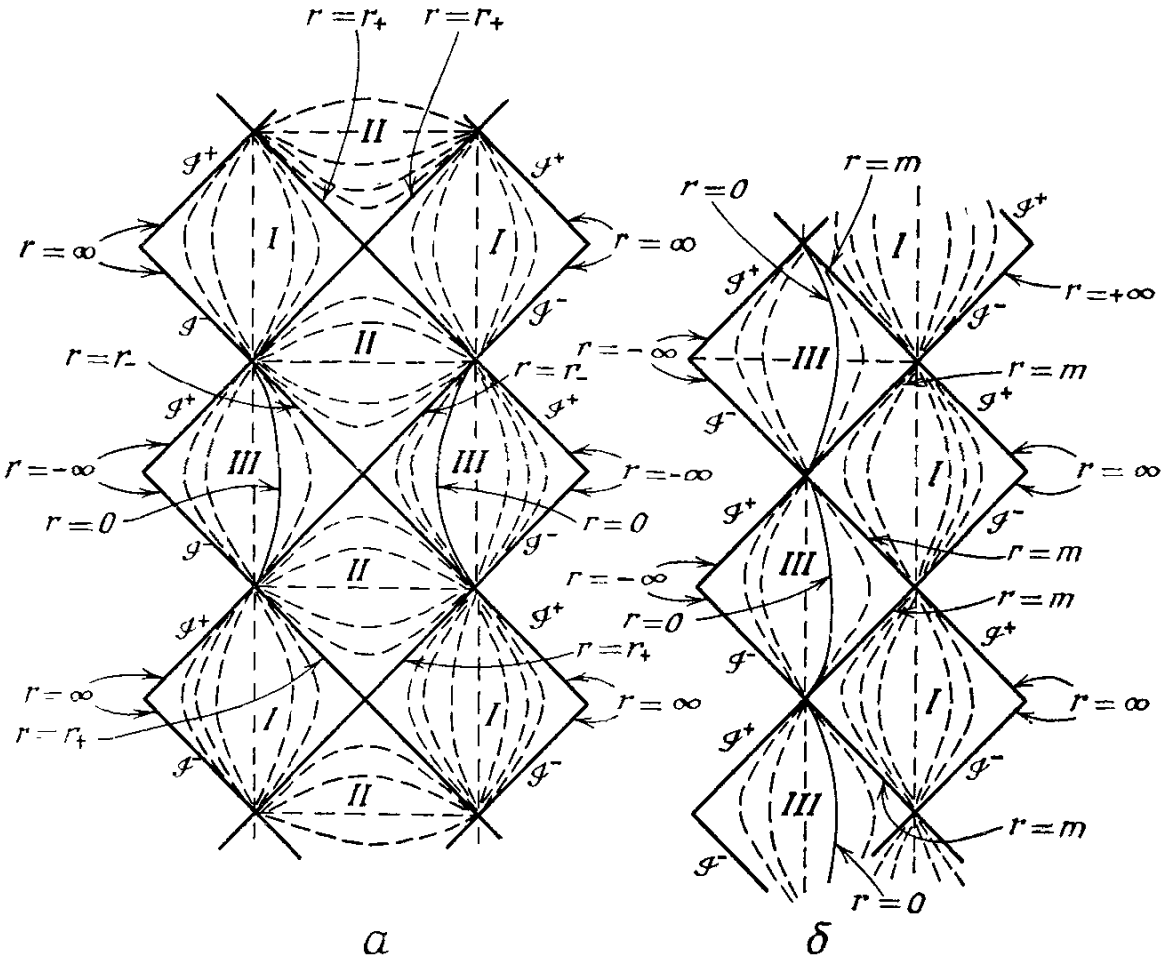
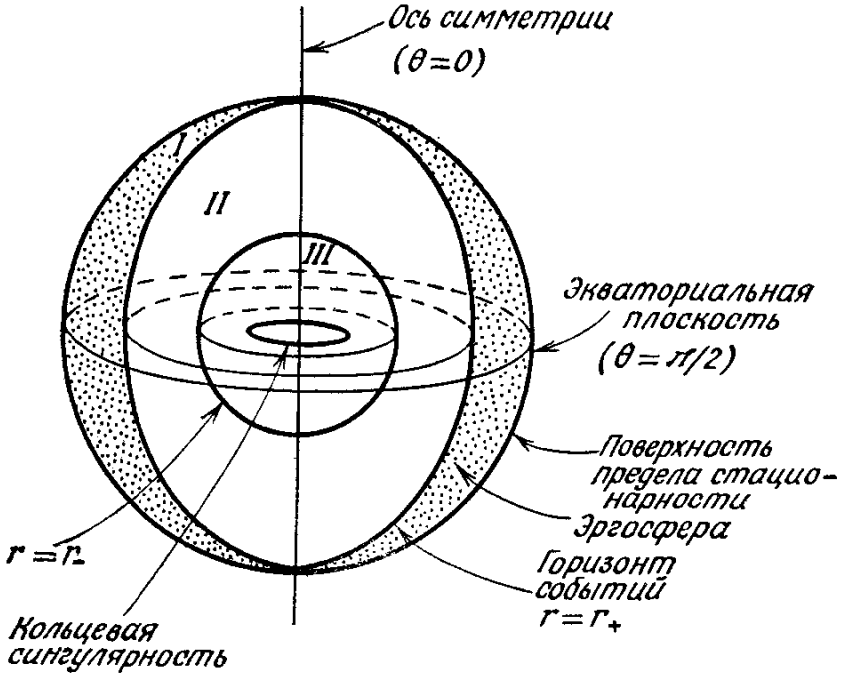


Рис. 28. Конформная структура решения Керра вдоль оси симметрии $0 < a^2 < m$ (*а*); $a^2 = m^2$ (*б*). Пунктирные линии соответствуют линиям $r = \text{const}$; области *I*, *II* и *III* в случае (*а*) разделяются поверхностями $r = r_+$ и $r = r_-$, а области *I* и *III* в случае (*б*) — поверхностью $r = m$. В обоих случаях структура пространства вблизи кольцевой сингулярности такая же, как на рис. 27.

нулю на оси симметрии. Орбиты вектора Киллинга K^a определяют станционарную систему отсчета: объект, движущийся вдоль одной их этих орбит, из бесконечности выглядит станционарным. Орбиты вектора Киллинга $\tilde{K}^a$ представляют собой замкнутые кривые, что соответствует вращательной симметрии решения Керра.

В решениях Шварцшильда и Райсснера — Нордстрема вектор Киллинга, времениподобный при больших r, времениподобен всюду в области *I* и становится изотропным на поверхностях $r = 2m$ и $r = r_+$ соответственно. Эти поверхности изотропны. Отсюда следует, что частица, которая пересекает одну из этих

поверхностей, в направлении будущего не может вернуться в прежнюю область. Эти поверхности являются границей той области решения, из которой частицы могут достичь бесконечности $\mathscr{I}^+$ данной области I, и они называются *горизонтами событий* этой $\mathscr{I}^+$. Для наблюдателя, движущегося в области I по какой-либо орбите вектора K^a, они являются фактически горизонтами событий в смысле, установленном в разд. 5.2.



Р и с. 29. Решение Керра при $0 < a^2 < m^2$; эргосфера лежит между предельной поверхностью стационарности и горизонтом $r = r_+$. Частицы могут уйти на бесконечность из области I (вне горизонта событий $r = r_+$), но не из областей II (между $r = r_+$ и $r = r_-$) и III ($r < r_-$; эта область содержит кольцевую сингулярность).

Наоборот, для решения Керра вектор Киллинга K^a пространственноподобен в некоторой области вне $r = r_+$, называемой *эргосферой* (рис. 29). Внешней границей этой области является поверхность $r = m + (m^2 - a^2 \cos\theta)^{1/2}$, на которой K^a изотропен. Она называется *предельной поверхностью стационарности*, так как является границей области, в которой частица, движущаяся по времениподобной кривой, может находиться на орбите вектора Киллинга и таким образом оставаться в покое относительно бесконечности. Предельная поверхность стационарности времениподобна, за исключением двух точек на оси, в которых она изотропна (в этих точках она соприкасается с поверхностью $r = r_+$). Там, где она времениподобна, ее могут пересекать и входящие, и выходящие частицы. Поэтому она не является горизонтом событий для $\mathscr{I}^+$. Горизонтом событий в действительности является поверхность $r = r_+ = m + (m^2 - a^2)^{1/2}$. Из рис. 30

видно, почему это так. На нем изображена экваториальная плоскость $\theta = \pi/2$; каждая точка на этом рисунке представляет собой орбиту вектора Киллинга K^a, т. е. стационарна относительно $\mathscr{I}^+$. Малые круги изображают положение фронта светового

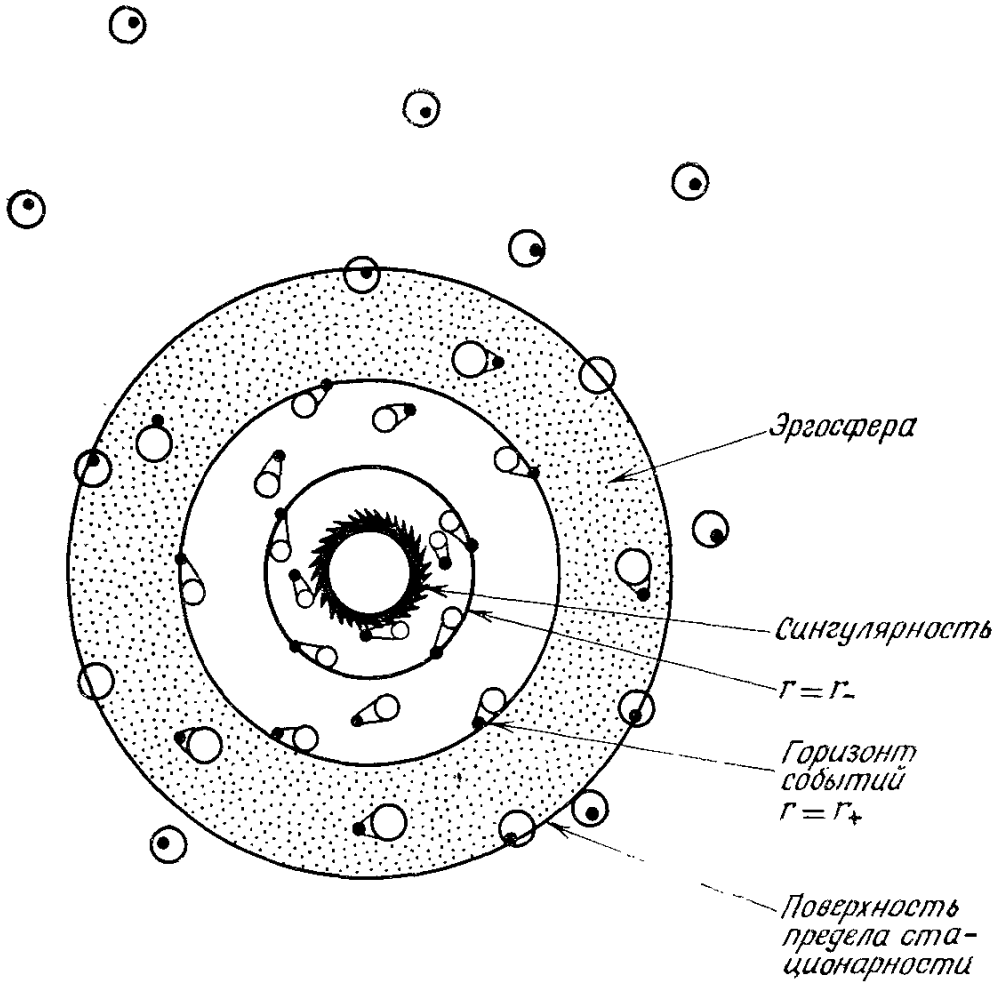


Рис. 30. Экваториальная плоскость решения Керра при $m^2 > a^2$. Круги изображают положение через короткое время вспышек света, излученных источниками, которые обозначены жирными точками.

импульса через короткое время после излучения из точек, изображенных черными кружками. Вне предельной поверхности стационарности вектор Киллинга K^a времениподобен, т. е. лежит внутри светового конуса. Это означает, что точка на рис. 30, изображающая орбиту точки вспышки, лежит внутри фронта волны света.

На предельной поверхности стационарности вектор K^a изотропен, и точка, изображающая орбиту точки вспышки, лежит на фронте волны. Однако фронт волны лежит частью вне, а частью внутри предельной поверхности стационарности, поэтому частица, движущаяся по времениподобной кривой, может с этой поверхности уйти в бесконечность. В эргосфере между предельной

поверхностью стационарности и поверхностью $r=r_+$ вектор Киллинга K^a пространственноподобен, и точка, изображающая орбиту испускания, лежит вне фронта волны. В этой области времениподобное или изотропное движение не может происходить по орбите вектора Киллинга и поэтому частица не может оставаться в покое относительно бесконечности. Однако положения волновых фронтов таковы, что частицы по-прежнему могут пересечь предельную поверхность стационарности и уйти на бесконечность. На поверхности $r = r_+$ вектор K^a по-прежнему пространственноподобен, но фронт волны, расходящийся от частицы на этой поверхности, лежит целиком внутри этой поверхности. Это означает, что частица, движущаяся по времениподобной кривой из какой-либо точки на или внутри поверхности $r = r_+$, не может выйти из области, охватываемой этой поверхностью, и достичь бесконечности. Поэтому поверхность $r = r_+$ является горизонтом событий для $\mathscr{I}^+$ и изотропна.

Хотя вектор Киллинга K^a пространственноподобен в эргосфере, модуль $K^a\tilde{K}^bK_{[a}\tilde{K}_{b]}$ *бивектора Киллинга* $K_{[a}\tilde{K}_{b]}$ отрицателен всюду вне $r = r_+$, за исключением оси $\tilde{K}^a = 0$, где он равен нулю. Поэтому на K^a и $\tilde{K}^a$ натягивается 2-поверхность и, следовательно, в каждой точке вне $r = r_+$, за исключением оси, имеется линейная комбинация K^a и $\tilde{K}^a$, которая времениподобна. Таким образом, решение в эргосфере в некотором смысле локально стационарно, хотя и не станционарно относительно бесконечности. Одной линейной комбинации K^a и $\tilde{K}^a$, которая была бы времениподобна всюду вне $r = r_+$, не существует. Величина бивектора Киллинга на $r = r_+$ обращается в нуль и положительна внутри этой поверхности. На $r = r_+$ оба вектора K^a и $\tilde{K}^a$ пространственноподобны, но существует их линейная комбинация, изотропная всюду на этой поверхности [25].

Рассмотренные нами свойства эргосферы и горизонта событий будут играть важную роль при рассмотрении черных дыр в разд. 9.2 и 9.3.

Подобно тому как решения Райсснера — Нордстрема можно рассматривать как решение Шварцшильда при наличии заряда, имеется семейство аналогичных решений Керра [23]. Их глобальные свойства весьма похожи на свойства решения Керра без заряда.

5.7. Модель Вселенной Гёделя

В 1949 г. Курт Гёдель опубликовал статью [62], которая заметно стимулировала изучение точных решений, более сложных, чем рассмотренные выше. Он дал точное решение уравнений Эйнштейна с материей в виде идеальной жидкости без давления. ($T_{ab} = \rho u_a u_b$, где ρ — плотность материи и u_a — нормированный

вектор 4-скорости.) Многообразием $\mathscr{M}$ является R^4, а метрика записывается в виде

$$ds^2 = -dt^2 + dx^2 - \frac{1}{2}\exp(2\sqrt{2}\,\omega x)\,dy^2 + dz^2 - 2\exp(\sqrt{2}\,\omega x)\,dt\,dy,$$

где $\omega > 0$ — постоянная. Уравнения поля удовлетворяются, если $\mathbf{u} = \partial/\partial x^0$ (т. е. $u^a = \delta^a{}_0$) и

$$4\pi\rho = \omega^2 = -\Lambda.$$

Постоянная ω фактически является угловой скоростью вращения вектора потока u^a.

Это пространство-время обладает 5-параметрической транзитивной группой изометрий, т. е. это пространство-время полностью однородно. (Группа транзитивна на $\mathscr{M}$, если она отображает любую точку $\mathscr{M}$ на любую другую точку $\mathscr{M}$.) Данная метрика является прямой суммой метрики $\mathbf{g}_1$:

$$ds_1^2 = -dt^2 + dx^2 - \frac{1}{2}\exp(2\sqrt{2}\,\omega x)\,dy^2 - 2\exp(\sqrt{2}\,\omega x)\,dt\,dy$$

на многообразии $\mathscr{M}_1 = R^3$, определяемом координатами (t, x, y), и метрики $\mathbf{g}_2$:

$$ds_2^2 = dz^2$$

на многообразии $\mathscr{M}_2 = R^1$, определяемом координатой z. Чтобы описать свойства этого решения, достаточно рассмотреть $(\mathscr{M}_1, \mathbf{g}_1)$.

Введем на $\mathscr{M}_1$ новые координаты (t', r, ϕ):

$$\exp(\sqrt{2}\,\omega x) = \operatorname{ch} 2r + \cos\phi \operatorname{sh} 2r,$$

$$\omega y \exp(\sqrt{2}\,\omega x) = \sin\phi \operatorname{sh} 2r,$$

$$\operatorname{tg}\tfrac{1}{2}(\phi + \omega t - \sqrt{2}\,t') = \exp(-2r)\operatorname{tg}\tfrac{1}{2}\phi;$$

тогда метрика $\mathbf{g}_1$ принимает вид

$$ds_1^2 = 2\omega^{-2}(-dt'^2 + dr^2 - (\operatorname{sh}^4 r - \operatorname{sh}^2 r)\,d\phi^2 + 2\sqrt{2}\operatorname{sh}^2 r\,d\phi\,dt'),$$

где $-\infty < t' < \infty$, $0 \leqslant r < \infty$ и $0 \leqslant \phi \leqslant 2\pi$, причем $\phi = 0$ отождествляется с $\phi = 2\pi$; вектор потока в этих координатах равен $\mathbf{u} = (\omega/\sqrt{2})\,\partial/\partial t'$. В этом выражении для метрики видна вращательная симметрия этого решения относительно оси $r = 0$. Путем того или иного выбора координат можно любую линию тока вещества сделать осью симметрии.

Поведение модели $(\mathscr{M}_1, \mathbf{g}_1)$ иллюстрирует рис. 31. Световые конусы на оси $r = 0$ содержат направление $\partial/\partial t'$ (вертикальное направление на диаграмме), но не содержат горизонтальных направлений $\partial/\partial r$ и $\partial/\partial\phi$. По мере удаления от оси световые конусы раскрываются и поворачиваются в направлении координатной линии ϕ, так что при радиусе $r = r_0 = \ln(1 + \sqrt{2})$ $\partial/\partial\phi$ является изотропным вектором, а окружность радиуса r_0 вокруг начала оказывается замкнутой изотропной кривой. При больших

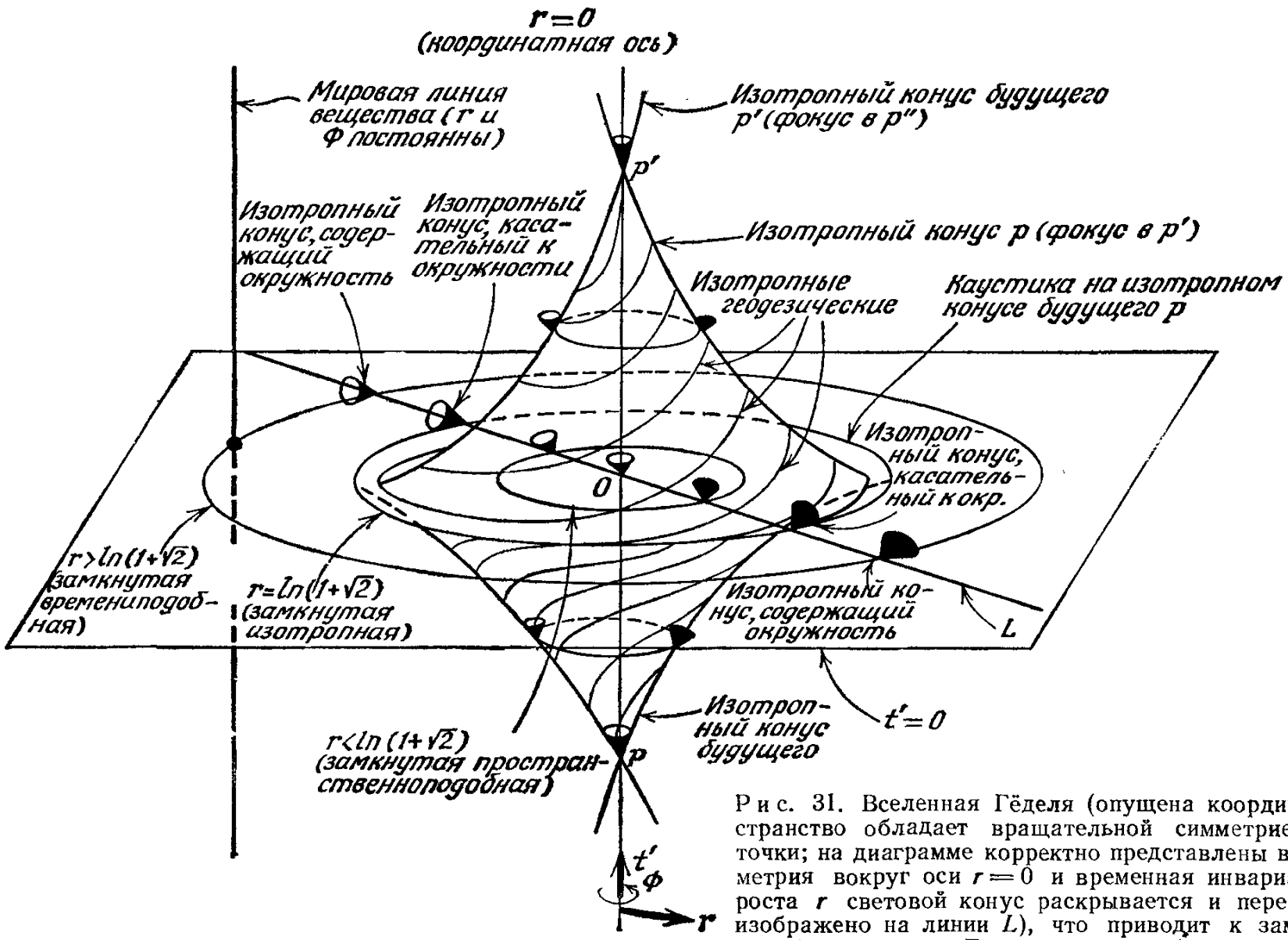


Рис. 31. Вселенная Гёделя (опущена координата z). Это пространство обладает вращательной симметрией вокруг любой точки; на диаграмме корректно представлены вращательная симметрия вокруг оси $r = 0$ и временная инвариантность. По мере роста r световой конус раскрывается и переворачивается (это изображено на линии L), что приводит к замкнутым времениподобным кривым. Диаграмма не отображает наглядно тот факт, что в действительности все точки пространства эквивалентны.

значениях r $\partial/\partial\phi$ становится времениподобным вектором, и окружности постоянных r и t' представляют собой замкнутые времениподобные кривые. Поскольку $(\mathscr{M}_1, \mathbf{g}_1)$ обладает 4-параметрической транзитивной группой изометрий, то получается, что через каждую точку $(\mathscr{M}_1, \mathbf{g}_1)$, а следовательно, и через каждую точку решений Гёделя $(\mathscr{M}, \mathbf{g})$ проходит замкнутая времениподобная геодезическая.

Все это дает основание думать, что решение Гёделя не имеет физического смысла. Существование замкнутых времениподобных кривых означает, что в $\mathscr{M}$ нет вложенных трехмерных поверхностей без края, которые были бы всюду пространственноподобны, ибо замкнутая времениподобная кривая, пересекшая такую поверхность, пересекает ее нечетное число раз. Это означает, что кривая не может быть непрерывным образом стянута в точку: непрерывная деформация может изменить число пересечений лишь на четное число. Это в свою очередь противоречит тому, что многообразие $\mathscr{M}$, будучи гомеоморфно R^4, односвязно. Существование замкнутых времениподобных линий означает также невозможность в $\mathscr{M}$ какой-либо космической координаты времени t, которая возрастала бы вдоль любой времениподобной или изотропной кривой, направленной в будущее.

Решение Гёделя геодезически полно. Поведение геодезических может быть описано с помощью разбиения на $(\mathscr{M}_1, \mathbf{g}_1)$ и $(\mathscr{M}_2, \mathbf{g}_2)$. Поскольку метрика $\mathbf{g}_2$ на $\mathscr{M}_2$ плоская, компонента в $\mathscr{M}_2$ вектора, касательного к геодезической, постоянна, т. е. координата z зависит линейно от аффинного параметра на геодезической. Поэтому достаточно описать поведение геодезических в $(\mathscr{M}_1, \mathbf{g}_1)$. Изотропные геодезические, проходящие через точку p на оси координат (рис. 31), первоначально расходятся (у оси), доходят до каустики $r = \ln(1+\sqrt{2})$ и затем сходятся к точке p' на оси. Поведение времениподобной геодезической аналогично: они достигают некоторого максимального значения r, меньшего чем $\ln(1+\sqrt{2})$, и снова сходятся к p'. Точку q с $r > \ln(1+\sqrt{2})$ можно соединить с p некоторой времениподобной кривой, но не времениподобной или изотропной геодезической.

Дальнейшие детали свойств решения Гёделя можно найти в работах [62; 96].

5.8. Пространство Тауба — НУТ

В 1951 г. Тауб получил пространственно-однородное решение уравнений Эйнштейна в пустом пространстве с топологией $R \times S^3$ и метрикой

$$ds^2 = -U^{-1}\,dt^2 + (2l)^2 U\,(d\psi + \cos\theta\,d\phi)^2 + \\ + (t^2 + l^2)(d\theta^2 + \sin^2\theta\,d\phi^2), \quad (5.32)$$

где

$$U(t) \equiv -1 + \frac{2(mt + l^2)}{t^2 + l^2},$$

m и l — положительные постоянные. Здесь θ, ϕ, ψ — координаты Эйлера на S^3 и, следовательно, $0 \leqslant \psi \leqslant 4\pi$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$. Эта метрика особенна при $t = t_\pm = (m \pm (m^2 + l^2)^{1/2}$, где $U = 0$. Ее можно продолжить через эти поверхности и получить пространство, найденное Ньюменом, Тамбурино и Унти [115], но прежде чем перейти к обсуждению этого расширения, рассмотрим простой двумерный пример с похожими свойствами, принадлежащий Мизнеру [105]. Это пространство обладает топологией $S^1 \times R^1$, а метрика $\mathbf{g}$ определяется формулой

$$ds^2 = -t^{-1}\,dt^2 + t\,d\psi^2,$$

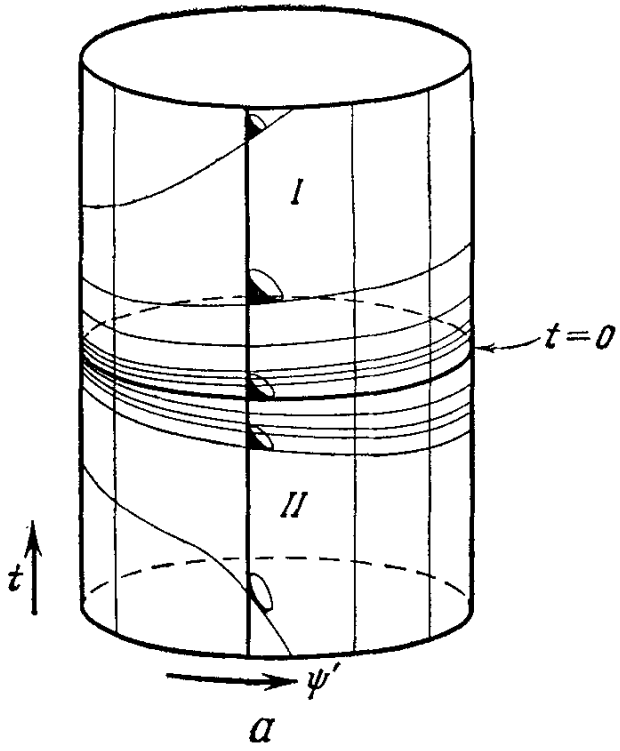
где $0 \leqslant \psi \leqslant 2\pi$. Эта метрика особенна при $t = 0$. Если принять, что многообразие $\mathscr{M}$ определяется координатами ψ и t, причем $0 < t < \infty$, то $(\mathscr{M}, \mathbf{g})$ можно расширить, вводя $\psi' = \psi - \ln t$. Тогда получаем метрику $\mathbf{g}'$:

$$ds^2 = +2\,d\psi'\,dt + t\,(d\psi')^2.$$

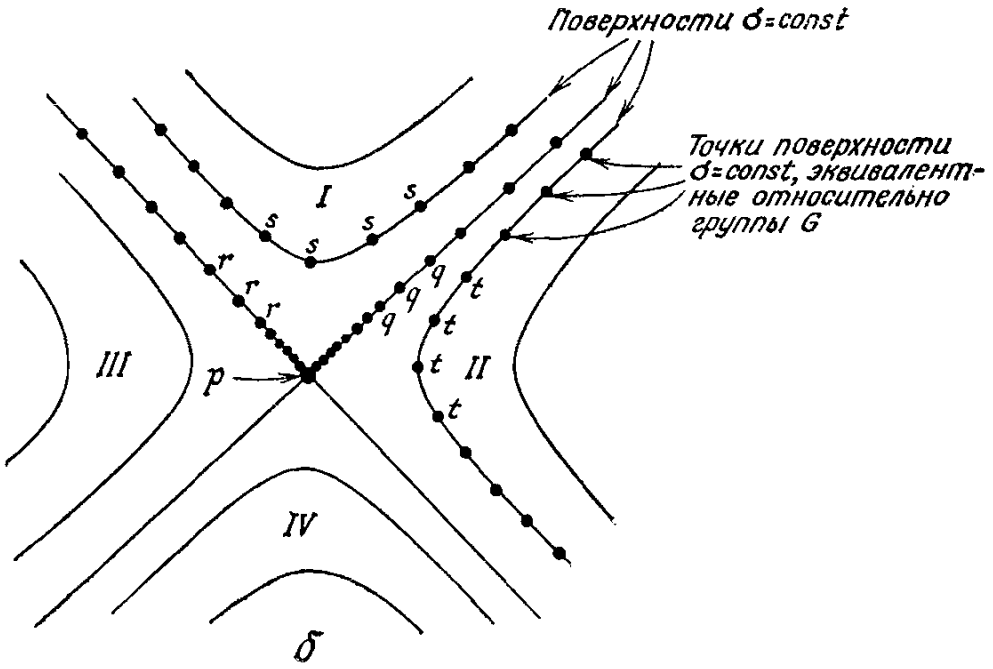
Метрика $\mathbf{g}'$ аналитична на многообразии $\mathscr{M}'$ с топологией $S^1 \times R^1$, которое определяется координатами ψ' и t, причем $-\infty < t < \infty$. Область $t > 0$ пространства $(\mathscr{M}', \mathbf{g}')$ изометрична $(\mathscr{M}, \mathbf{g})$. Основные черты пространства $(\mathscr{M}', \mathbf{g}')$ можно понять из рис. 32. В области $t < 0$ имеются замкнутые времениподобные линии, но их нет там, где $t > 0$. Одно семейство изотропных геодезических изображается на рис. 32 вертикальными линиями; они пересекают поверхность $t = 0$. Другое семейство приближается к $t = 0$ по спирали, но никогда не пересекает эту поверхность. Аффинные длины этих геодезических всегда конечны. Таким образом, расширение $(\mathscr{M}', \mathbf{g}')$ не симметрично относительно двух семейств изотропных геодезических, хотя исходное пространство $(\mathscr{M}, \mathbf{g})$ было симметричным. Однако можно задать другое расширение $(\mathscr{M}'', \mathbf{g}'')$, в котором эти два семейства меняются ролями. Для этого определим ψ'' как $\psi'' = \psi + \ln t$. Интервал, определяющий метрику $\mathbf{g}''$, равен

$$ds^2 = -2\,d\psi''\,dt + t\,(d\psi'')^2.$$

Метрика $\mathbf{g}''$ аналитична на многообразии $\mathscr{M}''$ с топологией $S^1 \times R^1$, определяемом ψ'' и неравенством $-\infty < t < \infty$. Область $t > 0$ пространства $(\mathscr{M}'', \mathbf{g}'')$ изометрична $(\mathscr{M}, \mathbf{g})$. Определив так ψ'', мы в некотором смысле «раскрутили» второе семейство изотропных геодезических так, чтобы они превратились в вертикальные линии, и теперь их можно продолжить через $t = 0$. Однако при этом мы «намотали» первое семейство, так



а



б

Рис. 32. Двумерный пример Мизнера. *а* — расширение области *I* за границу $t = 0$, в область *II*. Вертикальные изотропные геодезические полны, но закручивающиеся — нет; *б* — универсальное накрывающее пространство является пространством Минковского. Относительно дискретной подгруппы G группы Лоренца точки s эквивалентны. Аналогично, точки p, q, r эквивалентны. Рисунок (*а*) получается отождествлением эквивалентных точек в областях *I* и *II*.

что его геодезические стали спиральными, и их уже нельзя продолжить через $t = 0$. Таким образом, мы имеем два неэквивалентных локально нерасширяемых аналитических расширения пространства $(\mathscr{M}, g)$, причем обе геодезических неполны. Связь между этими двумя расширениями хорошо видна при переходе к накрывающему пространству пары $(\mathscr{M}, g)$.

Оказывается, что таким накрывающим пространством является область I двумерного пространства Минковского $(\tilde{\mathscr{M}}, \tilde{\eta})$, содержащаяся внутри светового конуса будущего точки p (рис. 32, *б*). Изометрии $(\tilde{\mathscr{M}}, \tilde{\eta})$, оставляющие неподвижной точку p, образуют однопараметрическую группу (группа Лоренца для метрики $\tilde{\eta}$), орбитами которой являются гиперболы $\{\sigma = \text{const}\}$, где $\sigma \equiv \tilde{t}^2 - \tilde{x}^2$, а $\tilde{t}$, $\tilde{x}$ — обычные координаты Минковского. Пространство $(\mathscr{M}, g)$ есть фактор-пространство пространства $(I, \tilde{\eta})$ по дискретной подгруппе G группы Лоренца, которая состоит из A^n (n целое), причем A отображает $(\tilde{t}, \tilde{x})$ в

$$(\tilde{t}\,\mathrm{ch}\,\pi + \tilde{x}\,\mathrm{sh}\,\pi,\ \tilde{x}\,\mathrm{ch}\,\pi + \tilde{t}\,\mathrm{sh}\,\pi),$$

иначе говоря, отождествляются точки

$$(\tilde{t}\,\mathrm{ch}\,n\pi + \tilde{x}\,\mathrm{sh}\,n\pi,\ \tilde{x}\,\mathrm{ch}\,n\pi + \tilde{t}\,\mathrm{sh}\,n\pi)$$

при всех целых n, и в $\mathscr{M}$ они соответствуют точке

$$t = \frac{1}{4}(\tilde{t}^2 - \tilde{x}^2), \qquad \psi = \mathrm{arth}\left(\frac{\tilde{x}}{\tilde{t}}\right).$$

Действие группы изометрии G в области I — собственно разрывное. Действие группы H на многообразии $\mathscr{N}$ называется *собственно разрывным*, если:

1) у каждой точки $q \in \mathscr{N}$ имеется окрестность $\mathscr{U}$, такая, что $A(\mathscr{U}) \cap \mathscr{U} = \varnothing$ для любого $A \in H$, который не является единицей группы;

2) точки $q, r \in \mathscr{N}$, которые при любом $A \in H$ удовлетворяют условию $Aq = r$, обладают соответственно окрестностями $\mathscr{U}$ и $\mathscr{U}'$, такими, что $B(\mathscr{U}) \cap \mathscr{U}' \neq \varnothing$ при любом $B \in H$.

Условие (1) означает, что фактор-пространство $\mathscr{N}/H$ является многообразием, а условие (2) — что это многообразие хаусдорфово. Таким образом, фактор-пространство $(I, \tilde{\eta})/G$ есть хаусдорфово пространство $(\mathscr{M}, g)$. Действие G собственно разрывно также в областях $I + II (\tilde{t} > -\tilde{x})$. Следовательно, $(I + II, \tilde{\eta})/G$ — тоже хаусдорфово пространство; на самом деле оно совпадает с $(\mathscr{M}', g')$. Аналогично $(I + III, \tilde{\eta})/G$ — хаусдорфово пространство $(\mathscr{M}'', g'')$. Отсюда видно, почему одно семейство изотропных геодезических может быть выбрано полным в расширении $(\mathscr{M}', g')$, а другое семейство — в расширении

$(\mathscr{M}'', \mathrm{g}'')$. Казалось бы, можно осуществить оба расширения одновременно. Однако действие группы G в области $(I+II+III)$, т. е. при $\tilde{t} > -|\tilde{x}|$, удовлетворяет лишь условию (1), а условие (2) для точек q на границе между I и II точек r на границе между I и III не выполняется. Поэтому фактор-пространство $(I+II+III, \tilde{\eta})/G$ не хаусдорфово, хотя и является многообразием.

Этот тип нехаусдорфова поведения отличается от примера, приведенного в разд. 2.1. Там можно было провести непрерывные кривые, которые разветвляются так, что две ее ветви уходят в разные области. Такое поведение мировой линии наблюдателя представляется крайне неудобным. В многообразии $(I+II+III)/G$ нет разветвляющихся кривых. Кривые из области I можно продолжить в II или в III, но не в обе эти области одновременно. Это могло бы служить отправной точкой для такого ослабления требования хаусдорфовости модели пространства-времени, при котором допускаются ситуации описанного здесь типа, но исключаются те, которые приводят к бифуркации кривых. Дальнейшие результаты по нехаусдорфовым моделям пространства-времени можно найти в работе [66].

Можно убедиться, что условию (1) удовлетворяет действие G на $\tilde{\mathscr{M}} - \{p\}$. Следовательно, пространство $(\tilde{\mathscr{M}} - \{p\}, \tilde{\eta})/G$ является в некотором смысле максимальным нехаусдорфовым расширением пространства $(\mathscr{M}, \mathbf{g})$. Однако и оно геодезически неполно, ибо есть геодезические, которые проходят через точку p, которую надо удалить. Если же p оставить, то действие группы не удовлетворяет условию (1) и поэтому фактор-пространство $\tilde{\mathscr{M}}/G$ не будет даже нехаусдорфовым многообразием. Рассмотрим, однако, расслоение линейных реперов $L(\tilde{\mathscr{M}})$, т. е. совокупность всех пар $(\mathbf{X}, \mathbf{Y})$ $\mathbf{X}, \mathbf{Y} \in T_q$, линейно-независимых векторов во всех точках $q \in \tilde{\mathscr{M}}$. Действие элемента A группы изометрий G на $\tilde{\mathscr{M}}$ индуцирует операцию A_* на $L(\tilde{\mathscr{M}})$, которая переводит репер $(\mathbf{X}, \mathbf{Y})$ в точке q в репер $(A_*\mathbf{X}, A_*\mathbf{Y})$ в точке $A(q)$. Эта операция удовлетворяет условию (1), поскольку даже для $(\mathbf{X}, \mathbf{Y}) \in \in T_p$, $A_*\mathbf{X} \neq \mathbf{X}$ и $A_*\mathbf{Y} \neq \mathbf{Y}$, если, конечно, A не есть тождественное преобразование; она удовлетворяет также условию (2), если даже $\mathbf{X}$ и $\mathbf{Y}$ лежат на изотропном конусе p. Таким образом, фактор-пространство $L(\tilde{\mathscr{M}})/G$ — хаусдорфово многообразие. Оно является расслоенным пространством над множеством $\tilde{\mathscr{M}}/G$, которое само не хаусдорфово и даже не многообразие. В определенном смысле $L(\tilde{\mathscr{M}})/G$ можно было бы рассматривать как расслоение линейных реперов для пространства $\tilde{\mathscr{M}}/G$. Тот факт, что расслоение линейных реперов может иметь нужные свойства, даже если само пространство ими не обладает, приводит к мыс-

ли о полезности изучения сингулярностей с точки зрения расслоения линейных реперов. Общий подход к этому указан в разд. 8.3.

Вернемся теперь к четырехмерному пространству Тауба $(\mathscr{M}, \mathbf{g})$, где $\mathscr{M}$ — это $R^1 \times S^3$, а метрика $\mathbf{g}$ дана формулой (5.32). Поскольку $\mathscr{M}$ односвязно, мы не можем, как в двумерном примере, построить накрывающее пространство. Однако можно получить похожий результат, рассматривая $\mathscr{M}$ как расслоенное пространство над S^2 со слоем $R^1 \times S^1$; проекция расслоения $\pi: \mathscr{M} \to S^2$ определяется как $(t, \psi, \theta, \phi) \to (\theta, \phi)$. Фактически это произведение оси t на расслоение Хопфа $S^3 \to S^2$ [162], у которого имеется слой S^1. Пространство $(\mathscr{M}, \mathbf{g})$ допускает 4-параметрическую группу изометрий, поверхностями транзитивности которой являются 3-сферы $\{t = \text{const}\}$. Эта группа изометрий отображает слои расслоения $\pi: \mathscr{M} \to S^3$ снова в слои, так что изометричны все пары $(\mathscr{F}, \tilde{\mathbf{g}})$, где $\mathscr{F}$ является слоем $(\mathscr{F} \approx \approx R^1 \times S^1)$ и $\tilde{\mathbf{g}}$ — метрика, индуцированная в слое четырехмерной метрикой $\mathbf{g}$ на $\mathscr{M}$. Слой $\mathscr{F}$ можно рассматривать как плоскость (t, ψ), а метрика $\tilde{\mathbf{g}}$ на $\mathscr{F}$ получается из (5.32) отбрасыванием членов с $\mathrm{d}\theta$ и $\mathrm{d}\phi$, т. е. $\tilde{\mathbf{g}}$ определяется квадратичной формой

$$\mathrm{d}s^2 = -U^{-1}\,\mathrm{d}t^2 + 4l^2U\,(\mathrm{d}\psi)^2. \tag{5.33}$$

Касательное пространство T_q в точке $q \in \mathscr{M}$ можно разложить на вертикальное подпространство V_q, которое касательно к слою $\mathscr{F}$ и натянуто на векторы $\partial/\partial t$ и $\partial/\partial\psi$ и на горизонтальное подпространство H_q, которое натянуто на векторы $\partial/\partial\theta$ и $\partial/\partial\phi - \cos\theta\,\partial/\partial\psi$. Любой вектор $\mathbf{X} \in T_q$ можно представить как сумму векторов $\mathbf{X}_V$, лежащего в V_q, и $\mathbf{X}_H$, лежащего в H_q. Тогда метрику $\mathbf{g}$ в T_q можно записать так:

$$g(\mathbf{X}, \mathbf{Y}) = g_V(\mathbf{X}_V, \mathbf{Y}_V) + (t^2 + l^2)\, g_H(\pi_*\mathbf{X}, \pi_*\mathbf{Y}_H), \tag{5.34}$$

где $g_V \equiv \tilde{\mathbf{g}}$, а $\mathbf{g}_H$ — обычная метрика на 2-сфере: $\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$. Итак, хотя метрика $\mathbf{g}$ и не является прямой суммой $\mathbf{g}_V$ и $(t^2 + l^2)\,\mathbf{g}_H$ (поскольку $R^1 \times S^3$ не есть прямое произведение S^2 и $R^1 \times S^1$), все же локально ее можно рассматривать в виде такой суммы.

Интересующая нас часть метрики $\mathbf{g}$ содержится в $\mathbf{g}_V$, и поэтому мы рассмотрим аналитические расширения пары $(\mathscr{F}, \mathbf{g}_V)$. Взяв их в такой же комбинации с метрикой $\mathbf{g}_H$ 2-сферы, как в (5.34), мы получим аналитические расширения пространства $(\mathscr{M}, \mathbf{g})$.

Метрика $\mathbf{g}_V$, определяемая формулой (5.33), имеет особенности при $t = t_\pm$, в которых $U = 0$. Однако, если взять многообразие $\mathscr{F}_0$, задаваемое ψ и $t_- < t < t_+$, $(\mathscr{F}_0, \mathbf{g}_V)$ можно расширить,

введя координату

$$\psi' = \psi + \frac{1}{2l}\int \frac{dt}{U(t)}.$$

Тогда получаем метрику $\mathbf{g}'_V$, для которой

$$ds^2 = 4l\,d\psi'\,(lU(t)\,d\psi' - dt).$$

Эта метрика аналитична на многообразии $\mathscr{F}'$ с топологией $S^1 \times R$, определяемой координатами ψ', t, причем $-\infty < t < \infty$. В области $t_- < t < t_+$ $(\mathscr{F}', \mathbf{g}_{V'})$ изометрична $(\mathscr{F}_0, \mathbf{g}_V)$. В этой области нет времениподобных кривых, но они есть при $t < t_-$ и $t > t_+$.

Это очень похоже на свойства рассмотренного ранее пространства $(\mathscr{M}', \mathbf{g}')$, за исключением того, что теперь возникают два горизонта (при $t = t_-$ и $t = t_+$) вместо одного горизонта (при $t = 0$). Одно семейство изотропных геодезических пересекает оба горизонта $t = t_-$ и $t = t_+$, а другое семейство описывает спирали вокруг этих поверхностей и неполно.

Как и прежде, можно произвести иное расширение введением координаты

$$\psi'' = \psi - \frac{1}{2l}\int \frac{dt}{U(t)}.$$

При этом получаем метрику $\mathbf{g}''_V$, для которой

$$ds^2 = 4l\,d\psi''\,(lU(t)\,d\psi'' + dt)$$

и которая аналитична на многообразии $\mathscr{F}''$, определяемом координатами ψ'' и t, $-\infty < t < \infty$. Эта метрика тоже изометрична $(\mathscr{F}_0, \mathbf{g}_V)$ в области $t_- < t < t_+$. Связь между разными расширениями снова можно продемонстрировать, перейдя к накрывающему пространству. Накрывающим пространством $\mathscr{F}_0$ является многообразие $\tilde{\mathscr{F}}_0$, определяемое координатами $-\infty < \psi < \infty$ и $t_- < t < t_+$. На $\tilde{\mathscr{F}}_0$ метрика $\mathbf{g}_V$ может быть записана в изотропной форме

$$ds^2 = 4l^2U(t)\,d\psi'\,d\psi'', \tag{5.35}$$

где $-\infty < \psi' < \infty$, $-\infty < \psi'' < \infty$. Ее можно расширить подобно тому, как это было сделано в случае решения Райсснера — Нордстрема. Выберем на $\mathscr{F}_0$ новые координаты (u_+, v_+) и (u_-, v_-):

$$u_\pm = \operatorname{arctg}\left(\exp\frac{\psi'}{\alpha_\pm}\right), \qquad v_\pm = \operatorname{arctg}\left(-\exp\left(-\frac{\psi''}{\alpha_\pm}\right)\right),$$

где

$$\alpha_+ = \frac{t_+ - t_-}{4l\,(mt + l^2)} \quad \text{и} \quad \alpha_- = \frac{t_+ - t_-}{4nl\,(mt + l^2)}.$$

Здесь $n > (mt_+ + l^2)/(mt_- + l^2)$ — некоторое целое число. Тогда метрика $\tilde{g}_V$, полученная из (5.35) этим преобразованием, аналитична на многообразии $\tilde{\mathcal{F}}$, изображенном на рис. 33, причем

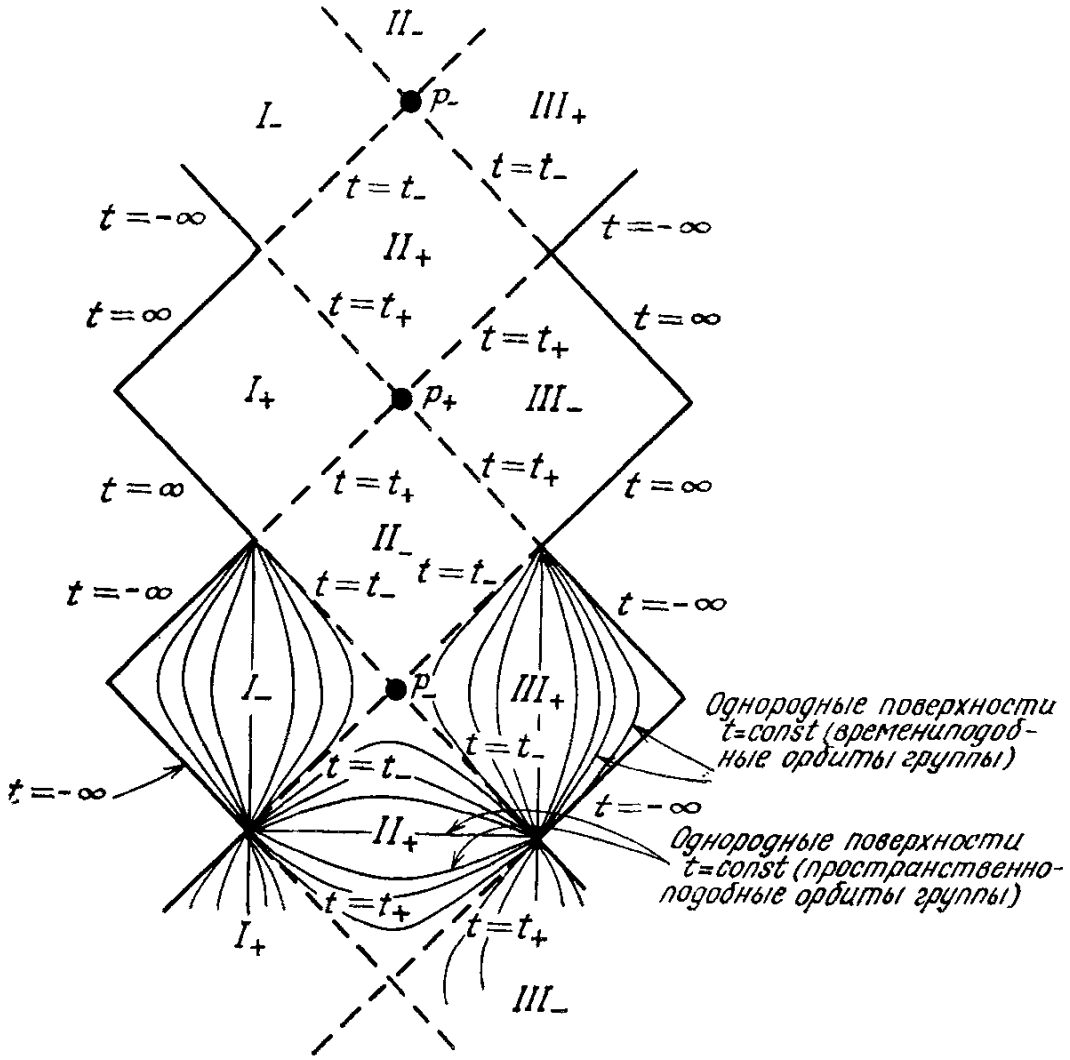


Рис. 33. Диаграмма Пенроуза максимально расширенного накрывающего пространства для двумерного сечения пространства Тауба — НУТ; изображены орбиты группы изометрии. Пространство Тауба — НУТ и его расширения получены из части этого пространства отождествлением точек при действии дискретной подгруппы группы изометрии.

координаты u_+, v_+ аналитичны везде, кроме поверхности $t = t_-$, а координаты u_-, v_- — везде, кроме $t = t_+$; на своих поверхностях неаналитичности (u_+, v_+) и (u_-, v_-) принадлежат по крайней мере классу C^3. Здесь имеется определенное сходство с расширением плоскости (t, r) решения Райсснера — Нордстрема.

Пространство $(\tilde{\mathcal{F}}, \tilde{g}_V)$ обладает однопараметрической группой изометрий, орбиты которой изображены на рис. 33. Вблизи точек p_+, p_- эта группа действует подобно группе Лоренца в двумерном пространстве Минковского (см. рис. 32, *б*). Пусть

G — дискретная подгруппа группы изометрии, порожденная нетривиальным элементом A последней. $(\mathscr{F}_0, \mathbf{g}_V)$ представляет собой фактор-пространство одной из областей $(II_+, \tilde{\mathbf{g}}_V)$ по G. $(\mathscr{F}', \mathbf{g}_V')$ и $(\mathscr{F}'', \mathbf{g}_V'')$ являются соответственно фактор-пространствами $(I_- + II_+ + III_-, \tilde{\mathbf{g}}_V)/G$ и $(I_+ + II_+ + III_+, \tilde{\mathbf{g}}_V)/G$. Построив фактор-пространство области $(I_+ + II_+ + I_-)$, мы могли бы получить хаусдорфово многообразие: это соответствует тому, чтобы расширять у поверхности $t = t_+$, как при построении $(\mathscr{F}', \mathbf{g}_V')$, а у поверхности $t = t_-$, как при построении $(\mathscr{F}'', \mathbf{g}_V'')$. Так же как в примере Мизнера, взяв фактор-пространство всего пространства $\tilde{\mathscr{F}}$ без точек p_+ и p_-, получаем нехаусдорфово многообразие, а взяв фактор-пространство $\tilde{\mathscr{F}}$, получим пространство, которое не хаусдорфово и не многообразие. Аналогично примеру Мизнера, фактор-пространство расслоения линейных реперов над $\mathscr{F}$ образует хаусдорфово многообразие.

Комбинируя эти расширения плоскости (t, ψ) с координатами (θ, ϕ), можно получить соответствующие расширения четырехмерного пространства $(\mathscr{M}, \mathbf{g})$. В частности, два расширения $(\mathscr{F}', \mathbf{g}_V')$ и $(\mathscr{F}'',$ приведут к двум различным локально нерасширяемым расширениям пространства $(\mathscr{M}, \mathbf{g})$, которые оба геодезически неполны.

Рассмотрим одно из этих расширений, скажем $(\mathscr{M}', \mathbf{g}')$. 3-сферы, являющиеся поверхностями транзитивности группы изометрии, пространственноподобны в области $t_- < t < t_+$ и времениподобны при $t > t_+$ и $t < t_-$. Поверхности транзитивности $t = t_-$ и $t = t_+$ изотропны и образуют горизонт Коши для любой пространственно-подобной поверхности, содержащейся в $t_- < t < t_+$, поскольку в областях $t < t_-$ и $t > t_+$ имеются времениподобные кривые, которые никогда не пересекают соответственно $t = t_-$ и $t = t_+$ (в этих областях существуют, в частности, замкнутые времениподобные кривые). Область пространства-времени $t_- \leqslant t \leqslant t_+$ компактна, и все же имеются времениподобные и изотропные геодезические, которые все время остаются внутри нее и неполны. Такого рода поведение будет рассмотрено позднее, в гл. 8.

Более подробное описание пространства Тауба — НУТ можно найти в работах [103, 109].

5.9. Прочие точные решения

В этой главе мы рассмотрели ряд точных решений и использовали их для того, чтобы дать примеры различных глобальных свойств, более общее исследование которых мы хотим провести позднее. Хотя локально известно много точных решений, лишь относительно малое число их изучено глобально. В заключение

этой главы отметим вкратце два других интересных семейства точных решений, глобальные свойства которых известны.

Первое из них составляют решения в виде *плоских волн* в пустом пространстве-времени. Они гомеоморфны R^4, и можно выбрать такие глобальные координаты (y, z, u, v), пробегающие значения от $-\infty$ до $+\infty$, в которых метрика имеет вид

$$ds^2 = 2\,du\,dv + dy^2 + dz^2 + H(z, y, u)\,du^2,$$

где

$$H = (y^2 - z^2) f(u) - 2yzg(u),$$

и $f(u)$, $g(u)$ — произвольные C^2-функции, характеризующие амплитуду и поляризацию волны. Эти пространства инвариантны относительно пятипараметрической группы изометрий, которая кратно-транзитивна на изотропных поверхностях $\{u = \text{const}\}$; особый подкласс, в котором $f(u) = \cos 2u$, $g(u) = \sin 2u$, допускает еще одно векторное поле Киллинга, т. е. такие однородные пространства инвариантны относительно 6-параметрической группы изометрий. Эти пространства не содержат замкнутых времениподобных или изотропных кривых, однако в них не может быть поверхностей Коши [124]. Локальные свойства этих пространств детально изучены Бонди, Пирани и Робинсоном [13], а глобальные — Пенроузом [124]. Оцват и Шюкинг [118] исследовали глобальные свойства такого пространства с более высокой симметрией. В работе [89] Хан и Пенруоз рассмотрели, как рассеиваются две ударные плоские волны и как это приводит к сингулярности.

Во-вторых, мы отметим 5-параметрическое семейство точных решений уравнений Эйнштейна — Максвелла без источников, найденное Картером [24] (см. также [37]). В это семейство в качестве частных случаев входят решения Шварцшильда, Райсснера — Нордстрема, Керра, решение Керра с зарядом, решение Тауба — НУТ и де Ситтера 1-го и 2-го родов. Описание некоторых их глобальных свойств дано в [22]. Несколько конкретных решений, тесно связанных с этим семейством, рассмотрены в [45, 90].

Глава 6

Причинная структура

Согласно постулату (а) разд. 3.2, из одной точки многообразия $\mathscr{M}$ к другой можно послать сигнал только в том случае, если эти точки могут быть соединены непространственноподобной геодезической. В этой главе мы займемся дальнейшим исследованием такого рода причинных отношений и получим ряд результатов, которые будут использованы в гл. 8 для доказательства существования сингулярностей.

В разд. 3.2 отмечалось, что изучение причинных отношений эквивалентно изучению конформной геометрии многообразия $\mathscr{M}$, т. е. множества всех метрик $\tilde{\mathbf{g}}$, конформных физической метрике $\mathbf{g}$ ($\tilde{\mathbf{g}} = \Omega^2\mathbf{g}$, где Ω — ненулевая C^r-функция). При такого рода конформном преобразовании метрики геодезическая кривая, вообще говоря, не останется геодезической, если только она не изотропная, но даже в последнем случае аффинный параметр вдоль кривой перестанет быть аффинным. Таким образом, в большинстве случаев геодезическая полнота (т. е. возможность продлить все геодезические до любых значений их аффинных параметров) будет зависеть от конкретного конформного множителя; следовательно (исключая специальные случаи, описанные в разд. 6.4), полнота не является свойством конформной геометрии. Кларке [33] и Зейферт [157] показали, что при выполнении некоторых физических разумных условий причинности любая лоренцева метрика конформна метрике, в которой все изотропные геодезические и все направленные в будущее времениподобные геодезические полны. Геодезическую полноту мы рассмотрим позднее, в гл. 8, где она служит основой для определения сингулярности.

В разд. 6.1 мы занимаемся вопросом ориентируемости времениподобных и пространственноподобных базисов. В разд. 6.2 даны определения основных причинных отношений и более широкое определение непространственноподобной кривой, допускающее не только кусочную дифференцируемость, но и просто непрерывность. Свойства границы будущего некоторого множества рассмотрены в разд. 6.3. В разд. 6.4 обсуждается ряд условий, которые исключают нарушение или близость к нарушению причинности. Тесно связанные между собой понятия области

Коши и глобальной гиперболичности введены в разд. 6.5 и 6.6, а в разд. 6.7 они использованы для доказательства существования непространственноподобных геодезических максимальной длины между определенными парами точек.

В разд. 6.8 излагается метод построения причинной границы пространства-времени, принадлежащий Героку, Кронхеймеру и Пенроузу. Конкретный пример такой границы дает класс асимптотически плоских пространств, который исследуется в разд. 6.9.

6.1. Ориентируемость

В ближайшей к нам области пространства-времени «стрела» времени четко задана направлением роста энтропии квазиизолированных термодинамических систем. Однако не вполне ясно, какова связь между этой «стрелой» времени и другими «стрелами» времени, которые определяются расширением Вселенной и условием излучения в электродинамике; интересующийся этим читатель может найти обсуждение этого вопроса в работах [63, 79, 81, 83]. Представляется физически разумным предположить, что существует локальная термодинамическая «стрела» времени, заданная непрерывным образом в каждой точке; но мы ограничимся требованием, чтобы имелась возможность задать непрерывно разделение непространственноподобных векторов на два класса, один из которых, по нашему выбору, назовем классом векторов, направленных в будущее, а другой — классом векторов, направленных в прошлое. Если такое деление возможно, то будем говорить, что пространство-время *ориентируемо по времени.*

В некоторых типах пространства-времени задать такую ориентацию по времени невозможно. Примером служит пространство-время, построенное из пространства де Ситтера (разд. 5.2) отождествлением точек, получаемых отражением относительно начала пятимерного объемлющего пространства. В этом пространстве существуют замкнутые кривые, не гомотопные нулю, при обходе по которым направление времени меняется на обратное. Однако ясно, что эта трудность могла бы быть устранена просто снятием отождествления точек, и это фактически всегда возможно: если пространство-время $(\mathscr{M}, \mathbf{g})$ не ориентируемо по времени, то оно имеет двукратное накрывающее пространство $(\tilde{\mathscr{M}}, \mathbf{g})$, которое ориентируемо. $\tilde{\mathscr{M}}$ можно определить как множество пар (p, α), где $p \in \mathscr{M}$, а α — одна из ориентаций времени в p. Тогда вместе с естественной структурой и проекцией $\pi: (p, \alpha) \to p\mathscr{M}$ является двукратным накрывающим многообразием для $\mathscr{M}$. Если $\tilde{\mathscr{M}}$ двусвязно, то $(\mathscr{M}, \mathbf{g})$ ориентируемо по времени. Если $\tilde{\mathscr{M}}$ односвязно, то $(\mathscr{M}, \mathbf{g})$ неориентируемо по времени.

но $(\tilde{\mathscr{M}}, \mathbf{g})$ — ориентируемо. В последующих разделах мы будем предполагать, что имеет место одно из двух: либо $(\mathscr{M}, \mathbf{g})$ ориентируемо по времени, либо мы рассматриваем ориентируемое по времени накрывающее пространство. Если мы сможем доказать наличие сингулярностей в накрывающем пространстве, то они должны быть и в $(\mathscr{M}, \mathbf{g})$.

Можно задаться таким вопросом: является ли пространство-время *пространственно-ориентируемым*, т. е. можно ли разделить непрерывным образом базисы из трех пространственноподобных осей на левые и правые? Герок [51] указал на интересную связь между пространственной и временной ориентируемостью, следующую из того факта, что некоторые эксперименты в физике элементарных частиц не инвариантны относительно зарядового сопряжения или пространственного отражения (или того и другого вместе). Вместе с тем имеются теоретические основания для уверенности в том, что все взаимодействия инвариантны относительно комбинации зарядового сопряжения, отражения и обращения времени (*CPT*-теорема, см. [164]). Если считать, что нарушение инвариантности относительно зарядового сопряжения и пространственного отражения в слабых взаимодействиях — не локальное явление, а происходит во всех точках пространства-времени, то из этого следует, что при обходе по любой замкнутой кривой такие характеристики, как знак заряда, ориентация базиса пространственноподобных осей и ориентация времени могут измениться или остаться неизменными лишь все три вместе. (Обычная теория Максвелла, в которой электромагнитное поле имеет определенный знак в каждой точке, не допускает изменения знака заряда при обходе по замкнутой кривой, не гомотопной нулю, без изменения ориентации времени. Однако можно было бы построить теорию, в которой поле двузначно и меняет свой знак при обходе по такой кривой. Такая теория была бы в согласии со всеми имеющимися экспериментальными данными.) В частности, если предположить, что пространство-время ориентируемо по времени, то оно должно быть и пространственно-ориентируемым. (Фактически этот вывод следует из одних лишь экспериментальных данных без использования *CPT*-теоремы.)

Герок [55] показал, что из возможности задать двухкомпонентные спинорные поля в каждой точке с необходимостью следует параллелизуемость пространства-времени, т. е. возможность ввести непрерывную систему базисов касательного пространства в каждой точке. (Дальнейшие следствия существования спинорных структур получены в [56].)

6.2. Причинные кривые

Принимая, что пространство-время ориентируемо по времени в смысле, разъясненном в предыдущем разделе, мы в каждой

точке можем разделить непространственноподобные векторы на направленные в будущее и в прошлое. Тогда для множеств $\mathcal{S}$ и $\mathcal{U}$ можно определить *хронологическое будущее* $I^+(\mathcal{S}, \mathcal{U})$ *множества* $\mathcal{S}$ *относительно* $\mathcal{U}$ как множество всех точек в $\mathcal{U}$, любую из которых можно достичь из $\mathcal{S}$ по направленной в будущее времениподобной кривой в $\mathcal{U}$. (Под кривой мы всегда имеем в виду кривую ненулевой протяженности; одна точка не считается кривой. Таким образом, $I^+(\mathcal{S}, \mathcal{U})$ не может содержать $\mathcal{S}$.) $I^+(\mathcal{S}, \mathcal{M})$ будем обозначать $I^+(\mathcal{S})$; $I^+(\mathcal{S})$ является открытым множеством: если $p \in \mathcal{M}$ можно достичь из $\mathcal{S}$ по направленной в будущее времениподобной кривой, то имеется малая окрестность p, которую тоже можно достичь по такой кривой.

Для этого определения существует дуальное, в котором «будущее» заменяется на «прошлое» и $+$ на $-$; чтобы не повторяться, мы будем считать дуальные определения и результаты самоочевидными.

Обозначим через $J^+(\mathcal{S}, \mathcal{U})$ *причинное будущее множества* $\mathcal{S}$ *относительно* $\mathcal{U}$; оно определяется как объединение $\mathcal{S} \cap \mathcal{U}$ с множеством всех тех точек в $\mathcal{U}$, которых можно достичь из $\mathcal{S}$ по направленным в будущее непространственноподобным кривым в $\mathcal{U}$. В разд. 4.5 мы видели, что непространственноподобная кривая между двумя точками, которая не является изотропной геодезической кривой, может быть деформирована во времениподобную кривую между этими точками. Поэтому, если $\mathcal{U}$ — открытое множество и $p, q, r \in \mathcal{U}$, то как из условия

$$q \in I^+(p, \mathcal{U}), \qquad r \in I^+(q, \mathcal{U}),$$

так и из условия

$$q \in I^+(p, \mathcal{U}), \qquad r \in J^+(q, \mathcal{U})$$

следует, что $r \in I^+(p, \mathcal{U})$. Отсюда $\overline{I^+}(p, \mathcal{U}) = \overline{J^+}(p, \mathcal{U})$ и $\dot{I}^+(p, \mathcal{U}) = \dot{J}^+(p, \mathcal{U})$, где для любого множества $\mathcal{K}$, $\bar{\mathcal{K}}$ означает замыкание,

$$\dot{\mathcal{K}} \equiv \bar{\mathcal{K}} \cap (\overline{\mathcal{M} - \mathcal{K}})$$

означает границу $\mathcal{K}$.

Будем снова вместо $I^+(\mathcal{S}, \mathcal{M})$ писать просто $I^+(\mathcal{S})$. $I^+(\mathcal{S})$ представляет собой область пространства-времени, в которой события из $\mathcal{S}$ могут причинно влиять: как видно из рис. 34, $I^+(\mathcal{S})$ не обязательно замкнутое множество, даже если $\mathcal{S}$ представляет собой одну точку. Кстати, этот пример иллюстрирует полезный способ построения пространства-времени с заданными свойствами причинности: начинаем с некоторого пространства-времени (при отсутствии других указаний это будет пространство Минковского), вырезаем из него какое-либо замкнутое множество и, если нужно, склеиваем его подходящим образом (т. е.

отождествляем точки $\mathscr{M}$). В результате получаем снова многообразие с лоренцевой метрикой, т. е. пространство-время, хотя оно и выглядит несколько неполным, поскольку некоторые точки вырезаны. Однако выше было отмечено, что эта неполнота может быть «исправлена» подходящим конформным преобразованием, которое передвинет вырезанные точки в бесконечность.

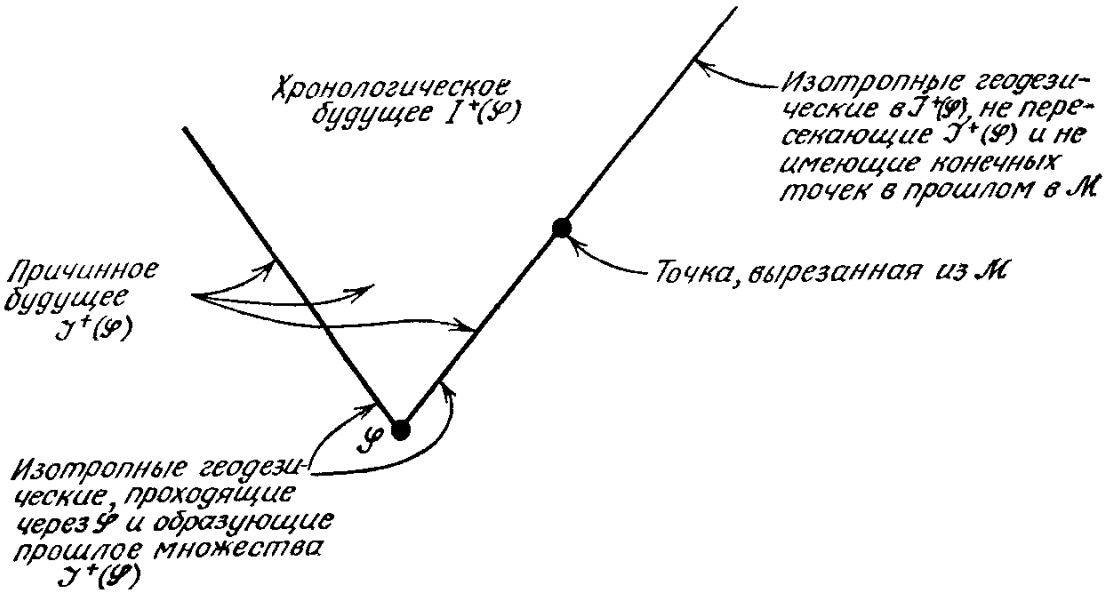


Рис. 34. При удалении из пространства Минковского одной точки причинное будущее $J^+(\mathscr{S})$ замкнутого множества $\mathscr{S}$ не обязательно замкнуто. Находящиеся за этой точкой части границы будущего $J^+(\mathscr{S})$ могут быть образованы изотропными геодезическими сегментами, не имеющими в $\mathscr{M}$ конечных точек в прошлом.

Контур будущего (*horismos*) *для* множества $\mathscr{S}$ *относительно* $\mathscr{U}$, обозначаемый через $E^+(\mathscr{S}, \mathscr{U})$, определяется как $J^+(\mathscr{S}, \mathscr{U}) - I^+(\mathscr{S}, \mathscr{U})$; будем писать $E^+(\mathscr{S})$ вместо $E^+(\mathscr{S}, \mathscr{M})$. (В некоторых работах отношения $p \in I^+(q)$, $p \in J^+(q)$) и $p \in E^+(q)$ обозначаются соответственно как $q \ll p$, $q < p$ и $q \to p$.) Если $\mathscr{U}$ — открытое множество, согласно предложению 4.5.10, точки $E^+(\mathscr{S}, \mathscr{U})$ должны лежать на направленных в будущее изотропных геодезических из $\mathscr{S}$; если же $\mathscr{U}$ — нормальная выпуклая окрестность точки p, то из предложения 4.5.1 следует, что $E^+(p, \mathscr{U})$ состоит из направленных в будущее изотропных геодезических в $\mathscr{U}$ с началом в p и образует в $\mathscr{U}$ границу как хронологического $I^+(p, \mathscr{U})$, так и причинного $J^+(p, \mathscr{U})$ будущего точки p. В частности, это имеет место в пространстве Минковского, где такую границу образует изотропный конус точки p. Но в более сложных случаях такое не обязательно.

Для последующего изложения удобно расширить определение времениподобных и непространственноподобных кривых, ослабив требование кусочной дифференцируемости до требова-

ния лишь непрерывности. Хотя в таком случае кривая может и не иметь касательного вектора, мы все же сможем говорить, что она непространственноподобна, если локально любые две точки кривой можно соединить кусочно дифференцируемой непространственноподобной кривой. Точнее, будем говорить, что непрерывная кривая $\gamma:\mathcal{F}\to\mathcal{M}$, где $\mathcal{F}$ — связный интервал из R^1, *направлена в будущее и непространственноподобна*, если для каждой $t\in\mathcal{F}$ существует ее окрестность G в $\mathcal{F}$ и нормальная выпуклая окрестность $\mathcal{U}$ точки $\gamma(t)$ в $\mathcal{M}$, такие, что для любого $t_1\in G_1$

$$\gamma(t_1)\in J^-(\gamma(t),\ \mathcal{U})-\gamma(t),$$

если $t_1<t$, и

$$\gamma(t_1)\in J^+(\gamma(t),\ \mathcal{U})-\gamma(t),$$

если $t<t_1$. Будем говорить, что γ *направлена в будущее и времениподобна*, если те же самые условия выполняются при замене J на I. Впредь, если не оговаривается иное, под времениподобной или непространственноподобной кривой мы будем подразумевать именно такую непрерывную кривую, а две кривые, которые получаются одна из другой перепараметризацией, будем считать эквивалентными. После такого обобщения мы можем доказать утверждение, которое неоднократно используется в оставшейся части главы. Сначала дадим несколько новых определений.

Точка p будет называться *конечной точкой в будущем* непространственноподобной кривой $\gamma:\mathcal{F}\to\mathcal{M}$, направленной в будущее, если для каждой окрестности $\mathcal{V}$ этой точки существует такое значение $t\in\mathcal{F}$, что $\gamma(t_1)\in\mathcal{V}$ для каждого $t_1\in\mathcal{F}$ при $t_1\geqslant t$. Непространственноподобная кривая *непродолжима в будущее* (соответственно *непродолжима в будущее в множестве* $\mathcal{S}$), если у нее нет конечной точки в будущем (соответственно нет конечной точки в будущем в множестве $\mathcal{S}$). Точка p будет называться *предельной точкой* бесконечной последовательности непространственноподобных кривых λ_n, если любую окрестность p пересекает бесконечное число этих кривых. Непространственноподобная кривая будет называться *предельной кривой* последовательности λ_n, если имеется подпоследовательность λ_n' последовательности λ_n, такая, что для каждой $p\in\lambda$, λ_n' сходится к p.

Лемма 6.2.1

Пусть $\mathcal{S}$ — некоторое открытое множество и пусть λ_n — бесконечная последовательность непространственноподобных кривых в $\mathcal{S}$, которые непродолжимы в $\mathcal{S}$. Если $p\in\mathcal{S}$ — предельная точка последовательности λ_n, то через p можно провести непространственноподобную кривую λ, которая непродолжима в будущее в $\mathcal{S}$ и является предельной кривой последовательности λ_n.

Достаточно рассмотреть случай $\mathscr{S} = \mathscr{M}$, поскольку $\mathscr{S}$ можно считать многообразием с лоренцевой метрикой. Допустим, что $\mathscr{U}_1$ — нормальная выпуклая координатная окрестность точки p

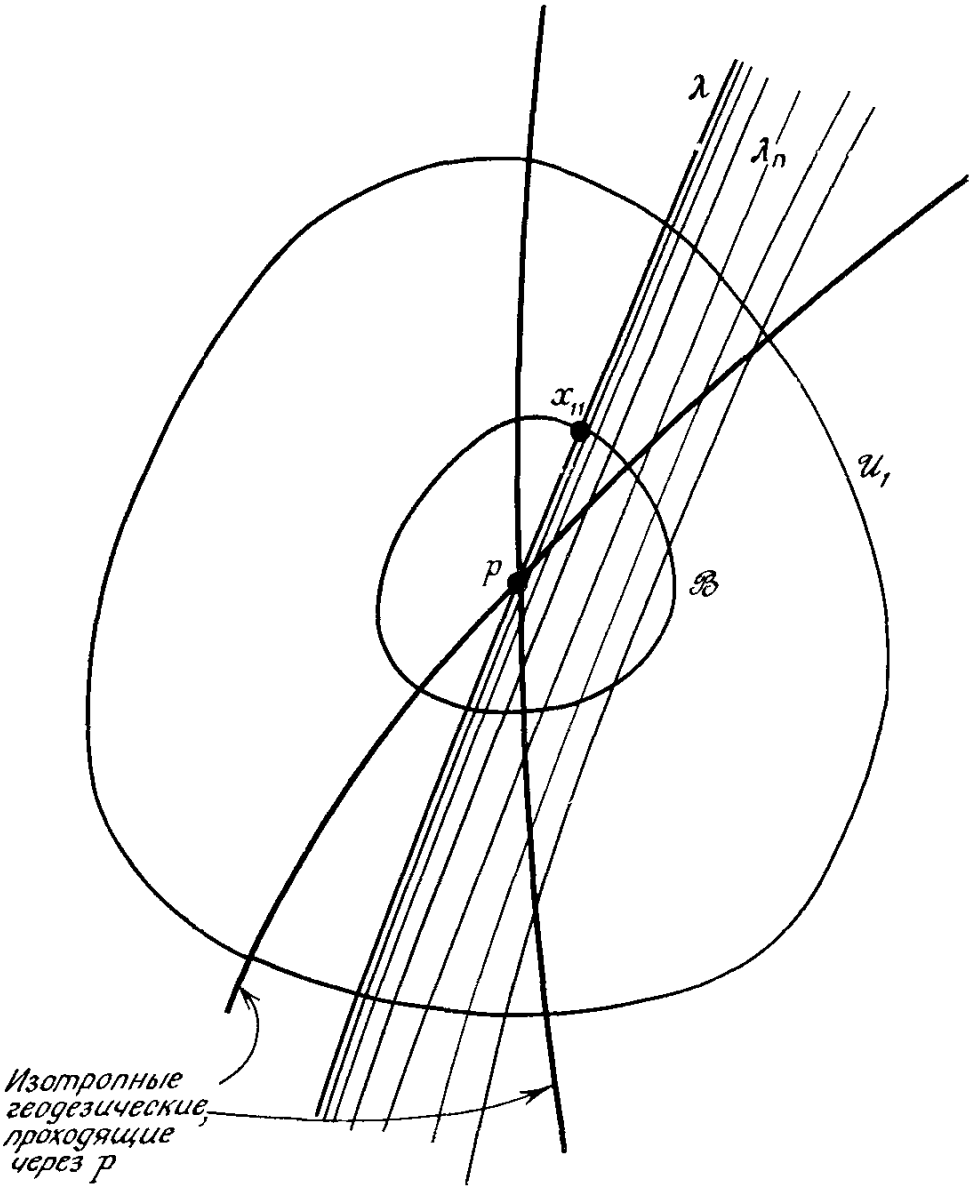


Рис. 35. Проходящая через точку p непространственноподобная предельная кривая λ для семейства непространственноподобных кривых λ_n; для этого семейства p является предельной точкой.

и что $\mathscr{B}(q, a)$ — открытый шар координатного радиуса a с центром в q. Пусть постоянная $b > 0$ такова, что $\mathscr{B}(p, b)$ существует, и пусть $\lambda(1, 0)_n$ — последовательность $\lambda_n \cap \mathscr{U}_1$, сходящаяся к p. Так как $\mathscr{B}(p, b)$ — компактное множество, оно содержит предельные точки последовательности $\lambda(1, 0)_n$. Любая такая предельная точка должна лежать либо в $J^-(p, \mathscr{U}_1)$, либо в $J^+(p, \mathscr{U}_1)$, поскольку в противном случае существовали бы

окрестности $\mathscr{V}_1$ точки y и $\mathscr{V}_2$ точки p, которые нельзя было бы соединить непространственноподобной кривой в $\mathscr{U}_1$. Пусть

$$x_{11} \in J^+(p, \mathscr{U}_1) \cap \dot{\mathscr{B}}(p, b)$$

— одна из этих предельных точек (рис. 35) и $\lambda(1,1)_n$ — сходящаяся к x_{11} подпоследовательность из $\lambda(1,0)_n$. Точка x_{11} будет точкой нашей предельной кривой λ. Продолжая по индукции, определим точку

$$x_{ij} \in J^+(p, \mathscr{U}_1) \cap \dot{\mathscr{B}}(p, i^{-1}jb)$$

как предельную точку последовательности $\lambda(i-1, i-1)_n$ при $j = 0$ и подпоследовательности $\lambda(i, j-1)$ при $i \geqslant j \geqslant r$; пусть также $\lambda(i,j)_n$ — подпоследовательность подпоследовательности $\lambda(1,1)$, сходящаяся к x_{ij}. Другими словами, мы делим интервал $[0, b]$ на все меньшие и меньшие отрезки и получаем точки нашей предельной кривой на соответствующих сферах с центром в p. Поскольку любые две точки x_{ij} $(j \geqslant i)$ разделены непространственноподобно, замыкание объединения всех x_{ij} $(j \geqslant i)$ даст непространственноподобную кривую λ от $p = x_{i0}$ до $x_{11} = x_{ii}$. Теперь остается построить такую подпоследовательность λ'_n последовательности λ_n, чтобы для каждого $q \in \lambda$ λ'_n сходилась к q. Это мы сделаем, выбирая в качестве λ'_m член подпоследовательности $\lambda(m,n)$, который пересекает каждый из шаров $\mathscr{B}(x_{mj}, m^{-1}b)$ при $0 \leqslant j \leqslant m$. Таким образом, λ будет предельной кривой последовательности λ_n от p до x_{11}. Пусть теперь $\mathscr{U}_2$ — нормальная выпуклая окрестность вокруг x_{11}; повторим построение, используя теперь последовательность λ'_n. Поступая таким образом, можно продолжить λ неограниченно. □

6.3. Ахрональные границы

Из предложения 4.5.1 следует, что в нормальной выпуклой окрестности $\mathscr{U}$ граница множеств $I^+(p, \mathscr{U})$ или $I^+(p, \mathscr{U})$ образована направленными в будущее изотропными геодезическими с началом p. Для установления свойств границ более общего типа введем понятия ахронального множества и множества будущего.

Множество $\mathscr{S}$ называется *ахрональным* (в литературе его иногда называют «полупространственноподобным»), если $I^+(\mathscr{S}) \cap \mathscr{S}$ пусто, иначе говоря, если никакие две точки $\mathscr{S}$ не разделены времениподобно. $\mathscr{S}$ называется *множеством будущего*, если $\mathscr{S} \supset I^+(\mathscr{S})$. Отметим, что если $\mathscr{S}$ — множество будущего, то $\mathscr{M} - \mathscr{S}$ — множество прошлого. Примерами множеств будущего служат $I^+(\mathscr{N})$ и $J^+(\mathscr{N})$, где $\mathscr{N}$ — любое множество. Примеры ахрональных множеств дает следующий фундаментальный результат.

Предложение 6.3.1

Если $\mathscr{S}$ — множество будущего, то $\dot{\mathscr{S}}$ (граница $\mathscr{S}$) представляет собой замкнутое вложенное ахрональное трехмерное C^1-многообразие.

Если $q \in \dot{\mathscr{S}}$, то любая окрестность q пересекает $\mathscr{S}$ и $\mathscr{M} - \mathscr{S}$. Если $p \in I^+(q)$, то у точки q существует окрестность в $I^-(p)$. Следовательно, $I^+(q) \subset \mathscr{S}$. Аналогично $I^-(q) \subset \mathscr{M} - \mathscr{S}$. Если $r \in I^+(q)$, то существует такая окрестность точки r, что

$$\mathscr{V} \subset I^+(q) \subset \mathscr{S}.$$

Таким образом, r не может принадлежать $\dot{\mathscr{S}}$. В некоторой окрестности $\mathscr{U}_\alpha$ точки q можно ввести нормальные координаты (x^1, x^2, x^3, x^4) с времениподобным вектором $\partial/\partial x^4$ так, чтобы кривые $\{x^i = \text{const}, i = 1, 2, 3\}$ пересекали и $I^+(q, \mathscr{U}_\alpha)$, и $I^-(q, \mathscr{U}_\alpha)$. Тогда каждая из этих кривых должна содержать точно одну точку $\dot{\mathscr{S}}$. Координата x^4 этих точек должна быть функцией Липшица от x^i $(i = 1, 2, 3)$, поскольку никакие две точки $\dot{\mathscr{S}}$ не разделены времениподобно. Поэтому взаимнооднозначное отображение $\phi: \mathscr{S} \cap \mathscr{U}_\alpha \to R^3$, задаваемое равенством $\phi_\alpha(p) = x^i(p)$ $(i = 1, 2, 3)$ для $p \in \mathscr{S} \cap \mathscr{U}_\alpha$, есть гомеоморфизм. Таким образом, $(\dot{\mathscr{S}} \cap \mathscr{U}_\alpha, \phi_\alpha)$ есть C^1-атлас для $\dot{\mathscr{S}}$. □

Множество со свойствами $\dot{\mathscr{S}}$, перечисленными в предложении 6.3.1, будем называть *ахрональной границей*. Такое множество можно разбить на четыре непересекающихся подмножества $\dot{\mathscr{S}}_N$, $\dot{\mathscr{S}}_+$, $\dot{\mathscr{S}}_-$, $\dot{\mathscr{S}}_0$ следующим образом: для точки $q \in \dot{\mathscr{S}}$ найдутся или не найдутся точки p, $r \in \dot{\mathscr{S}}$, такие, что $p \in E^-(q) - q$, $r \in E^+(q) - q$ и различные возможности определяют эти подмножества множества $\dot{\mathscr{S}}$ по следующей схеме:

$q \in$	$\exists p$	$\nexists p$	
	$\dot{\mathscr{S}}_N$	$\dot{\mathscr{S}}_-$	$\exists r$
	$\dot{\mathscr{S}}_+$	$\dot{\mathscr{S}}_0$	$\nexists r$

Если $q \in \dot{\mathscr{S}}_N$, то $r \in E^+(p)$, поскольку $r \in I^+(p)$ и согласно предложению 6.3.1 $r \in I^+(p)$. Это означает, что в $\mathscr{S}$ имеется изотропный геодезический сегмент, проходящий через q. Если $q \in \dot{\mathscr{S}}_+$ (соответственно $\dot{\mathscr{S}}_-$), то q является конечной точкой в будущем (соответственно прошлом) изотропной геодезической в $\dot{\mathscr{S}}$. Подмножество $\dot{\mathscr{S}}_0$ пространственноподобно (точнее непричинно). Это разбиение иллюстрируется рис. 36.

Полезное условие, определяющее принадлежность точки одному из множеств $\dot{\mathscr{S}}_N$, $\dot{\mathscr{S}}_+$ и $\dot{\mathscr{S}}_-$ дает следующая лемма, принадлежащая Пенроузу [128]:

Лемма 6.3.2

Пусть $\mathscr{W}$ — некоторая окрестность $q \in \dot{\mathscr{S}}$, где $\mathscr{S}$ — множество будущего. Тогда

1) из $I^+(q) \subset I^+(\mathscr{S} - \mathscr{W})$ следует, что $q \in \dot{\mathscr{S}}_N \cup \dot{\mathscr{S}}_+$,

2) из $I^-(q) \subset I^-(\mathscr{M} - \mathscr{S} - \mathscr{W})$ следует, что $q \in \dot{\mathscr{S}}_N \cup \dot{\mathscr{S}}_-$.

Достаточно доказать пункт (1), поскольку $\dot{\mathscr{S}}$ можно считать также и границей множества прошлого $\mathscr{M} - \mathscr{S}$. Пусть $\{x_n\}$ — бесконечная последовательность точек в $I^+(q) \cap \mathscr{W}$, сходящаяся

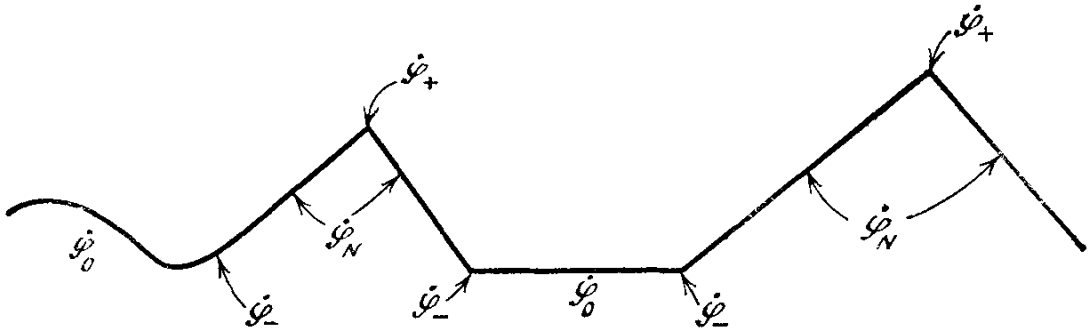


Рис. 36. Ахрональная граница $\dot{\mathscr{S}}$ может быть разделена на четыре множества: пространственноподобное $\dot{\mathscr{S}}_0$, изотропное $\dot{\mathscr{S}}_N$ и $\dot{\mathscr{S}}_+$ (соответственно $\dot{\mathscr{S}}_-$)— конечная точка будущего (соответственно прошлого) изотропной геодезической в $\dot{\mathscr{S}}$.

к q. Если $I^+(q) \subset I^+(\mathscr{S} - \mathscr{W})$, от каждого x_n до $\mathscr{S} - \mathscr{W}$ можно провести направленную в прошлое времениподобную кривую. По лемме 6.2.1 существует направленная в прошлое предельная кривая λ от q до $\overline{(\mathscr{S} - \mathscr{W})}$. Поскольку $I^-(q)$ — открытое множество и содержится в $\mathscr{M} - \mathscr{S}$, $I^-(q) \cap \mathscr{S}$ пусто. Отсюда λ должна быть времениподобной геодезической и лежать в $\dot{\mathscr{S}}$. □

В качестве иллюстрации предыдущих результатов рассмотрим $\dot{J}^+(\mathscr{K}) = \dot{I}^+(\mathscr{K})$, границу замкнутого множества будущего $\mathscr{K}$. Согласно предложению 6.3.1, она является ахрональным многообразием, и по лемме 6.3.2 каждая точка множества $\dot{J}(\mathscr{K}) - \mathscr{K}$ принадлежит $[\dot{J}^+(\mathscr{K})]_N$ или $[\dot{J}^+(\mathscr{K})]_+$. Это означает, что $\dot{J}(\mathscr{K}) - \mathscr{K}$ порождается изотропными геодезическими сегментами, которые могут иметь конечную точку в будущем в $\dot{J}^+(\mathscr{K}) - \mathscr{K}$, но если они имеют конечные точки в прошлом, то обязательно только в самом $\mathscr{K}$. Как видно из рис. 36, может существовать изотропная геодезическая, порождающая сегменты,

которые вообще не имеют конечных точек в прошлом, а уходят в бесконечность. Очевидно, что это несколько надуманный пример, но Пенроуз [124] показал, что аналогичное поведение имеет место в таком простом случае, как плоские волны; другими примерами служат решения де Ситтера 2-го рода (разд. 5.2) и Райснера — Нордстрема (разд. 5.5). В разд. 6.6 мы увидим, что такое поведение связано с отсутствием поверхностей Коши для этих решений.

Будем говорить, что открытое множество $\mathcal{U}$ — *причинно-простое*, если для каждого компактного множества

$$\dot{J}^+(\mathcal{K}) \cap \mathcal{U} = E^+(\mathcal{K}) \cap \mathcal{U} \quad \text{и} \quad \dot{J}^-(\mathcal{K}) \cap \mathcal{U} = E^-(\mathcal{K} \cap \mathcal{U}).$$

Это эквивалентно утверждению, что $J^+(\mathcal{K})$ и $J^-(\mathcal{K})$ замкнуты в $\mathcal{U}$.

6.4. Условия причинности

Постулат (а) разд. 3.2 требует только локальной причинности; глобальные вопросы остаются при этом открытыми. Следовательно, мы не исключили такую возможность, что в больших масштабах могут найтись замкнутые времениподобные геодезические (т. е. времениподобные окружности S^1). Но существование таких кривых, по-видимому, привело бы к возможности логических парадоксов: можно представить ситуацию, когда космонавт на звездолете может совершить круг по такой кривой и, возвратившись обратно к своему старту, помешать самому себе стартовать в первый раз. Конечно, здесь есть противоречие только с позиций элементарных представлений о свободе воли, но их не так легко отбросить, поскольку эти представления опираются на уверенность, что мы вольны осуществить любой возможный эксперимент. Не исключено, что можно построить теорию, в которой допускались бы замкнутые времениподобные кривые, а представление о свободе воли было бы как-то изменено (см., например, [150]), но проще представить себе, что пространство-время удовлетворяет требованию, которое мы называем *хронологическим условием*, а именно, чтобы в пространстве-времени не было замкнутых времениподобных кривых. Однако мы должны иметь в виду возможность существования точек (может быть, там, где плотность или кривизна очень велики), в которых это условие не выполняется. Множество всех таких точек будем называть множеством, *нарушающим хронологию*; оно характеризуется следующим предложением:

Предложение 6.4.1 (Картер)
Множество многообразия $\mathcal{M}$, нарушающее хронологию, есть объединение пересекающихся множеств вида $I^+(q) \cap I^-(q)$, $q \in \mathcal{M}$.

Если q принадлежит множеству многообразия $\mathscr{M}$, нарушающему хронологию, то должна существовать направленная в будущее времениподобная кривая λ с конечными точками в прошлом и будущем в q. Если $r \in I^-(q) \cap I^+(q)$, то найдутся направленные в прошлое и будущее времениподобные кривые μ_1 и μ_2, соединяющие q и r. Тогда $\mu_1^{-1} \circ \lambda \circ \mu_2$ будет направленной в будущее времениподобной кривой с конечными точками в прошлом и будущем r. Более того, если

$$r \in [I^-(q) \cap I^+(q)] \cap [I^-(p) \cap I^+(p)],$$

то

$$p \in I^-(q) \cap I^+(q) = I^-(p) \cap I^+(p).$$

Для завершения доказательства заметим, что каждая точка r, в которой нарушается хронология, принадлежит множеству $I^+(r) \cap I^+(r)$. □

Предложение 6.4.2

Если $\mathscr{M}$ компактно, то множество, нарушающее хронологию в $\mathscr{M}$, является непустым.

$\mathscr{M}$ может быть покрыто открытыми множествами вида $I^+(q)$, $q \in \mathscr{M}$. Если хронологическое условие выполняется в q, то $q \notin I^+(q)$. Таким образом, если бы хронологическое условие выполнялось в каждой точке, нельзя было бы покрыть $\mathscr{M}$ конечным числом множеств вида $I^+(q)$. □

Из полученного результата, по-видимому, разумно сделать вывод, что пространство-время некомпактно. Другой аргумент против компактности состоит в том, что никакое компактное четырехмерное многообразие с лоренцевой метрикой на нем не может быть односвязным. (Существование лоренцевой метрики означает, что характеристика Эйлера $\chi(\mathscr{M}) = 0$ ([162], p. 270).) Как известно, $\chi = \sum_{n=0}^{4} (-1)^n B_n$, где $B_n \geqslant 0$ — n-е число Бетти многообразия $\mathscr{M}$. В силу дуальности ([160], p. 297) $B_n = B_{4-n}$. Так как $B_0 = B_4 = 1$, это означает, что $B_1 \neq 0$, откуда в свою очередь следует, что $\pi_1(\mathscr{M}) \neq 0$ ([160], p. 398). Таким образом, компактное пространство-время является в действительности компактным многообразием, в котором отождествлены некоторые точки. Представляется физически разумным не отождествлять точки, а считать, что пространством-временем является накрывающее пространство.

Мы будем говорить, что выполнено *условие причинности*, если не существует замкнутых непространственноподобных кривых. Аналогично предложению 6.4.1 имеем

Предложение 6.4.3

Множество точек, в которых нарушается условие причинности, есть объединение непересекающихся множеств вида $J^-(q) \cap J^+(q)$, $q \in \mathscr{M}$. □

В частности, если в точке $q \in \mathscr{M}$ условие причинности нарушено, а хронологическое условие выполняется, должна существовать замкнутая изотропная геодезическая кривая γ, проходящая через q. Пусть v — аффинный параметр на γ (рассматриваемый как отображение открытого интервала из R^{-1} на $\mathscr{M}$) и пусть $\ldots, v_{-1}, v_0, v_1, v_2, \ldots$ — последовательные значения v и q. Тогда мы можем сравнивать в точке q касательный вектор $\partial/\partial v|_{v=v_0}$ и касательный вектор $\partial/\partial v|_{v=v_1}$, полученный параллельным переносом при обходе по γ. Поскольку оба вектора имеют одинаковое направление, то они должны быть пропорциональны $\partial/\partial v|_{v=v_1} = a\partial/\partial v|_{v=v_0}$. Множитель a имеет следующий смысл: аффинное расстояние $v_{n+1} - v_n$, покрываемое при n-м обходе по γ, равно $a^{-n}(v_1 - v_0)$. Отсюда при $a > 1$ v никогда не достигает значения $(v_1 - v_0)(1 - a^{-1})^{-1}$, и, таким образом, кривая γ геодезически неполна в будущем, несмотря на то, что можно совершить бесконечное число оборотов. Аналогично, если $a < 1$, то γ неполна в прошлом, в то время как при $a = 1$ она полна в обоих направлениях. В двумерной модели пространства Тауба — НУТ, описанной в разд. 5.7, существуют замкнутые изотропные геодезические, которые служат примером случая с $a > 1$. Поскольку множитель a является конформным инвариантом, эта неполнота не зависит от конформного множителя. Однако такого рода поведение может встретиться лишь при нарушении причинности в том или ином смысле; если выполняется условие сильной причинности (см. ниже), подходящее конформное преобразование метрики превращает все изотропные геодезические в полные [33].

Множитель a приобретает более широкий смысл в силу следующего результата.

Предложение 6.4.4

Пусть γ — замкнутая изотропная геодезическая кривая, которая неполна в будущем; тогда найдется вариация γ, которая сдвигает каждую точку γ в будущее, и при этом получается замкнутая времениподобная кривая.

Согласно разд. 2.6, на $\mathscr{M}$ можно построить поле линейных элементов $(\mathbf{V}, -\mathbf{V})$, нормированное так, что $g(\mathbf{V}, \mathbf{V}) = -1$. Поскольку мы предполагаем, что $\mathscr{M}$ ориентируемо во времени, можно согласованным образом выбрать одно из направлений $(\mathbf{V}, -\mathbf{V})$ и получить в результате направленное в будущее времениподобное векторное поле $\mathbf{V}$. Тогда можно ввести положи-

тельно-определенную метрику $\mathbf{g}'$:

$$g'(\mathbf{X}, \mathbf{Y}) = g(\mathbf{X}, \mathbf{Y}) + 2g(\mathbf{X}, \mathbf{V})\, g(\mathbf{Y}, \mathbf{V}).$$

Пусть t — неаффинный параметр на γ, который равен нулю в некоторой точке $q \in \gamma$ и для которого $g(\mathbf{V}, \partial/\partial t) = -1/\sqrt{2}$. Тогда t измеряет собственное расстояние вдоль γ в метрике $\mathbf{g}'$ и пробегает значения $-\infty < t < \infty$. Рассмотрим вариацию γ с вектором вариации $\partial/\partial u$, равным $x\mathbf{V}$, где x — функция $x(t)$. Согласно разд. 4.5,

$$\frac{1}{2}\frac{\partial}{\partial u} g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) = \frac{\mathrm{d}}{\mathrm{d}t} g\left(\frac{\partial}{\partial u}, \frac{\partial}{\partial t}\right) - g\left(\frac{\partial}{\partial u}, \frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial t}\right) =$$

$$= -2^{-1/2}\left(\frac{\mathrm{d}x}{\mathrm{d}t} - xf\right),$$

где $f(\partial/\partial t) = (\mathrm{D}/\partial t)(\partial/\partial t)$. Допустим теперь, что v — аффинный параметр на γ. Тогда $\partial/\partial v$ будет пропорционален $\partial/\partial t$: $\partial/\partial v = h\partial/\partial t$, где $h^{-1}\mathrm{d}h/\mathrm{d}t = -f$. После одного обхода по γ вектор $\partial/\partial v$ умножается на $a > 1$. Следовательно,

$$\oint f\, \mathrm{d}t = -\ln a \leqslant 0.$$

Поэтому, если мы возьмем функцию $x(t)$ равной

$$\exp\left(\int_0^t f(t')\, \mathrm{d}t' + b^{-1} t \ln a\right),$$

где $b = \oint \mathrm{d}t$, то получим вариацию γ в будущее и замкнутую времениподобную кривую. □

Предложение 6.4.5

Если

а) $R_{ab}K^aK^b \geqslant 0$ для любого изотропного вектора $\mathbf{K}$;

б) выполняется типовое условие: каждая изотропная геодезическая с касательным вектором $\mathbf{K}$ содержит точку, в которой $K_{[a}R_{b]\,cd\,[e}K_{f]}K^cK^d \neq 0$;

в) в $\mathscr{M}$ выполняется хронологическое условие,

то в $\mathscr{M}$ выполняется условие причинности.

Если бы существовали неполные замкнутые изотропные кривые, то в силу предыдущего предложения их можно было бы варьировать так, чтобы получить замкнутые времениподобные кривые. Если бы существовали полные замкнутые изотропные линии, то по предложению 4.4.5 они содержали бы сопряженные точки и, согласно предложению 4.5.12, их снова можно было бы варьировать к замкнутым времениподобным кривым. □

Из этого видно, что в физически реалистических решениях условие причинности и хронологическое условие эквивалентны.

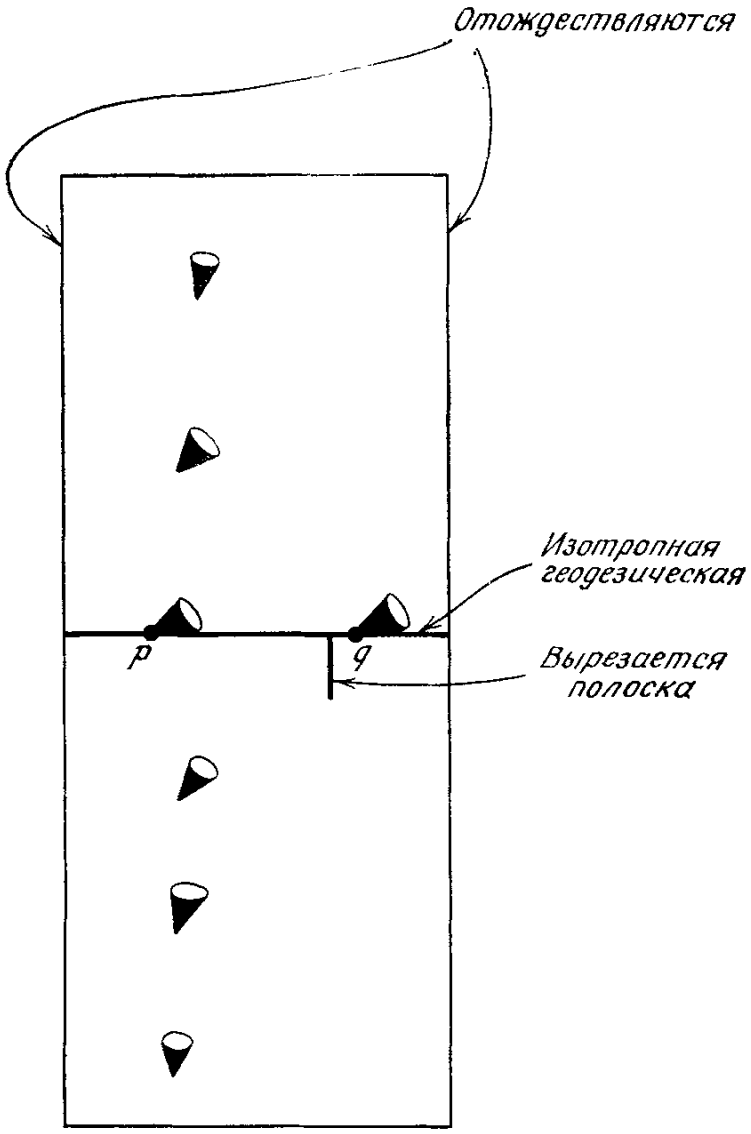


Рис. 37. Пространство, в котором всюду выполняются условие причинности и условие, выделяющее прошлое, но условие, выделяющее будущее, в p или q не выполняется (действительно, $I^+(p) = I^+(q)$). Наклон изотропных конусов на цилиндре увеличивается, пока одно изотропное направление не станет горизонтальным, и затем конусы снова выпрямляются; из пространства удалена полоса, так что она разрывает эту изотропную геодезическую, которая иначе была бы замкнута.

Подобно отказу от замкнутых времениподобных линий, представляется разумным исключить ситуации, в которых будут существовать непространственноподобные кривые, возвращающиеся сколь угодно близко к своей исходной точке или проходящие сколь угодно близко от таких непространственноподобных кривых, которые в свою очередь проходят бесконечно близко от исходной точки первой кривой и т. д. Картер [27] указал, что в

действительности имеется более чем счетная бесконечная иерархия таких условий причинности более высокого порядка; они зависят от числа и порядка рассматриваемых предельных процессов. Мы опишем первые три из этих условий и затем сформулируем предельно-сильное условие причинности.

Говорят, что в $p \in \mathscr{M}$ выполняется *условие, выделяющее будущее* (соответственно *прошлое*) [94], если в каждой окрестности p

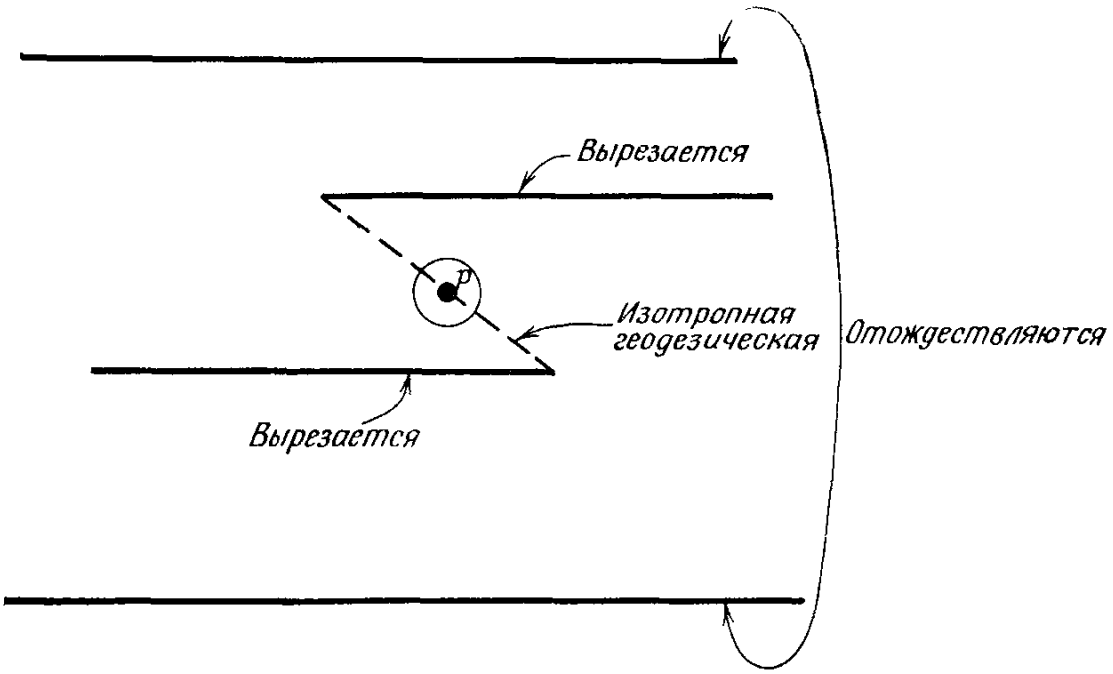


Рис. 38. Пространство-время, которое удовлетворяет условию причинности и условиям, выделяющим прошлое и будущее, но в точке p не удовлетворяет условию сильной причинности. С поверхности цилиндра удалены две полоски; световые конусы проходят под углом $\pm 45°$.

содержится такая окрестность p, которую никакая направленная в будущее (соответственно в прошлое) из p непространственноподобная кривая не пересекает более одного раза. Эквивалентное утверждение состоит в том, что из равенства $I^+(q) = I^+(p)$ (соответственно из $I^-(q) = I^-(p)$) следует, что $q = p$. На рис. 37 дан пример ситуации, когда условие причинности и условие, выделяющее прошлое, выполнены всюду, но условие, выделяющее будущее, не выполняется в точке p.

Говорят, что в p выполняется *условие сильной причинности*, если каждая окрестность p содержит такую окрестность p, которую никакая непространственноподобная кривая не пересекает более одного раза. Пример нарушения этого условия дан на рис. 38.

Предложение 6.4.6

Если условия (а)—(в) предложения 6.4.5 выполнены и, кроме того,

г) $\mathscr{M}$ — изотропно геодезически полное многообразие, то на $\mathscr{M}$ выполняется условие сильной причинности.

Допустим, что в $p \in \mathscr{M}$ условие сильной причинности не выполняется. Пусть $\mathscr{U}$ — нормальная выпуклая окрестность точки p и пусть $V_n \subset \mathscr{U}$ — бесконечная последовательность окрестностей p, таких, что любая окрестность p содержит все V_n с достаточно большими n. Тогда для каждого V_n будет существовать направленная в будущее непространственноподобная кривая λ_n, которая выйдет за пределы $\mathscr{U}$ и затем вернется в V_n. По лемме 6.2.1 найдется непродолжимая непространственноподобная кривая λ, которая будет предельной кривой для последовательности λ_n. Никакие две точки λ не могут иметь времениподобное разделение, иначе их можно было бы соединить некоторой λ_n, так чтобы получилась замкнутая непространственноподобная кривая. Таким образом, λ должна быть изотропной геодезической. Но по условиям (а), (б) и (г) λ должна содержать сопряженные точки и, следовательно, точки с времениподобным разделением. □

Следствие

На $\mathscr{M}$ будут выполняться также условия, выделяющие прошлое и будущее, ибо они содержатся в условии сильной причинности.

С этими тремя условиями причинности более высокого порядка тесно связано явление *захвата*. При построении направленной в будущее непространственноподобной кривой γ, которая непродолжима в будущее, возможны следующие варианты:

1) кривая войдет и останется в компактном множестве $\mathscr{S}$;

2) кривая не будет оставаться внутри какого-либо компактного множества, но будет постоянно возвращаться в компактное множество $\mathscr{S}$;

3) кривая не останется внутри какого-либо компактного множества $\mathscr{S}$ и не будет возвращаться более конечного числа раз в любое такое множество.

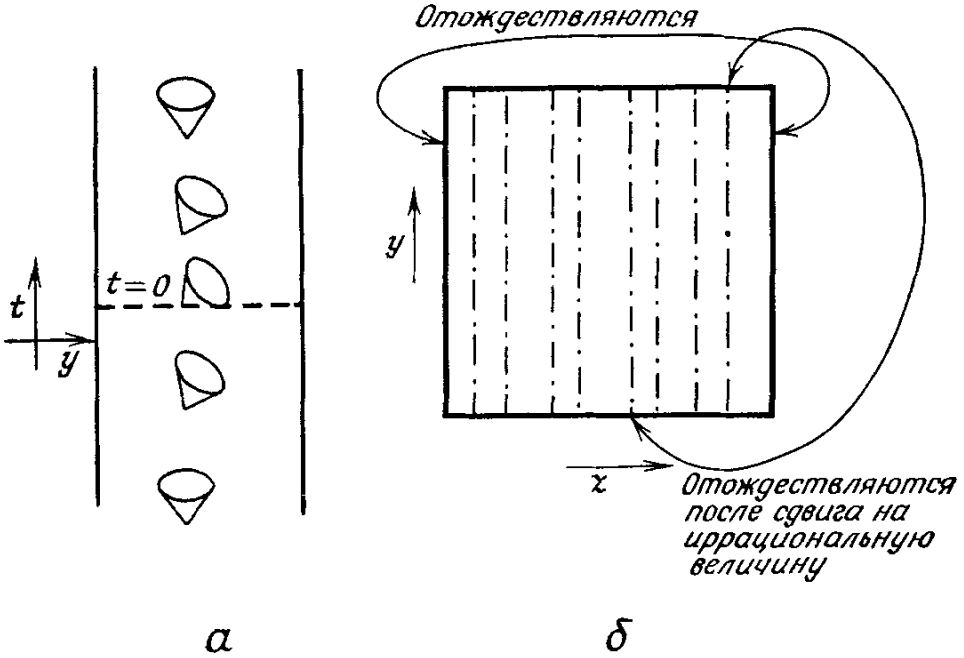
В третьем случае можно считать, что γ уходит к краю пространства-времени, т. е. или на бесконечность, или в сингулярность. В первом и во втором случае мы будем говорить, что *будущее* γ соответственно *полностью* или *частично* захвачено множеством $\mathscr{S}$. Может показаться, что захват имеет место только при нарушении условия причинности, но пример, принадлежащий Картеру (рис. 39), показывает, что это не обязательно. Тем не менее справедливо следующее предложение:

Предложение 6.4.7

Если условие сильной причинности выполняется на компактном множестве $\mathscr{S}$, то не может быть непродолжимых в будущее

непространственноподобных кривых с будущим, полностью или частично захваченным множеством $\mathscr{P}$.

$\mathscr{P}$ можно покрыть конечным числом выпуклых нормальных координатных окрестностей $\mathscr{U}_i$ с компактным замыканием, таких, что любую $\mathscr{U}_i$ любая непространственноподобная кривая пересекает не более одного раза. (Такие окрестности мы будем



Р и с. 39. Пространство с захваченными непространственноподобными линиями но без замкнутых непространственноподобных линий. Многообразие представляет собой $R^1 \times S^1 \times S^1$ и описывается координатами (t, y, z), причем (t, y, z) отождествляется с $(t, y, z+1)$ и с $(y, y+1, z+a)$, где a — иррациональное число. Лоренцева метрика имеет вид

$$ds^2 = (\operatorname{ch} t - 1)^2 (d t^2 - d y^2) + dt\, dy - d z^2.$$

а — сечение $\{z = \text{const}\}$, показывающее ориентацию изотропных конусов; *б* — сечение $t = 0$, показывающее часть изотропной геодезической.

называть *окрестностями локальной причинности.*) Любая непродолжимая в будущее непространственноподобная кривая, которая пересекает одну из этих окрестностей, должна ее покинуть и не возвращаться в нее. □

Предложение 6.4.8

Если на компактном множестве $\mathscr{P}$ выполняются условия, выделяющие будущее или прошлое, то не может существовать непродолжимая в будущее непространственноподобная кривая, будущее которой полностью захвачено $\mathscr{P}$. (Этот результат приведен здесь, поскольку он интересен сам по себе, но в дальнейшем он нам не понадобится.)

Пусть $\{\mathscr{V}_\alpha\}$, $(\alpha = 1, 2, 3)$ — счетный базис из открытых множеств в $\mathscr{M}$ (т. е. любое открытое множество в $\mathscr{M}$ можно предста-

вить как некоторое объединение множеств $\mathscr{V}_\alpha$). Поскольку на $\mathscr{S}$ выполняется условие, выделяющее будущее или прошлое, любая точка $p \in \mathscr{S}$ будет иметь нормальную выпуклую координатную окрестность $\mathscr{U}$, такую, что никакая направленная в будущее (соответственно в прошлое) непространственноподобная кривая, выходящая из p, не пересекает $\mathscr{U}$ более одного раза. Введем функцию $f(p)$, равную наименьшему значению α, при котором $\mathscr{V}_\alpha$ содержит p и сама содержится в некоторой окрестности $\mathscr{U}$.

Допустим, что имеется непродолжимая в будущее непространственноподобная кривая с будущим, полностью заключенным в $\mathscr{S}$. Пусть точка $q \in \lambda$ такова, что $\lambda' = \lambda \cap J^+(q)$ содержится в $\mathscr{S}$. Определим $\mathscr{A}_0$ как замкнутое непустое множество, состоящее из всех точек $\mathscr{S}$, которые являются предельными точками λ. Пусть $p_0 \in \mathscr{A}_0$ — та точка, в которой $f(p_0)$ равна наименьшему значению $f(p)$ на $\mathscr{A}_0$. Тогда через p_0 можно провести непродолжимую непространственноподобную кривую γ_0, каждая точка которой будет предельной точкой λ'. Никакие две точки γ_0 не могут иметь времениподобное разделение, иначе некоторый сегмент λ' можно было бы деформировать так, чтобы получить замкнутую непространственноподобную кривую. Следовательно, γ_0 будет непродолжимой изотропной геодезической, которая полностью захвачена в $\mathscr{S}$ и в направлении будущего, и в направлении прошлого. Пусть $\mathscr{A}_1$ — замкнутое множество, состоящее из всех предельных точек области $\gamma_0 \cap J^+(p_0)$ (или, в случае выполнения на $\mathscr{S}$ условия, выделяющего прошлое, $\gamma_0 \cap J^-(p_0)$). $\mathscr{A}_1 \subset \mathscr{A}_0$, поскольку каждая такая точка будет также и предельной точкой λ'. Ввиду того что $\mathscr{V}_{f(p_0)}$ не может содержать каких-либо предельных точек множества $\gamma_0 \cap J^+(p_0)$ (соответственно $\gamma_0 \cap J^-(p_0)$), $\mathscr{A}_1$ будет точно меньше $\mathscr{A}_0$. Так мы получим бесконечную последовательность замкнутых множеств $\mathscr{A}_0 \supset \mathscr{A}_1 \supset \supset \mathscr{A}_2 \supset \mathscr{A}_\beta \ldots$. Каждое $\mathscr{A}_\beta$ не пусто, ибо является множеством всех предельных точек изотропной геодезической с будущим (соответственно с прошлым), полностью захваченным в $\gamma_{\beta-1} \cap J^+(p_{\beta-1})$ (соответственно в $\gamma_{\beta-1} \cap J^-(p_{\beta-1})$). Пусть $\mathscr{K} = \bigcap_\beta \mathscr{A}_\beta$. Поскольку $\mathscr{S}$ компактно, $\mathscr{K}$ — непустое множество: пересечение любого конечного числа множеств $\mathscr{A}_\beta$ непусто. Допустим, $r \in \mathscr{K}$. Тогда $f(r) = f(p_\beta)$ для некоторого β. Но $\mathscr{V}_{f(p_\beta)} \cap \cap \mathscr{A}_{\beta+1}$ пусто, и поэтому r не может принадлежать множеству $\mathscr{A}_{\beta+1}$, а значит, и не может принадлежать $\mathscr{K}$. Отсюда ясно, что существование непродолжимой в будущее непространственноподобной кривой с полностью захваченным в $\mathscr{S}$ будущим невозможно.

Причинные отношения в $(\mathscr{M}, \mathbf{g})$ можно использовать для введения на $\mathscr{M}$ топологии, называемой *топологией Александрова*.

Эта топология, в которой по определению множество открыто в том и только в том случае, если оно является объединением одного или больше множеств вида $I^+(p) \cap I^-(q)$, $p, q \in \mathscr{M}$. $I^+(p) \cap I^-(q)$ открыто в топологии многообразия; поэтому любое множество, открытое в топологии Александрова, будет открытым и в топологии многообразия, хотя обратное не обязательно верно.

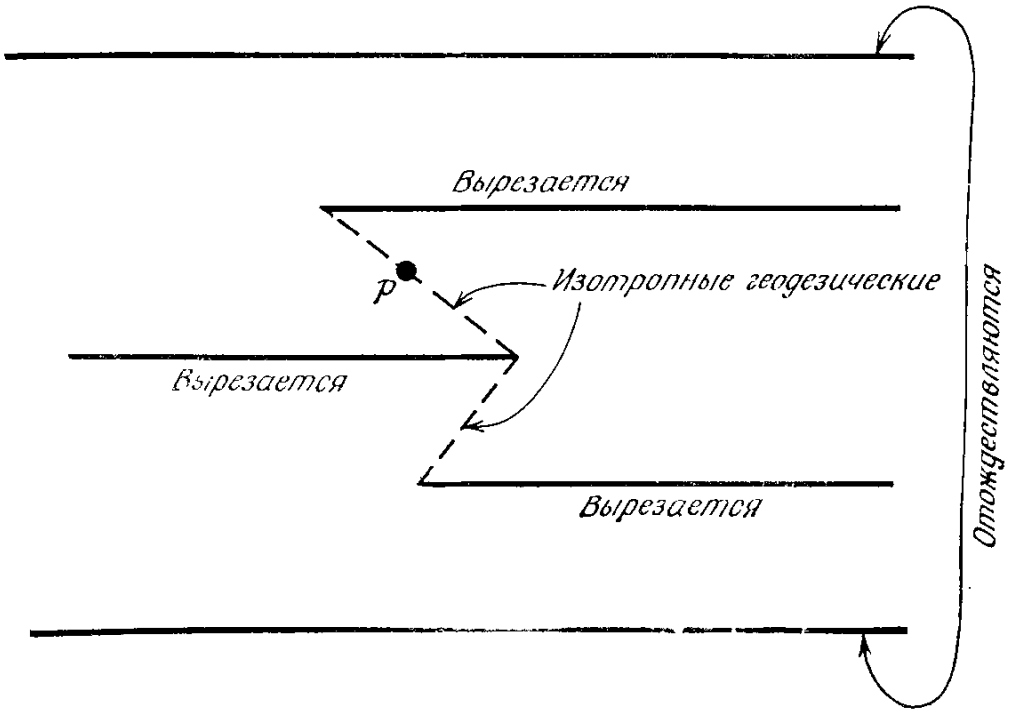


Рис. 40. Пространство, удовлетворяющее условию сильной причинности, в котором, однако, малейшее изменение метрики может привести к появлению замкнутых времениподобных линий, проходящих через p. С поверхности цилиндра удалены три полоски; световые конусы проходят под углом $\pm 45°$.

Допустим, однако, что на $\mathscr{M}$ выполняется условие сильной причинности. Тогда вокруг любой точки $r \in \mathscr{M}$ найдется окрестность локальной причинности $\mathscr{U}$. Топология Александрова пары $(\mathscr{U}, \mathbf{g}|_{\mathscr{U}})$, которая сама рассматривается как пространство-время, очевидно, совпадает с топологией многообразия в окрестности $\mathscr{U}$. Отсюда следует, что топология Александрова многообразия $\mathscr{M}$ совпадает с топологией многообразия, так как $\mathscr{M}$ можно покрыть окрестностями локальной причинности. Это означает, что при выполнении условия сильной причинности топологическая структура пространства-времени может быть установлена изучением причинных соотношений.

Как видно из рис. 40, даже наложение условия сильной причинности не ликвидирует все причинные патологии; пространство-время находится на грани нарушения хронологического условия в том смысле, что малейшее изменение метрики может

привести к появлению замкнутых кривых. Такого рода ситуация не выглядит физически реалистической, поскольку общая теория относительности, по-видимому, является классическим пределом еще неизвестной квантовой теории пространства-времени, а в

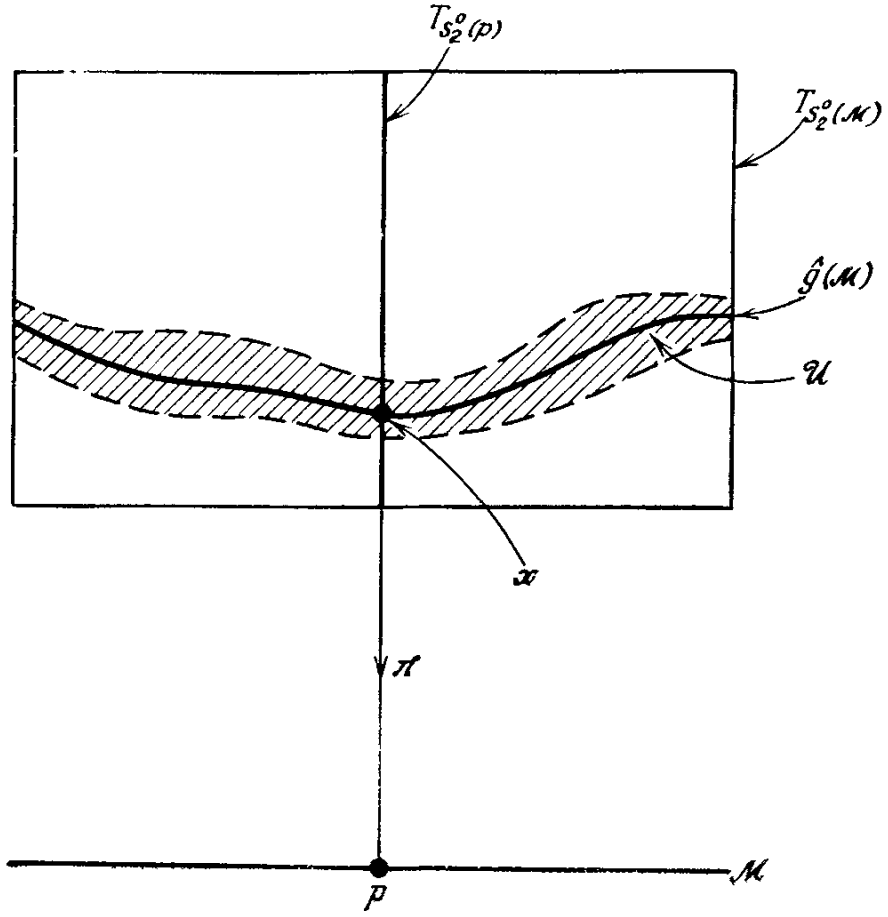


Рис. 41. Открытое множество $\mathcal{U}$ в открытой C^0-топологии на пространстве $T^0_{S2}(\mathcal{M})$ симметричных тензоров типа (0,2) на $\mathcal{M}$.

этой теории принцип неопределенности должен воспрепятствовать тому, чтобы метрика имела точное значение в каждой точке. Поэтому, чтобы быть физически осмысленным, то или иное свойство пространства-времени должно обладать определенного рода устойчивостью, т. е. этим свойством должны обладать и «близкие» пространства. Для придания точного смысла слову «близкий» необходимо задать некоторую топологию на множестве всех пространств-времен, т. е. всех некомпактных четырехмерных многообразий и лоренцевых метрик на них. Мы не будем касаться проблемы объединения многообразий с различными топологиями в односвязное топологическое пространство, а рассмотрим лишь введение топологии в множество всех лоренцевых C^r-метрик ($r \geqslant 1$) на данном многообразии. Это можно сделать разными способами в зависимости от того, требуем ли мы, чтобы

«близкая» метрика была близка только по своим значениям (C^0-топология) или же еще и по производным до k-го порядка (C^k-топология), а также от того, требуется ли близость везде (открытая топология) или только на компактных множествах (компактная открытая топология).

Для наших целей интересна *открытая C^0-топология*. Ее можно определить следующим образом. Пространства $T^0_{S2}(p)$ симметричных тензоров типа (0, 2) в каждой точке $p \in \mathscr{M}$ образуют многообразие (с естественной структурой). $T^0_{S2}(\mathscr{M})$ — расслоение симметричных тензоров типа (0, 2) над $\mathscr{M}$. Лоренцева метрика $\mathbf{g}$ на $\mathscr{M}$ есть установление соответствия между некоторым элементом $T^0_{S2}(\mathscr{M})$ и каждой точкой $p \in \mathscr{M}$, и, следовательно, ее можно рассматривать как такое отображение или сечение $\hat{g}: \mathscr{M} \to T^0_{S2}(\mathscr{M})$, для которого $\pi \circ \hat{g} = 1$, где π — проекция $T^0_{S2}(\mathscr{M}) \to \mathscr{M}$, переводящая $x \in T^0_{S2}(p)$ в p. Пусть $\mathscr{U}$ — открытое множество в $T^0_{S2}(\mathscr{M})$, и пусть $O(\mathscr{U})$ — множество всех лоренцевых C^0-метрик $\mathbf{g}$, таких, что $\hat{g}(\mathscr{M})$ содержится в $\mathscr{U}$ (рис. 41). Тогда открытое множество лоренцевых C^r-метрик в открытой C^0-топологии определяется как объединение одного или более множеств вида $O(\mathscr{U})$.

Будем говорить, что на $\mathscr{M}$ выполняется *условие устойчивой причинности*, если пространственно-временная метрика $\mathbf{g}$ имеет такую открытую окрестность в открытой C^0-топологии, что в любой метрике, принадлежащей этой окрестности, нет замкнутых времениподобных кривых. Использование C^k-топологии не привело бы к отличиям, но компактная открытая топология недопустима, так как в ней каждая окрестность любой метрики содержит замкнутые времениподобные кривые. Другими словами, условие устойчивой причинности означает, что в каждой точке можно слегка расширить световой конус, не получая при этом замкнутых времениподобных кривых.

Предложение 6.4.9

Условие устойчивой причинности выполняется всюду в $\mathscr{M}$, если и только если существует функция f на $\mathscr{M}$, градиент которой всюду времениподобен.

Замечание. Функцию f можно мыслить как род космического времени в том смысле, что она возрастает вдоль каждой направленной в будущее непространственноподобной кривой.

Доказательство. Из существования функции f с везде времениподобным градиентом следует условие устойчивой причинности, поскольку замкнутые времениподобные линии невозможны в любой метрике $\mathbf{h}$, которая близка к $\mathbf{g}$ настолько, что изотропный конус каждой точки $p \in \mathscr{M}$ в метрике $\mathbf{h}$ пересекает проходящую

через p поверхность $\{f = \text{const}\}$ только в p. Чтобы показать справедливость обратного утверждения, введем на $\mathscr{M}$ объемную меру μ (она не связана с объемной мерой, задаваемой метрикой $\mathbf{g}$), по которой полный объем $\mathscr{M}$ равен единице. Один из способов введения этой меры состоит в следующем: выберем для $\mathscr{M}$ счетный атлас $(\mathscr{U}_\alpha, \phi_\alpha)$, в котором $\overline{\phi_\alpha(\mathscr{U}_\alpha)}$ компактно в R^4. Пусть μ_0 — естественная евклидова мера на R^4, и пусть f_α — разложение единицы для атласа $(\mathscr{U}_\alpha, \phi_\alpha)$. Тогда μ можно определить как

$$\sum_\alpha f_\alpha 2^{-\alpha} [\mu_0(\mathscr{U}_\alpha)]^{-1} \phi_\alpha^* \mu_0.$$

Теперь, если выполняется условие устойчивой причинности, найдется семейство лоренцевых C^r-метрик $\mathbf{h}(a)$, $a \in [0, 3]$ со следующими свойствами:

1) $\mathbf{h}(0)$ является пространственно-временной метрикой $\mathbf{g}$;

2) для каждого $a \in [0, 3]$ в метрике $\mathbf{h}(a)$ нет замкнутых времениподобных кривых;

3) если a_1, $a_2 \in [0, 3]$ и $a_1 < a_2$, то каждый вектор, непространственноподобный в метрике $\mathbf{h}(a_1)$, будет времениподобным в метрике $\mathbf{h}(a_2)$.

Пусть $\theta(p, a)$, $p \in \mathscr{M}$ — объем области $I^-(p, \mathscr{M}, \mathbf{h}(a))$ по мере μ, где через $I^-(\mathscr{S}, \mathscr{U}, \mathbf{h})$ обозначено прошлое $\mathscr{S}$ относительно $\mathscr{U}$ в метрике $\mathbf{h}$. Для данного значения $a \in (0, 3)$ $\theta(p, a)$ будет ограниченной функцией, возрастающей вдоль каждой непространственноподобной кривой. Однако эта функция может не быть непрерывной: как видно из рис. 42, может оказаться, что небольшое изменение положения позволяет наблюдателю увидеть прошлое за «ширмой» и таким образом увеличить объем прошлого. Следовательно, нам нужно каким-то образом сгладить $\theta(p, a)$ так, чтобы получить непрерывную функцию, возрастающую вдоль каждой направленной в будущее и непространственноподобной в метрике $\mathbf{h}(0)$ кривой. Это можно сделать, проводя усреднение по области изменения a. Пусть

$$\bar{\theta}(p) = \int_1^2 \theta(p, a)\, da;$$

покажем, что $\bar{\theta}(p)$ непрерывна на $\mathscr{M}$.

Сначала нужно показать, что $\bar{\theta}(p)$ полунепрерывна сверху. Зададим $\varepsilon > 0$; пусть $\mathscr{B}$ — шар с центром в p, такой, что объем $\mathscr{B}$ по мере μ меньше, чем $\varepsilon/2$. По свойству (3) для a_1, $a_2 \in [0, 3]$ с $a_1 < a_2$ в $\mathscr{B}$ найдется окрестность $\mathscr{F}(a_1, a_2)$ точки p, для которой

$$[I^-(\mathscr{F}(a_1 a_2), \bar{\mathscr{B}}, \mathbf{h}(a_1)) \cap \dot{\mathscr{B}}] \subset [I^-(p, \bar{\mathscr{B}}, \mathbf{h}(a_2)) \cap \dot{\mathscr{B}}].$$

Пусть n — положительное целое число, большее чем $2\varepsilon^{-1}$. Тогда мы определим множество $\mathcal{G} = \bigcap_i \mathcal{F}\left(1+\frac{1}{2}in^{-1},\ 1+\frac{1}{2}(i+1)n^{-1}\right)$, $i=0, 1, \ldots, 2n$. $\mathcal{G}$ является окрестностью p и содержится в $\mathcal{F}(a, a+n^{-1})$ при любом $a \in [1, 2]$. Поэтому $I^-(q, \mathcal{M}, \mathbf{h}(a)) - \bar{\mathcal{B}}$

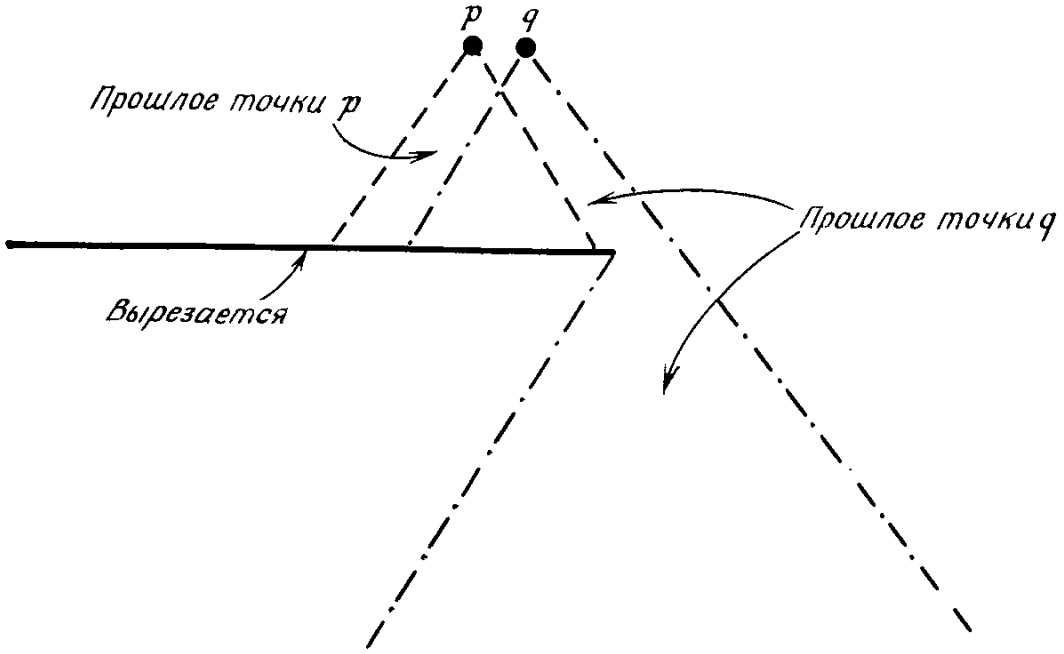


Рис. 42. Небольшой сдвиг точки из p в q приводит к большому изменению объема прошлого этой точки. Световые конусы проходят под углом ±45°; удалена полоска.

будет содержаться в $I^-(p, \mathcal{M}, \mathbf{h}(a+n^{-1})) - \bar{\mathcal{B}}$, если $q \in \mathcal{G}$ и $a \in [1, 2]$. Отсюда

$$\theta(q, a) \leqslant \theta\left(p,\ a+\tfrac{1}{2}\varepsilon\right) + \tfrac{1}{2}\varepsilon$$

и, следовательно, $\bar{\theta}(q) \leqslant \bar{\theta}(p) + \varepsilon$, из чего видно, что $\bar{\theta}$ полунепрерывна сверху. Аналогично доказывается, что она полунепрерывна снизу. Чтобы получить дифференцируемую функцию, можно усреднить $\bar{\theta}$ с подходящей сглаживающей функцией по окрестности каждой точки. Выбирая достаточно малую окрестность, можно получить функцию f, градиент которой везде времениподобен по метрике $\mathbf{g}$. Детали этой процедуры сглаживания можно найти в работе [167]. □

Пространственноподобные поверхности $\{f = \text{const}\}$ можно рассматривать в качестве поверхностей одновременности в пространстве-времени, хотя они, конечно, не единственны. Если эти поверхности компактны, то они все диффеоморфны одна другой, но это не обязательно справедливо, если некоторые из них некомпактны.

6.5. Области Коши

В теории Ньютона имеется мгновенное дальнодействие; поэтому, чтобы предсказать события в будущих точках пространства-времени, мы должны знать состояние всей Вселенной в настоящий момент времени, а также наложить некоторые граничные условия на бесконечности, вроде того, что потенциал стремится к нулю. Напротив, из постулата (а) разд. 3.2 следует, что в теории относительности события в различных точках пространства-времени могут быть причинно связаны только в том случае, если их можно соединить непространственноподобной кривой. Таким образом, знание соответствующих данных на замкнутом множестве $\mathscr{S}$ (если данные известны на открытом множестве, то на его замыкании они получаются в силу непрерывности) позволит устанавливать события в области $D^+(\mathscr{S})$ будущего $\mathscr{S}$, называемой *областью Коши будущего* множества $\mathscr{S}$ и определяемой как множество всех таких точек $p \in \mathscr{M}$, для которых каждая непродолжимая в прошлое непространственноподобная кривая, проходящая через p, пересекает $\mathscr{S}$ (отметим, что $D^+(\mathscr{S}) \supset \mathscr{S}$).

Пенроуз [127, 128] определяет область Коши множества $\mathscr{S}$ немного иначе, а именно как множество всех точек $p \in \mathscr{M}$, таких, что каждая непродолжимая в прошлое времениподобная кривая, проходящая через p, пересекает $\mathscr{S}$. Такое множество мы будем обозначать $\tilde{D}^+(\mathscr{S})$. Справедлив следующий результат:

Предложение 6.5.1

$\tilde{D}^+(\mathscr{S}) = \overline{D^+}(\mathscr{S})$.

Очевидно, $\tilde{D}^+(\mathscr{S}) \supset D^+(\mathscr{S})$. Если точка $q \in \mathscr{M} - \tilde{D}^+(\mathscr{S})$, то у нее есть окрестность $\mathscr{U}$, не пересекающая $\mathscr{S}$. Из q можно провести непродолжимую в прошлое кривую λ, которая не пересекает $\mathscr{S}$. Если $r \in \lambda \cap I^-(q, \mathscr{U})$, то $I^+(r, \mathscr{U})$ является открытой окрестностью q в $\mathscr{M} - \tilde{D}^+(\mathscr{S})$. Следовательно, $\mathscr{M} - \tilde{D}^+(\mathscr{S})$ — открытое множество, а $\tilde{D}^+(\mathscr{S})$ — замкнутое. Допустим, что существует точка $p \in \tilde{D}^+(\mathscr{S})$, у которой имеется окрестность $\mathscr{V}$, не пересекающая $D^+(\mathscr{S})$. Выберем точку $x \in I^-(p, \mathscr{V})$. Из нее можно провести непродолжимую непространственноподобную кривую γ, которая не пересекает $\mathscr{S}$. Пусть y_n — последовательность точек на γ, которая не сходится к какой-либо точке, причем y_{n+1} лежит в прошлом y_n. Пусть $\mathscr{W}_n$ — нормальные выпуклые окрестности точек y_n, причем $\mathscr{W}_{n+1}$ не пересекается с $\mathscr{W}_n$. Пусть, наконец, z_n есть последовательность точек, такая, что

$$z_{n+1} \in I^+(y_{n+1}, \mathscr{W}_{n+1}) \cap I^-(z_n, \mathscr{M} - \mathscr{S}).$$

Тогда должна существовать непродолжимая времениподобная кривая с началом в p, которая проходит через каждую точку z_n и не пересекает $\mathscr{S}$. Но это противоречит условию, что $p \in \tilde{D}^+(\mathscr{S})$. Следовательно, $\overline{\tilde{D}^+(\mathscr{S})}$ содержится в замыкании $D^+(\mathscr{S})$, и поэтому $\tilde{D}^+(\mathscr{S}) = D^+(\mathscr{S})$. □

Граница области $D^+(\mathscr{S})$ в будущем, т. е. $\overline{D^+(\mathscr{S})} - I^-(D^+(\mathscr{S}))$, очерчивает пределы области, которая может быть предсказана

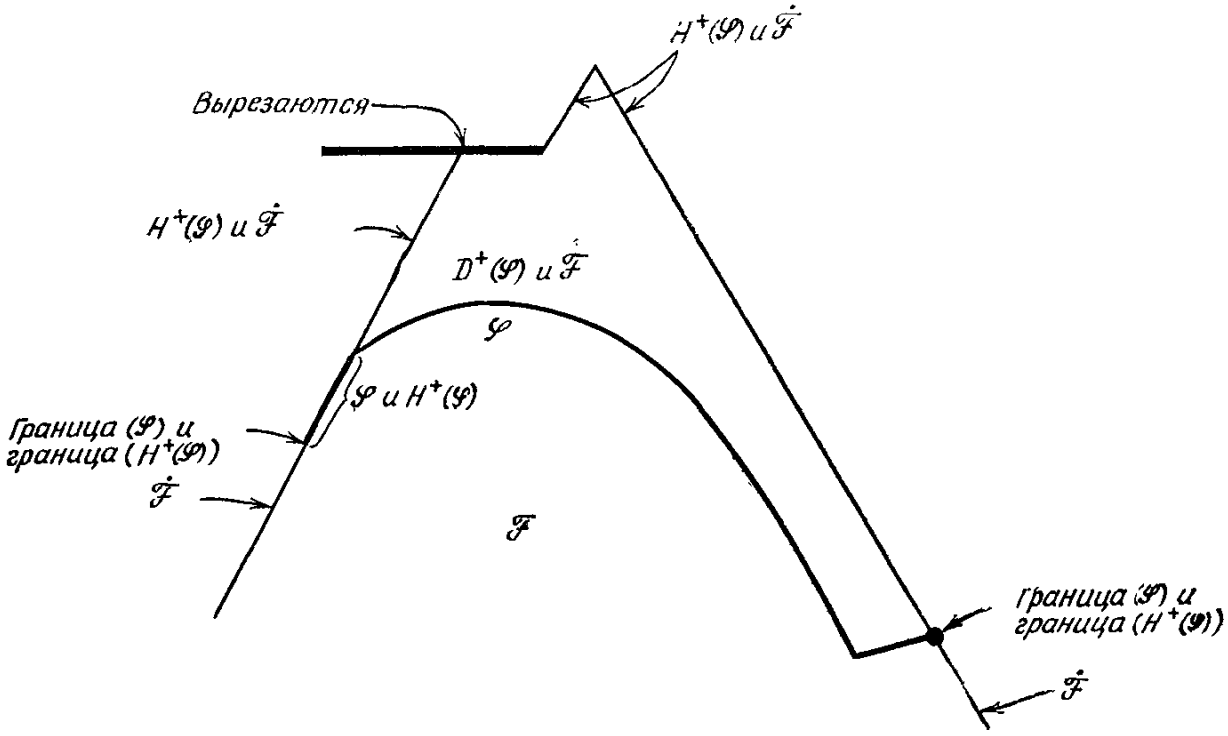


Рис. 43. Область Коши будущего $D^+(\mathscr{S})$ и горизонт Коши в будущем $H^+(\mathscr{S})$ замкнутого множества $\mathscr{S}$, которое частично изотропно и частично пространственноподобно. Отметим, что горизонт $H^+(\mathscr{S})$ не обязательно связен. Изотропные линии проходят под углом $\pm 45°$; удалена полоска.

по известным данным на $\mathscr{S}$. Это замкнутое ахрональное множество мы будем называть *горизонтом Коши в будущем* множества $\mathscr{S}$ и обозначим его через $H^+(\mathscr{S})$; как видно из рис. 43, $H^+(\mathscr{S})$ пересекает $\mathscr{S}$, если множество $\mathscr{S}$ изотропно или имеет «границу». Чтобы придать этому выражению точный смысл, определим границу $(\overline{\mathscr{S}})$ ахронального множества $\mathscr{S}$ как множество всех точек $q \in \overline{\mathscr{S}}$, в каждой окрестности $\mathscr{U}$ которых есть точки $p \in I^-(q, \mathscr{U})$ и $r \in I^+(q, \mathscr{U})$, которые можно соединить непересекающей $\mathscr{S}$ времениподобной кривой в $\mathscr{U}$. В силу аргументов, аналогичных тем, что были приведены в доказательстве предложения 6.3.1, получаем, что $\mathscr{S}$ является трехмерным вложенным C^1-подмногообразием, если непустое ахрональное множество $\mathscr{S}$ не имеет границы.

Предложение 6.5.2

Для замкнутого ахронального множества $\mathscr{S}$ граница $(H^+(\mathscr{S}))$ совпадает с границей $(\mathscr{S})$.

Пусть $\mathscr{U}_n$ — последовательность окрестностей точки q, принадлежащей границе $(H^+(\mathscr{S}))$, такая, что любая окрестность q содержит все $\mathscr{U}_n$ с достаточно большим n. В каждой окрестности $\mathscr{U}_n$ найдутся точки $\tilde{p} \in I^-(q, \mathscr{U}_n)$ и $r_n \in I^+(q, \mathscr{U}_n)$, которые можно соединить времениподобной кривой λ_n, не пересекающей $H^+(\mathscr{S})$. Это означает, что λ_n не может пересечь $D^+(\mathscr{S})$. По предложению 6.5.1 $q \in \tilde{D}^+(\mathscr{S})$ и, следовательно, $I^-(q) \subset I^-(\tilde{D}^+(\mathscr{S})) \subset I^-(\mathscr{S}) \cup \tilde{D}^+(\mathscr{S})$. Таким образом, p_n должна лежать в $I^-(\mathscr{S})$. Кроме того, каждая непродолжимая в направлении прошлого времениподобная кривая с началом в q должна пересечь $\mathscr{S}$. Поэтому при каждом n на каждой времениподобной линии в $\mathscr{U}_n$, соединяющей q и p_n, должна быть точка множества $\mathscr{S}$ и, значит, q должна принадлежать $\bar{\mathscr{S}}$. Так как кривые λ_n не пересекают $\mathscr{S}$, q лежит на границе $(\mathscr{S})$. Аналогично доказывается обратное утверждение. □

Предложение 6.5.3

Пусть $\mathscr{S}$ — замкнутое ахрональное множество. Тогда $H^+(\mathscr{S})$ порождается изотропными геодезическими сегментами, которые или вообще не имеют конечных точек в прошлом, или имеют конечную точку в прошлом на границе $(\mathscr{S})$.

Множество $\mathscr{F} \equiv \tilde{D}^+(\mathscr{S}) \cup I^-(\mathscr{S})$ является множеством прошлого. Следовательно, в силу предложения 6.3.1, $\dot{\mathscr{F}}$ — ахрональное C^1-многообразие. $H^+(\mathscr{S})$ — замкнутое подмногообразие многообразия $\dot{\mathscr{F}}$. Пусть $q \in (H^+(\mathscr{S})$ — граница $(\mathscr{S}))$. Если $q \notin \mathscr{S}$, то $q \in I^+(\mathscr{S})$, поскольку $q \in \tilde{D}^+(\mathscr{S})$. В силу ахрональности $\mathscr{S}$ найдется нормальная выпуклая окрестность $\mathscr{W}$ точки q, которая не пересекает $I^-(\mathscr{S})$. Пусть теперь $q \in \mathscr{S}$, и пусть $\mathscr{W}$ — нормальная выпуклая окрестность q, такая, что ни одна точка из $I^+(q, \mathscr{W})$ не может быть соединена с какой-либо точкой из $I^-(q, \mathscr{W})$ времениподобной кривой, не пересекающей $\mathscr{S}$. В любом случае, если p — какая-либо точка в $I^+(q)$, должна найтись направленная в прошлое времениподобная кривая от p до некоторой точки из $\mathscr{M} - \mathscr{F} - \mathscr{W}$, поскольку иначе p лежала бы в $D^+(\mathscr{S})$. Отсюда, согласно условию (1) леммы 6.3.2, примененному к множеству будущего $\mathscr{M} - \mathscr{F}$, имеем $q \in \dot{\mathscr{F}}_N \cup \dot{\mathscr{F}}_+$. □

Следствие

Если граница $(\mathscr{S}) = \varnothing$, то $H^+(\mathscr{S})$, когда он не пуст, является ахрональным трехмерным вложенным C^1-многообразием, которое порождено изотропными геодезическими сегментами, не имеющими конечных точек в прошлом.

Непричинное множество $\mathscr{S}$ без края мы будем называть *частичной поверхностью Коши*. Иначе говоря, частичная поверхность Коши — это пространственноподобная гиперповерхность, которую ни одна непространственноподобная кривая не пересекает более одного раза. Допустим, что существует связная пространственноподобная гиперповерхность $\mathscr{S}$ (без края), пересекаемая непространственноподобной кривой в точках p_1 и p_2. Тогда p_1 и p_2 можно соединить кривой μ, лежащей в $\mathscr{S}$, и $\mu \cup \lambda$ будет замкнутой кривой, которая пересекает $\mathscr{S}$ только один раз. Эту кривую нельзя непрерывным образом стянуть в нуль, поскольку такая деформация может изменить число пересечений кривой с $\mathscr{S}$ только на четное число. Таким образом, $\mathscr{M}$ не может быть односвязным. Это означает, что мы могли бы «развернуть» $\mathscr{M}$, перейдя к односвязному универсальному накрывающему многообразию для $\mathscr{M}$, в котором каждая связная компонента образа $\mathscr{S}$ является пространственноподобной гиперповерхностью (без края), а следовательно, частичной поверхностью Коши в $\tilde{\mathscr{M}}$ (рис. 44). Однако переход к универсальному накрывающему многообразию может «развернуть» $\mathscr{M}$ больше, чем требуется, и дать в результате такую частичную поверхность Коши, которая будет некомпактной даже тогда, когда $\mathscr{S}$ компактно. Для целей последующих глав было бы желательно иметь накрывающее многообразие, которое развертывает $\mathscr{M}$ так, чтобы каждая связная компонента образа $\mathscr{S}$ была частичной поверхностью Коши, но при этом оставалась гомеоморфной $\mathscr{S}$. Такое накрывающее многообразие можно построить по крайней мере двумя различными способами.

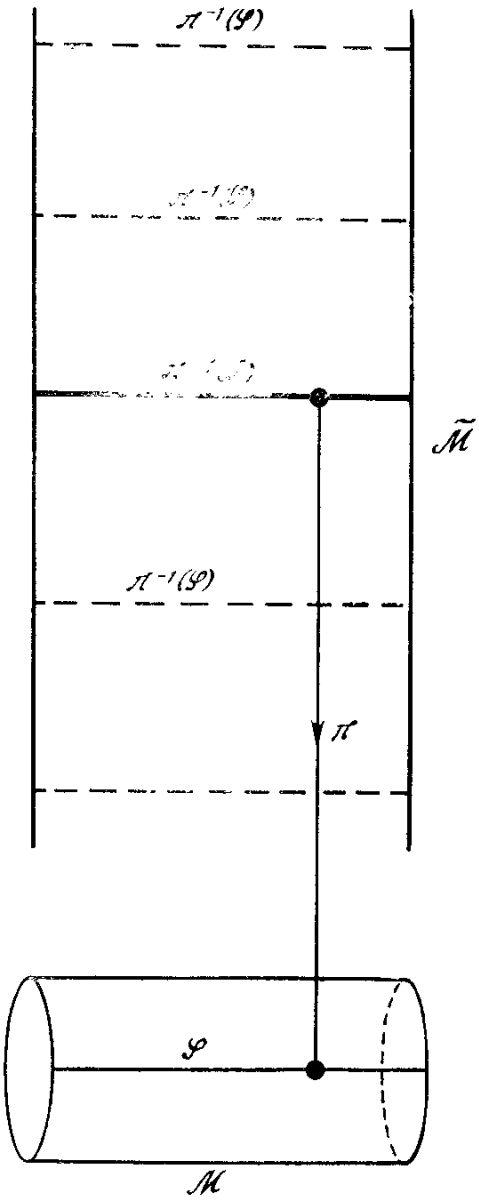


Рис. 44. $\mathscr{S}$ — связная пространственноподобная гиперповерхность без края в $\mathscr{M}$. Она не является частичной поверхностью Коши, однако каждый образ $\pi^{-1}(\mathscr{S})$ поверхности $\mathscr{S}$ в универсальном накрывающем многообразии $\tilde{\mathscr{M}}$ многообразия $\mathscr{M}$ представляет собой частичную поверхность Коши в $\tilde{\mathscr{M}}$.

Напомним, что универсальное накрывающее многообразие можно определить как множество всех пар $(p, [\lambda])$, где $p \in \mathscr{M}$ и $[\lambda]$ — класс эквивалентности кривых в $\mathscr{M}$ от некоторой фиксированной точки $q \in \mathscr{M}$ до p, гомотопных по модулю q и p. Накрывающее многообразие $\mathscr{M}_H$ определяется как множество всех пар $(p, [\lambda])$, где $[\lambda]$ — класс эквивалентности кривых от $\mathscr{S}$ до p, гомотопных по модулю $\mathscr{S}$ и p (т. е. концы на поверхности $\mathscr{S}$ могут сдвигаться по ней). Можно сказать, что $\mathscr{M}_H$ — это наибольшее накрывающее многообразие, в котором каждая связная компонента образа $\mathscr{S}$ гомеоморфна $\mathscr{S}$. Накрывающее многообразие $\mathscr{M}_G$ [52] по определению — это множество всех пар $(p, [\lambda])$, где на этот раз $[\lambda]$ является классом эквивалентности кривых от фиксированной точки q до p, которые пересекают $\mathscr{S}$ одинаковое число раз, причем пересечение в направлении будущего считается положительным, а в направлении прошлого — отрицательным. Можно сказать, что $\mathscr{M}_G$ является наименьшим накрывающим многообразием, в котором каждая связная компонента образа $\mathscr{S}$ делит $\mathscr{M}_G$ на две части. В обоих случаях топологическая и дифференциальная структура накрывающего многообразия фиксируется требованием, чтобы проекция, отображающая $(p, [\lambda])$ в p, локально была диффеоморфизмом.

Введем $D(\mathscr{S}) = D^+(\mathscr{S}) \cup D^-(\mathscr{S})$. Частичная поверхность Коши $\mathscr{S}$ называется *глобальной поверхностью Коши* (или просто *поверхностью Коши*), если $D(\mathscr{S}) = \mathscr{M}$. Таким образом, поверхность Коши — это пространственноподобная гиперповерхность, которую каждая непространственноподобная кривая пересекает точно один раз. Примерами поверхности Коши в пространстве Минковского служат поверхности $\{x^4 = \text{const}\}$, но гиперболоиды

$$\{(x^4)^2 - (x^3)^2 - (x^2)^2 - (x^1)^2 = \text{const}\}$$

являются только частичными поверхностями Коши, поскольку изотропный конус будущего или прошлого в начале системы координат будут горизонтами Коши для этих поверхностей (см. разд. 5.1 и рис. 13). То, что данная поверхность является поверхностью Коши, — это не только свойство самой поверхности, но также и свойство объемлющего пространства-времени в целом. Например, достаточно вырезать из пространства Минковского только одну точку, как полученное пространство-время вообще не будет допускать ни одной поверхности Коши.

Если для рассматриваемого многообразия $\mathscr{M}$ существует поверхность Коши, то, имея нужные данные на ней, мы могли бы определить состояние Вселенной в любой момент прошлого или будущего. Однако мы не можем собрать такие данные, не побывав в будущем каждой точки поверхности, что в большинстве случаев невозможно. Мы не располагаем вескими физическими

доводами для уверенности в том, что Вселенная допускает поверхность Коши; имеется целый ряд известных точных решений уравнения Эйнштейна, которые ее не допускают. Среди них описанные в гл. 5 пространство де Ситтера 2-го рода, плоские волны, пространство Тауба — НУТ, решение Райсснера — Нордстрема. Последнее представляет собой особо интересный случай: поверхность $\mathscr{S}$ на рис. 25 позволяет предсказать события во внешних областях I, где $r > r_+$, и в соседних областях II, где $r_- < r < r_+$, но при этом существует горизонт Коши $r = r_-$. Точки соседней области III не лежат в $D^+(\mathscr{S})$, ибо в этой области имеются непространственноподобные кривые, которые непродолжимы в направлении прошлого и не пересекают $r = r_-$, а стремятся к точкам i^+ (их можно считать расположенными на бесконечности) или к сингулярности $r = 0$ (которую нельзя рассматривать как часть пространства-времени, см. разд. 8.1). В эту область может поступать дополнительная информация из бесконечности или сингулярности, которая нарушит любые предсказания, сделанные на основе одних только данных на $\mathscr{S}$. Итак, в общей теории относительности наша способность предсказывать будущее ограничена как трудностями, связанными с необходимостью знать данные на всей пространственноподобной гиперповерхности, так и тем, что даже такого знания может оказаться недостаточно. Однако, несмотря на эти ограничения, все же можно предсказать, правда, при определенных условиях существование сингулярностей.

6.6. Глобальная гиперболичность

С областями развития тесно связано свойство глобальной гиперболичности [99]. Множество $\mathscr{N}$ называется *глобально гиперболическим*, если на $\mathscr{N}$ выполняется предположение о сильной причинности и если для любых двух точек p, $q \in (\mathscr{N})$, $J^+(p) \cap \cap J^-(q)$ компактно и содержится в $\mathscr{N}$. В некотором смысле можно считать, что это определение говорит о том, что ни одна точка из $J^+(p) \cap J^-(q)$ не принадлежит краю пространства-времени, т. е. бесконечности или сингулярности. Термин «глобальная гиперболичность» возник в связи с тем, что волновое уравнение с источником в виде δ-функции в $p \in \mathscr{N}$ имеет в $\mathscr{N}$ единственное решение, обращающееся в нуль вне $\mathscr{N} - J^+(p, \mathscr{N})$ (см. гл. 7).

Напомним, что $\mathscr{N}$ называют *причинно-простым*, если для любого компактного множества $\mathscr{K}$, содержащегося в $\mathscr{N}$, $J^+(\mathscr{K}) \cap \cap \mathscr{N}$ и $J^-(\mathscr{K}) \cap \mathscr{N}$ замкнуты в $\mathscr{N}$.

Предложение 6.6.1

Открытое глобальное гиперболическое множество $\mathscr{N}$ — причинно простое.

Пусть p — любая точка множества $\mathcal{N}$. Допустим, что есть точка

$$q \in \left(\overline{J^+(p)} - J^-(p)\right) \cap \mathcal{N}.$$

Поскольку $\mathcal{N}$ открыто, должна быть точка $r \in \left(I^+(q) \cap \mathcal{N}\right)$. Но тогда $q \in J^+(p) \cap J^-(r)$, что невозможно, так как $J^+(p) \cap J^-(r)$ должно быть компактным и поэтому замкнутым. Таким образом, $J^+(p) \cap \mathcal{N}$ и $J^-(p) \cap \mathcal{N}$ замкнуты в $\mathcal{N}$.

Допустим теперь, что существует точка $q \in \left(\overline{J^+(\mathcal{K})} - J^+(\mathcal{K})\right) \cap \mathcal{N}$. Пусть q_n — бесконечная последовательность точек $I^+(q) \cap \mathcal{N}$, сходящаяся к q, причем $q_{n+1} \in I^-(q_n)$. Для каждого n $J^-(q_n) \cap \mathcal{K}$ будет компактным непустым множеством. Поэтому $\bigcap_n \{J^-(q_n) \cap \mathcal{K}\}$ будет непустым множеством. Пусть p — точка этого множества. Тогда q_n при всех n будут лежать в $J^+(p)$. Но $J^+(p)$ замкнуто. Поэтому $J^+(p)$ содержит q. □

Следствие

Если $\mathcal{K}_1$ и $\mathcal{K}_2$ — компактные множества в $\mathcal{N}$, $J^+(\mathcal{K}_1) \cap J^-(\mathcal{K}_2)$ компактно.

Можно найти конечное число точек $p_i \in \mathcal{N}$, для которых

$$\left\{\bigcup_i J^+(p_i)\right\} \supset \mathcal{K}_1.$$

Аналогично, существует конечное число точек q_j, таких, что $\mathcal{K}_2$ содержится в

$$\bigcup_j J^-(q_j).$$

Тогда $J^+(\mathcal{K}_1) \cap J^-(\mathcal{K}_2)$ содержится в

$$\bigcup_{i,j} \{J^+(p_i) \cap J^-(q_j)\}$$

и замкнуто. □

В работе [99] Лере дал иное, но эквивалентное введенному выше определение глобальной гиперболичности; сейчас мы его сформулируем. Для точек $p, q \in \mathcal{M}$, для которых в $J^+(p) \cap J^-(q)$ выполняется условие сильной причинности, мы введем $C(p, q)$ — пространство всех (непрерывных) непространственноподобных кривых, соединяющих p и q; при этом принимается, что две кривые $\gamma(t)$ и $\lambda(u)$ представляют одну и ту же точку пространства $C(p, q)$, если они получаются одна из другой перепараметризацией, т. е. если существует непрерывная монотонная функция $f(u)$, такая, что $\gamma(f(u)) = \lambda(u)$. (Пространство $C(p, q)$ можно

ввести, даже если на $J^+(p) \cap J^-(q)$ условие сильной причинности не выполняется.) Топология $C(p, q)$ определяется заданием окрестностей кривой $\gamma \subset C(p, q)$, состоящих из всех тех кривых в $C(p, q)$, точки которых в $\mathscr{M}$ лежат в окрестности $\mathscr{W}$ точек γ (рис. 45). Определение Лере [99] состоит в том, что множество $\mathscr{N}$ называется глобально гиперболическим, если $C(p, q)$ компактно для всех $p, q \in \mathscr{N}$. Эквивалентность этих определений видна из следующего результата:

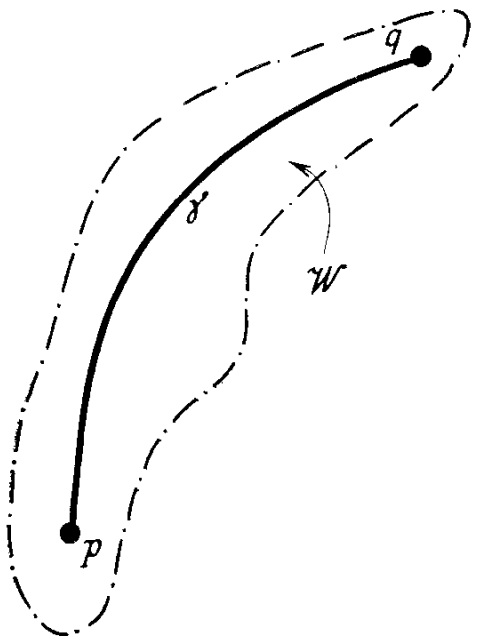


Рис. 45. Окрестность $\mathscr{W}$ точек кривой γ в $\mathscr{M}$. Окрестность γ в $C(p, q)$ состоит из всех кривых от p до q, точки которых лежат в $\mathscr{W}$.

Предложение 6.6.2 (*Зейферт* [*156*], *Герок* [*52*])
Пусть на открытом множестве $\mathscr{N} = J^-(\mathscr{N}) \cap J^+(\mathscr{N})$ выполняется сильная причинность. Тогда для глобальной гиперболичности $\mathscr{N}$ необходимо и достаточно, чтобы пространство $C(p, q)$ было компактно для всех $p, q \in \mathscr{N}$.

Предположим сначала, что $C(p, q)$ компактно. Пусть r_n — бесконечная последовательность точек $J^+(p) \cap J^-(q)$, и пусть λ_n — последовательность непространственноподобных кривых от p до q, проходящих через соответствующую r_n. Ввиду компактности $C(p, q)$ существует кривая λ, к которой в топологии $C(p, q)$ сходится некоторая подпоследовательность λ'_n. Пусть $\mathscr{U}$ — такая окрестность кривой λ в $\mathscr{M}$, что $\overline{\mathscr{U}}$ компактно. Тогда $\mathscr{U}$ будет содержать все λ'_n, а значит, и все r'_n при достаточно больших n; следовательно, имеется точка $r \in \mathscr{U}$, которая является предельной точкой последовательности r'_n. Очевидно, r лежит на λ. Таким образом, любая бесконечная последовательность в $J^+(p) \cap J^-(q)$ имеет предельную точку в $J^+(p) \cap J^-(q)$. Поэтому $J^+(p) \cap J^-(q)$ компактно.

Обратно, допустим $J^+(p) \cap J^-(q)$ компактно. Пусть λ_n — бесконечная последовательность непространственноподобных кривых от p до q. По лемме 6.2.1, примененной к открытому множеству $\mathscr{M} - q$, в нем существует направленная в будущее непространственноподобная кривая λ с началом в p, которая непродолжима в $\mathscr{M} - q$ и для которой найдется подпоследовательность λ'_n, сходящаяся к r при каждом $r \in \lambda$. Кривая λ должна иметь конечную точку q в будущем, поскольку, согласно предложению 6.4.7, она не может быть полностью заключена в будущем в компактном множестве $J^+(p) \cap J^-(p)$ и не может покинуть это множество иначе как в q.

Пусть $\mathscr{U}$ — любая окрестность λ в $\mathscr{M}$, и пусть r_i $(1 \leqslant i \leqslant k)$ — конечное множество точек на λ, таких, что $r_1 = p$, $r_k = q$ и каждая r_i имеет окрестность $\mathscr{V}_i$, удовлетворяющую условию $J^+(\mathscr{V}_i) \cap \cap J^-(\mathscr{V}_i) \subset \mathscr{U}$. Тогда для достаточно больших n $\lambda'_n \subset \mathscr{U}$. Следовательно, λ'_n сходится к λ в топологии $C(p, q)$, и пространство компактно. □

Связь между глобальной гиперболичностью и областями Коши устанавливается следующими утверждениями:

Предложение 6.6.3

Если $\mathscr{S}$ — замкнутое ахрональное множество, то

$$\operatorname{int} D(\mathscr{S}) \equiv D(\mathscr{S}) - \dot{D}(\mathscr{S}),$$

если оно не пусто, является глобально гиперболическим множеством.

Докажем сначала несколько лемм.

Лемма 6.6.4

Если $p \in D^+(\mathscr{S}) - H^+(\mathscr{S})$, то каждая непродолжимая в прошлое непространственноподобная кривая, проходящая через p, пересекает $I^-(\mathscr{S})$.

Пусть $p \in D^+(\mathscr{S}) - H^+(\mathscr{S})$ и γ — непродолжимая в будущее непространственноподобная кривая, проходящая через q. Тогда можно найти точку $q \in D^+(\mathscr{S}) \cap I^+(p)$ и непродолжимую в прошлое непространственноподобную кривую λ, проходящую через нее, такие, что для каждой точки $x \in \lambda$ найдется точка $y \in \gamma$, лежащая в области $I^-(x)$. Так как λ пересекает $\mathscr{S}$ в некоторой точке x_1, то существует $y_1 \in \gamma \cap I^-(\mathscr{S})$. □

Следствие

Если $p \in \operatorname{int}(D(\mathscr{S}))$, то любая непродолжимая непространственноподобная кривая, проходящая через p, пересекает $I^-(\mathscr{S})$ и $I^+(\mathscr{S})$.

Имеем $\operatorname{int}(D(\mathscr{S})) = D(\mathscr{S}) - \{H^+(\mathscr{S}) \cup H^-(\mathscr{S})\}$. Если $p \in I^+(\mathscr{S})$ или $I^-(\mathscr{S})$, результат получается сразу. Если $p \in D^+(\mathscr{S}) - I^+(\mathscr{S})$,

то $p \in \mathscr{S} \subset D^-(\mathscr{S})$, и отсюда снова следует то же утверждение. □

Лемма 6.6.5
В $\operatorname{int}(D(\mathscr{S}))$ выполняется условие сильной причинности.

Допустим, что через $p \in \operatorname{int} D(\mathscr{S})$ проходит непространственноподобная кривая λ. Согласно предыдущему результату, должны существовать точки $q \in \lambda \cap I^-(\mathscr{S})$ и $r \in \lambda \cap I^+(\mathscr{S})$. Поскольку точка $r \in J^-(q)$, то она принадлежит также и $I^-(\mathscr{S})$, что противоречит ахрональности $\mathscr{S}$. Таким образом, в $\operatorname{int} D(\mathscr{S})$ выполняется условие причинности. Теперь допустим, что в p нарушается условие сильной причинности. Тогда, как и в лемме 6.4.6, должна существовать бесконечная последовательность направленных в будущее непространственноподобных кривых λ_n, которая сходится к проходящей через q непродолжимой изотропной геодезической γ. При этом существовали бы точки $q \in \gamma \cap I^-(\mathscr{S})$ и $r \in \gamma \cap I^+(\mathscr{S})$, и, следовательно, нашлась бы некоторая λ_n, которая пересекает $I^+(\mathscr{S})$, а значит, и $I^-(\mathscr{S})$. Это противоречит ахрональности $\mathscr{S}$. □

Доказательство предложения 6.6.3. Мы хотим показать, что $C(p, q)$ компактно, если $p, q \in \operatorname{int}(D(\mathscr{S}))$. Рассмотрим сначала тот случай, когда $p, q \in I^-(\mathscr{S})$, и предположим, что $p \in J^-(q)$. Пусть λ_n — бесконечная последовательность непространственноподобных кривых, соединяющих p и q. Согласно лемме 6.2.1, будет существовать направленная в будущее непространственноподобная предельная кривая с началом в p, которая непродолжима в $\mathscr{M} - q$. Она должна иметь конечную точку в будущем в q, так как иначе она пересекла бы $\mathscr{S}$, что невозможно, поскольку $q \in I^-(\mathscr{S})$. Рассмотрим теперь случай, когда $p \in J^-(\mathscr{S})$, $q \in J^+(\mathscr{S}) \cap J^+(p)$. Если конечной точкой предельной кривой λ является q, то она и есть искомая предельная точка в $C(p, q)$. Если q не является конечной точкой λ, то на ней должна быть точка $y \in I^+(\mathscr{S})$, поскольку λ непродолжима в $\mathscr{M} - q$. Пусть λ'_n — подпоследовательность, которая сходится к r для любой точки $r \in \lambda$ между p и q. Пусть $\hat{\lambda}$ — направленная в прошлое предельная кривая подпоследовательности λ'_n с началом в q. Если конечной точкой в прошлом кривой $\hat{\lambda}$ является точка p, то она и будет искомой предельной точкой в $C(p, q)$. Допустим, что p не является конечной точкой кривой $\hat{\lambda}$ и она не проходит через y. Тогда на ней должна быть некоторая точка $z \in I^-(\mathscr{S})$. Пусть λ''_n — подпоследовательность λ'_n, сходящаяся к r для любой точки $r \in \hat{\lambda}$ между q и z. Пусть $\mathscr{V}$ — открытая окрестность кривой $\hat{\lambda}$, не содержащая y. Тогда для достаточно больших n все $\lambda''_n \cap J^+(\mathscr{S})$ должны принадлежать $\mathscr{V}$. Однако это невозможно,

ибо y — предельная точка λ''_n. Поэтому должна существовать непространственноподобная кривая от p до q, которая будет предельной точкой последовательности λ_n в пространстве $C(p, q)$.

Случаи, когда $p, q \in I^-(\mathscr{S})$ и $p \in J^-(\mathscr{S})$, $q \in J^+(\mathscr{S})$, вместе с дуальными им случаями, исчерпывают все возможные комбинации. Следовательно, во всех этих случаях мы получаем непространственноподобную кривую от p до q, являющуюся предельной точкой последовательности λ_n в топологии пространства $C(p, q)$. □

Подобными же рассуждениями можно доказать

Предложение 6.6.6

Если $q \in \mathrm{int}(D(\mathscr{S}))$, то $J^+(\mathscr{S}) \cap J^-(q)$ или компактно, или пусто. □

Чтобы показать, что глобально гиперболична вся область $D(\mathscr{S})$, а не только ее внутренняя часть, необходимо наложить некоторые дополнительные условия.

Предложение 6.6.7

Если $\mathscr{S}$ — замкнутое ахрональное множество, такое, что в $J^+(\mathscr{S}) \cap J^-(\mathscr{S})$ одновременно удовлетворяются и условие сильной причинности, и одно из следующих условий:

1) непричинность (что имеет место, если и только если $\mathscr{S}$ непричинно);

2) компактность —

тогда $D(\mathscr{S})$ — глобально гиперболическое множество.

Допустим, что в какой-то точке $q \in D(\mathscr{S})$ условие сильной причинности не выполняется. Тогда в силу утверждения, аналогичного лемме 6.6.5, через q должна проходить непродолжимая изотропная геодезическая, в каждой точке которой нарушена сильная причинность. Но это невозможно, поскольку такая кривая пересекала бы $\mathscr{S}$. Следовательно, на $D(\mathscr{S})$ выполняется сильная причинность.

Если $p, q \in I^-(\mathscr{S})$, то выполняется условие предложения 6.6.3. Когда $p \in J^-(\mathscr{S})$, $q \in J^+(\mathscr{S})$, можно, как и при доказательстве предложения 6.6.3, провести из p направленную в будущее предельную кривую λ, а из q — направленную в прошлое предельную кривую $\hat{\lambda}$ и выбрать подпоследовательность λ''_n, которая сходится к r для каждой точки r на λ или $\hat{\lambda}$. В случае (1) λ пересекает $\mathscr{S}$ в единственной точке x. В любой окрестности x будут точки кривых λ''_n с достаточно большими n и, следовательно, ввиду ахрональности $\mathscr{S}$ в ней будет точка $x''_n \equiv \lambda'''_n \cap \mathscr{S}$. Отсюда, x''_n будет сходиться к x. Аналогично, x''_n будет сходиться к $\hat{x} \equiv \hat{\lambda} \cap \mathscr{S}$. Таким образом, $\hat{x} = x$ и можно соединить λ с $\hat{\lambda}$, что приводит к непространственноподобной предельной кривой в $C(p, q)$.

В случае (2) предположим, что q не является конечной точкой в будущем кривой λ. Тогда λ покинет область $I^-(\mathscr{S})$, поскольку она пересекает $\mathscr{S}$ и в силу предложения 6.4.7 должна выйти из компактного множества $J^+(\mathscr{S}) \cap J^-(\mathscr{S})$. Поэтому на λ можно найти точку x, которая не лежит в $I^-(\mathscr{S})$. Для каждого n выберем точку $x''_n \in \mathscr{S} \cap \lambda''_n$. Поскольку $\mathscr{S}$ компактно, найдутся некоторая точка $y \in \mathscr{S}$ и подпоследовательность λ'''_n, такие, что соответствующие точки x'''_n сходятся к y. Допустим, что y не лежит на λ. Тогда при достаточно больших n x'''_n будет лежать в будущем любой окрестности $\mathscr{U}$ точки x. Это означает, что $x \in \overline{I^-(\mathscr{S})}$. Однако последнее невозможно, поскольку x находится в $I^+(\mathscr{S})$, но не в компактном множестве $J^+(\mathscr{S}) \cap J^-(\mathscr{S})$. Поэтому λ пройдет через y. Аналогично, $\hat{\lambda}$ тоже пройдет через y. Итак, эти кривые можно соединить и получить предельную кривую. □

Из предложения 6.6.3 следует, что существование поверхности Коши для открытого множества $\mathscr{N}$ означает глобальную гиперболичность $\mathscr{N}$. Из следующего предложения видна справедливость обратного утверждения.

Предложение 6.6.8 (Герок [57])
Если открытое множество $\mathscr{N}$ глобально гиперболично, то оно, рассматриваемое как многообразие, гомеоморфно $R^1 \times \mathscr{S}$, где $\mathscr{S}$ — трехмерное многообразие и $\{a\} \times \mathscr{S}$ есть поверхность Коши в многообразии $\mathscr{N}$ для каждого $a \in R^1$.

Как и при доказательстве 6.4.9, введем на $\mathscr{N}$ меру μ так, чтобы полный объем $\mathscr{N}$ по этой мере был равен единице. Для $p \in \mathscr{N}$ введем функцию $f^+(p)$ — объем области $I^+(p, \mathscr{N})$ по мере μ. Ясно, что $f^+(p)$ — ограниченная функция на $\mathscr{N}$: она убывает вдоль любой направленной в будущее непространственноподобной кривой. Мы покажем, что глобальная гиперболичность означает непрерывность $f^+(p)$ на $\mathscr{N}$, так что нам не нужно, как при доказательстве предложения 6.4.9, «усреднять» объем будущего. Достаточно лишь доказать, что $f^+(p)$ непрерывна на любой непространственноподобной кривой λ.

Пусть $r \in \lambda$, и пусть x_n — бесконечная последовательность точек на λ, лежащих строго в прошлом относительно r. Пусть $\mathscr{F} = \bigcap_n I^+(x_n, \mathscr{N})$. Допустим, что $f^+(p)$ не полунепрерывна сверху в $r \in \lambda$. При этом должна существовать точка $q \in \mathscr{F} - I^+(r, \mathscr{N})$. Тогда $r \notin I^-(q, \mathscr{N})$; но каждая точка $x_n \in I^-(q, \mathscr{N})$ и, следовательно, $r \in \overline{I^-}(q, \mathscr{N})$, что невозможно, поскольку, согласно предложению 6.6.1, $I^-(q, \mathscr{N})$ замкнуто в $\mathscr{N}$. Подобным же образом доказывается полунепрерывность снизу.

По мере движения точки p в будущее вдоль непродолжимой непространственноподобной кривой λ на $\mathscr{N}$ значение $f^+(p)$ должно стремиться к нулю. Предположим, что имеется некоторая точка, которая лежит в будущем любой точки. Тогда направленная в будущее кривая λ войдет в компактное множество $J^+(r) \cap \cap J^-(q)$ и останется внутри его при любом $r \in \lambda$, что невозможно в силу предложения 6.4.7, так как на $\mathscr{N}$ выполняется условие сильной причинности.

Рассмотрим теперь определенную на $\mathscr{N}$ функцию $f(p) = f^-(p)/f^+(p)$. Любая поверхность $f = \text{const}$ будет непричинным множеством и, согласно предложению 6.3.1, вложенным в $\mathscr{N}$ трехмерным C^1-многообразием. Она будет также поверхностью Коши для $\mathscr{N}$, потому что вдоль любых непространственноподобных кривых f^- и f^+ стремятся к нулю в прошлом и в будущем соответственно. На $\mathscr{N}$ можно ввести времениподобное векторное поле $\mathbf{V}$ и задать непрерывное отображение β, которое переводит точки $\mathscr{N}$ вдоль интегральных кривых поля $\mathbf{V}$ на пересечения этих кривых с $\mathscr{S}(f = 1)$. Тогда $(\ln f(p), \beta(p))$ есть гомеоморфизм $\mathscr{N}$ на $R^1 \times \mathscr{S}$, который можно превратить в диффеоморфизм, если сгладить f, как при доказательстве предложения 6.4.9. □

Таким образом, если все пространство-время глобально гиперболично, оно обладает весьма скучной топологией.

6.7. Существование геодезических

Важность глобальной гиперболичности для гл. 8 связана со следующим результатом:

Предложение 6.7.1

Пусть p и q лежат на глобально гиперболическом множестве $\mathscr{N}$, причем $q \in J^+(p)$. Тогда p и q можно соединить непространственноподобной кривой, длина которой не меньше длины любой другой непространственноподобной кривой от p до q.

Мы дадим два доказательства этого утверждения: первое, принадлежащее Авезу [4] и Зейферту [156], основано на использовании компактности $C(p, q)$, а второе (применимое лишь при открытом $\mathscr{N}$) представляет собой метод построения самой этой геодезической.

Пространство $C(p, q)$ содержит плотное подмножество $C'(p, q)$, состоящее из всех времениподобных C^1-кривых от p до q. Длина одной такой кривой определяется (ср. с разд. 4.5) как

$$L[\lambda] = \int_p^q \left(-g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right)\right)^{1/2} \mathrm{d}t,$$

где t есть C^1-параметр на λ. Функция L не является непрерывной на $C'(p, q)$, поскольку любая окрестность λ содержит зигзагообразную кусочно почти изотропную кривую произвольно малой длины (рис. 46). Эта потеря непрерывности возникает от того, что мы пользуемся C^0-топологией, согласно которой две кривые близки, если в $\mathscr{M}$ близки их точки, но не требуется близости их касательных векторов. Можно было бы наделить

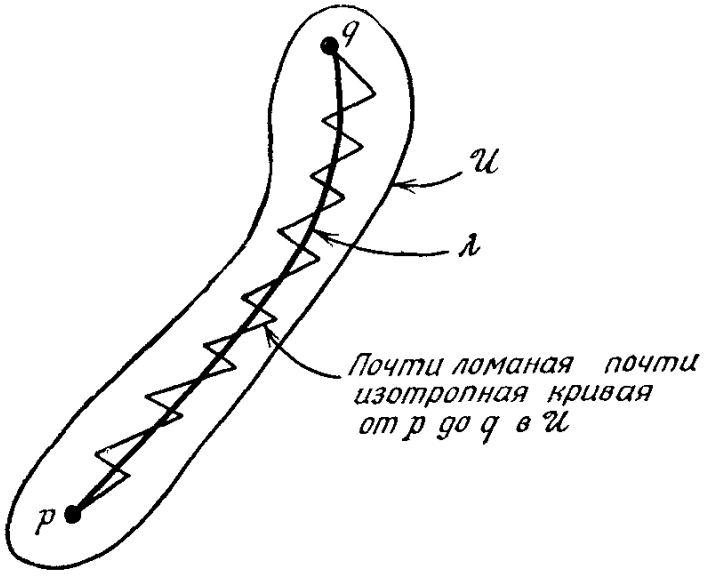


Рис. 46. $\mathscr{U}$ — открытая окрестность времениподобной кривой λ от p до q. В $\mathscr{U}$ существуют времениподобные кривые от p до q, которые приближенно являются ломаными изотропными кривыми и имеют сколь угодно малую длину.

$C'(p, q)$ C^1-топологией и таким образом придать L непрерывность, но мы не делаем этого, ибо $C'(p, q)$ некомпактно; компактное пространство получается только после присоединения всех непрерывных непространственноподобных кривых. Вместо этого мы пользуемся C^0-топологией и распространим определение L на $C(p, q)$.

Из-за сигнатуры метрики увеличение извилистости времениподобной кривой сокращает ее длину. Таким образом, L не полунепрерывна снизу. Однако имеет место

Лемма 6.7.2
L полунепрерывна сверху в C^0-топологии на $C'(p, q)$.

Рассмотрим времениподобную C^1-кривую $\lambda(t)$ от p до q, причем в качестве параметра выберем длину дуги от p. В достаточно малой окрестности $\mathscr{U}$ кривой λ можно найти функцию f, которая равна t и такова, что поверхности $\{f=\text{const}\}$ пространственноподобны и ортогональны к $\partial/\partial t$ [т. е. $g^{ab}f_{;b}|_\lambda = (\partial/\partial t)^a$]. Один способ задания такой функции состоит в построении пространственноподобных геодезических, ортогональных к λ. Для

достаточно малой окрестности $\mathcal{U}$ кривой λ. они дают однозначное отображение $\mathcal{N}$ на λ., и значение f в некоторой точке p окрестности $\mathcal{U}$ можно определить как значение t в той точке λ., на которую отображается p. Любую кривую μ в $\mathcal{N}$ можно параметризовать функцией f. Касательный вектор $(\partial/\partial f)_\mu$ к μ можно представить в виде

$$\left(\left(\frac{\partial}{\partial f}\right)_\mu\right)^a = g^{ab} f_{;b} + k^a,$$

где $\mathbf{k}$ — пространственноподобный вектор, лежащий на поверхности $\{f = \text{const}\}$, т. е. $k^a f_{;a} = 0$. Тогда

$$g\left(\left(\frac{\partial}{\partial f}\right)_\mu, \left(\frac{\partial}{\partial f}\right)_\mu\right) = g^{ab} f_{;a} f_{;b} + g_{ab} k^a k^b \geqslant g^{ab} f_{;a} f_{;b}.$$

Однако на λ $g^{ab} f_{;a} f_{;b} = -1$. Отсюда, при данном $\varepsilon > 0$ можно выбрать настолько малую окрестность $\mathcal{U}' \subset \mathcal{U}$, что в $\mathcal{U}'$ $g^{ab} f_{;a} f_{;b} > -1 + \varepsilon$. Поэтому для любой кривой μ в $\mathcal{U}'$

$$L[\mu] \leqslant (1 + \varepsilon)^{1/2} L[\lambda]. \qquad \square$$

Теперь мы определим длину непрерывной непространственноподобной кривой λ от p до q следующим образом. Пусть $\mathcal{U}$ — окрестность λ в $\mathcal{M}$ и пусть $l(\mathcal{U})$ — наименьшая верхняя грань длин времениподобных кривых в $\mathcal{U}$ от p до q. Тогда мы определим $L[\lambda]$ как наибольшую нижнюю грань $l(\mathcal{U})$ для всех окрестностей $\mathcal{U}$ кривой λ в $\mathcal{M}$. Это определение длины применимо ко всем кривым от p до q, в каждой окрестности которых существует времениподобная C^1-кривая, т. е. оно применимо ко всем точкам в $C(p, q)$, которые принадлежат замыканию $C'(p, q)$. Как мы знаем (разд. 4.5), непространственноподобная кривая от p до q, если она не является изотропной геодезической кривой без изломов, может быть деформирована во времениподобную кусочно C^1-кривую от p до q, а изломы этой кривой можно округлить так, чтобы получить времениподобную C^1-кривую от p до q. Следовательно, точки в $C(p, q) - \overline{C'(p, q)}$ — это изотропные геодезические без изломов (не содержащие сопряженных точек), и мы положим, что их длина равна нулю.

Такое определение L превращает ее в полунепрерывную сверху функцию на компактном пространстве $\overline{C'(p, q)}$. [На самом деле, пока непрерывная непространственноподобная кривая удовлетворяет локальному условию Липшица, она дифференцируема почти всюду. Поэтому длину все еще можно было бы определить как

$$\int \left(-g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right)\right)^{1/2} dt,$$

и это не противоречило бы введенному выше определению.] Если $\overline{C'(p, q)} = \varnothing$, но $C(p, q)$ — непустое множество, то p и q соединяются изотропной геодезической без изломов; кроме нее между p и q не существует иных непространственноподобных кривых. Если $\overline{C'(p, q)} \neq \varnothing$, это множество содержит некоторую точку, в которой L достигает своего максимального значения, т. е. найдется непространственноподобная кривая γ от p до q, длина которой больше или равна длине любой другой такой кривой. Согласно предложению 4.5.3, γ должна быть геодезической кривой, поскольку иначе можно было бы найти точки $x, y \in \gamma$, которые лежали бы в нормальной выпуклой координатной окрестности и могли бы быть соединены геодезическим сегментом большей длины, чем часть γ между x и y. □

Для другого, конструктивного доказательства мы введем функцию $\mathrm{d}(p, q)$ от $p, q \in \mathscr{M}$, равную нулю, если $q \notin J^+(p)$, а в противном случае равную наименьшей верхней грани длин, кусочно непространственноподобных кривых от p до q, направленных в будущее. [Отметим, что $\mathrm{d}(p, q)$ может быть бесконечной.] Для множеств $\mathscr{S}$ и $\mathscr{U}$ введем $\mathrm{d}(\mathscr{S}, \mathscr{U})$, равную наименьшей верхней грани $\mathrm{d}(p, q)$ при $p \in \mathscr{S}$, $q \in \mathscr{U}$.

Допустим, что $q \in I^+(p)$ и $\mathrm{d}(p, q)$ конечно. Тогда для любого $\delta > 0$ можно подобрать времениподобную кривую λ от p до q длиной $\mathrm{d}(p, q) - \frac{1}{2}\delta$ и такую окрестность $\mathscr{U}$ точки q, что λ можно деформировать во времениподобную кривую длины $\mathrm{d}(p, q) - \delta$, соединяющую p с любой точкой $r \in \mathscr{U}$. Следовательно, функция $\mathrm{d}(p, q)$ там, где она конечна, полунепрерывна снизу. В общем случае $\mathrm{d}(p, q)$ не полунепрерывна сверху, но имеется

Лемма 6.7.3

Функция $\mathrm{d}(p, q)$ конечна и непрерывна по p и q, когда p и q принадлежат глобально гиперболическому множеству $\mathscr{N}$.

Сначала мы докажем, что $\mathrm{d}(p, q)$ конечна. Поскольку на компактном множестве $J^+(p) \cap J^-(q)$ выполняется условие сильной причинности, его можно покрыть конечным числом таких локально-причинных множеств, каждое из которых не содержит непространственноподобные кривые длиннее некоторого ε. Поскольку любая непространственноподобная кривая от p до q может войти в каждую окрестность не более одного раза, она должна иметь конечную длину.

Предположим теперь, что $p, q \in \mathscr{N}$; существует такое постоянное $\delta > 0$, что каждая окрестность q содержит точку $r \in \mathscr{N}$, для которой

$$\mathrm{d}(p, r) > \mathrm{d}(p, q) + \delta.$$

Пусть x_n — бесконечная последовательность точек в $\mathscr{N}$, сходящаяся к q, для которой $\mathrm{d}(p, x_n) > \mathrm{d}(p, q) + \delta$. Тогда из каждой x_n до p можно провести непространственноподобную кривую λ_n, длина которой больше, чем $\mathrm{d}(p, q) + \delta$. По лемме 6.2.1 существует направленная в прошлое непространственноподобная кривая λ, которая проходит через q и является предельной кривой последовательности λ_n. Пусть $\mathscr{U}$ — локально-причинная окрестность точки q. Тогда λ не может пересечь $I^-(q) \cap \mathscr{U}$, ибо в этом случае одну из кривых λ_n можно было бы деформировать в непространственноподобную кривую от p до q с длиной больше $\mathrm{d}(p, q)$. Поэтому $\lambda \cap \mathscr{U}$ обязательно будет изотропной геодезической с началом в q, и в каждой точке $x \in \lambda \cap \mathscr{U}$ функция $\mathrm{d}(p, x)$ будет иметь разрыв больше, чем δ. Повторением этих же аргументов можно показать, что λ — изотропная геодезическая и что в каждой точке $x \in \lambda$ функция $\mathrm{d}(p, x)$ имеет разрыв больше, чем δ. Отсюда видно, что p не является конечной точкой кривой λ, так как по предложению 4.5.3 $\mathrm{d}(p, x)$ непрерывна в локально-причинной окрестности точки p. С другой стороны, λ должна быть непродолжимой в $\mathscr{M} - p$, и поэтому, если бы p не была его конечной точкой, то она должна была бы, согласно предложению 6.4.7, выйти из компактного множества $J^+(p) \cap \cap J^-(q)$. Это доказывает, что $\mathrm{d}(p, q)$ полунепрерывна сверху на $\mathscr{N}$. □

Используя функцию расстояния, легко построить геодезическую максимальной длины от p до q в случае открытого $\mathscr{N}$. Пусть $\mathscr{U} \subset \mathscr{N}$ — локально-причинная окрестность p, не содержащая q, и пусть $x \in J^+(p) \cap J^-(q)$ такова, что функция $\mathrm{d}(p, r) + \mathrm{d}(r, q)$, $r \in \mathscr{U}$ максимизируется при $r = x$. Проведем из p через точку x направленную в будущее геодезическую γ. Для всех точек r на γ между p и x будет выполняться соотношение $\mathrm{d}(p, r) + \mathrm{d}(r, q) = \mathrm{d}(p, q)$. Допустим, что имеется точка $y \in \in J^-(q) - q$, являющаяся последней точкой на γ, в которой это соотношение выполняется. Пусть $\mathscr{V} \subset \mathscr{N}$ — локально-причинная окрестность точки y, не содержащая q, и пусть точка $z \in J^+(y) \cap \cap J^-(q) \cap \mathscr{V}$ такова, что $\mathrm{d}(y, r) + \mathrm{d}(r, q)$, $r \in \mathscr{V}$ достигает своего максимального значения $\mathrm{d}(y, q)$ при $r = z$. Если $z \notin \gamma$, то

$$\mathrm{d}(p, z) > \mathrm{d}(p, y) + \mathrm{d}(y, z) \quad \text{и} \quad \mathrm{d}(p, z) + \mathrm{d}(z, q) > \mathrm{d}(p, q),$$

что невозможно. Отсюда видно, что соотношение

$$\mathrm{d}(p, r) + \mathrm{d}(r, q) = \mathrm{d}(p, q)$$

должно выполняться для всех $r \in \gamma \cap J^-(q)$. Ввиду компактности $J^+(p) \cap J^-(q)$ кривая γ должна покинуть $J^-(q)$ в некоторой точке y. Допустим, $y \neq q$; тогда y должна лежать на направленной в прошлое изотропной геодезической с началом в q. Соединив γ

с λ, мы получили бы непространственноподобную кривую от p до q, варьируя которую, можно было бы получить кривую более длинную, чем $\mathrm{d}(p, q)$; но это невозможно. Следовательно, γ — геодезическая кривая от p до q длины $\mathrm{d}(p, q)$. □

Следствие

Если $\mathscr{S}$ — частичная C^2-поверхность Коши, то к каждой точке $q \in D^+(\mathscr{S})$ можно провести направленную в будущее времениподобную геодезическую кривую длины $\mathrm{d}(\mathscr{S}, q)$, ортогональную $\mathscr{S}$ и не содержащую между $\mathscr{S}$ и q сопряженную к $\mathscr{S}$ точку.

Согласно предложению 6.5.2, $H^+(\mathscr{S})$ и $H^-(\mathscr{S})$ не пересекают $\mathscr{S}$ и, следовательно, не принадлежат $D(\mathscr{S})$.

Поэтому и в силу предложения 6.6.3 $D(\mathscr{S}) = \operatorname{int} D(\mathscr{S})$ — глобально гиперболическое множество. По предложению 6.6.6 $\mathscr{S} \cap J^-(q)$ компактно, и, следовательно, $\mathrm{d}(p, q)$, $p \in \mathscr{S}$ достигнет своего максимального значения $\mathrm{d}(\mathscr{S}, q)$ в некоторой точке $r \in \mathscr{S}$. При этом между r и q будет существовать геодезическая кривая γ длины $\mathrm{d}(\mathscr{S}, q)$, которая, согласно лемме 4.5.5 и предложению 4.5.9, должна быть ортогональна к $\mathscr{S}$ и не должна иметь между $\mathscr{S}$ и q точку, сопряженную к $\mathscr{S}$. □

6.8. Причинная граница пространства-времени

В этом разделе мы изложим в общих чертах метод присоединения границы к пространству-времени, предложенный Героком, Кронхеймером и Пенроузом [60]. Это построение зависит только от причинной структуры $(\mathscr{M}, \mathbf{g})$, т. е. в нем не различаются граничные точки, находящиеся на конечном расстоянии и на бесконечности. В разд. 8.3 мы дадим описание другого построения, при котором присоединяется граница, представляющая лишь сингулярные точки. К сожалению, похоже, что между этими построениями нет явной связи.

Будем предполагать, что $(\mathscr{M}, \mathbf{g})$ удовлетворяет условию причинности. Тогда любая точка p в $(\mathscr{M}, \mathbf{g})$ однозначно определяется ее хронологическим будущим $I^+(p)$ или прошлым $I^-(p)$, т. е.

$$I^-(p) = I^-(q) \Leftrightarrow I^+(p) = I^+(q) \Leftrightarrow p = q.$$

Хронологическое прошлое $\mathscr{W} \equiv I^-(p)$ любой точки $p \in \mathscr{M}$ обладает следующими свойствами:

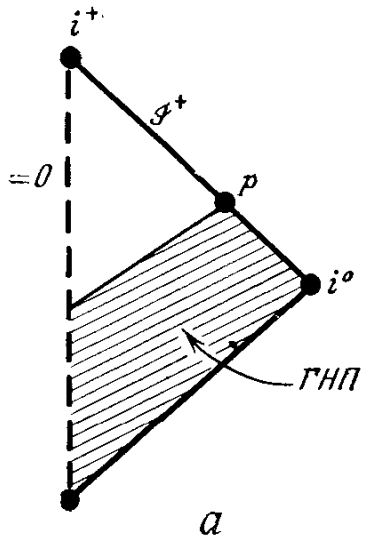
1) $\mathscr{W}$ — открытое множество;

2) $\mathscr{W}$ — множество прошлого, т. е. $I^-(\mathscr{W}) \subset \mathscr{W}$;

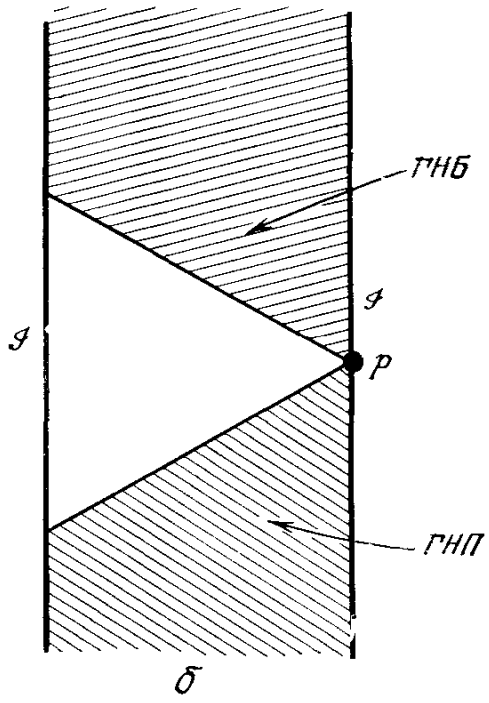
3) $\mathscr{W}$ не может быть представлено в виде объединения двух собственных подмножеств, обладающих свойствами (1) и (2).

Множество со свойствами (1)—(3) мы будем называть *неразложимым множеством прошлого*, сокращенно НП. (Определение, данное в [60], не включает свойства (1), тем не менее оно эквивалентно нашему определению, поскольку в [60] под «мно-

жеством прошлого» подразумевается множество, которое равно своему хронологическому прошлому, а не только содержит его.)



а



б

Рис. 47. Диаграммы Пенроуза пространства Минковского и пространства де Ситтера 2-го рода (ср. рис. 15 и 20), на которых изображены: ГНП, представляющее на $\mathscr{I}^+$ точку p пространства Минковского (*а*), и ГНП и ГНБ, представляющее на $\mathscr{I}$ точку p пространства де Ситтера 2-го рода (*б*).

Аналогично определяется *неразложимое множество будущего* (НБ).

Все НП делятся на два класса: *собственные НП* (*СНП*), которые являются прошлым каких-либо точек в $\mathscr{M}$, и *граничные НП*

(*ГНП*), которые не являются прошлым каких-либо точек в $\mathscr{M}$. Идея состоит в том, чтобы рассматривать эти ГНП и соответствующим образом определенные ГНБ в качестве представителей точек, лежащих на причинной границе (*с-граница*) пространства $(\mathscr{M}, g)$. Например, в пространстве Минковского заштрихованную область на рис. 47, *а* можно рассматривать в качестве представителя точки p на $\mathscr{I}^+$. Заметим, что в этом примере многообразие $\mathscr{M}$ в целом само является и ГНП, и ГНБ, причем эти НП и НБ можно считать представителями точек i^+ и i^- соответственно. Всем точкам конформной границы пространства Минковского, за исключением i^0, могут быть поставлены в соответствие ГНП или ГНБ. В некоторых случаях, таких, как пространство де Ситтера 2-го рода, где конформная граница времениподобна, представителями одной и той же точки границы могут быть и ГНБ и ГНП (рис. 47, *б*).

ГНП можно охарактеризовать и как прошлое непродолжимых в будущее времениподобных кривых. Это означает, что прошлое $I^-(\gamma)$ непродолжимой в будущее кривой γ можно рассматривать в качестве представителя конечной точки γ в будущем на c-границе. Другая кривая γ' в будущем будет иметь ту же конечную точку, если и только если $I^-(\gamma) = I^-(\gamma')$.

Предложение 6.8.1 (Герок, Кронхеймер и Пенроуз)
Множество $\mathscr{W}$ есть ГНП тогда и только тогда, когда существует непродолжимая в будущее времениподобная кривая γ, для которой $I^-(\gamma) = \mathscr{W}$.

Предположим сначала, что имеется кривая γ, для которой $I^-(\gamma) = \mathscr{W}$. Пусть $\mathscr{W} = \mathscr{U} \cup \mathscr{V}$, где $\mathscr{U}$ и $\mathscr{V}$ — некоторые открытые множества прошлого. Нам нужно показать, что или $\mathscr{U} \subset \mathscr{V}$, или $\mathscr{V} \subset \mathscr{U}$. Предположим противное: $\mathscr{U}$ не содержится в $\mathscr{V}$ и $\mathscr{V}$ не содержится в $\mathscr{U}$. Тогда можно найти точку $q \in \mathscr{U} - \mathscr{V}$ и точку $r \in \mathscr{V} - \mathscr{U}$. Поскольку $q, r \in I^-(\gamma)$, должны найтись такие точки q', $r' \in \gamma$, что $q \in I^-(q')$ и $r \in I^-(r')$. Но независимо от того, какое из множеств, $\mathscr{U}$ или $\mathscr{V}$, содержит более позднюю из точек q', r', она будет содержать также и q, и r, что противоречит первоначальному определению q и r.

Обратно, допустим, что $\mathscr{W}$ есть ГНП. Тогда можно построить времениподобную кривую γ, для которой $I^-(\gamma) = \mathscr{W}$. При этом, если p — какая-либо точка множества $\mathscr{W}$, то

$$\mathscr{W} = I^-\left(\mathscr{W} \cap I^+(p)\right) \cup I^-\left(\mathscr{W} - I^+(p)\right).$$

Однако $\mathscr{W}$ неразложимо, и потому или $\mathscr{W} = I^-(\mathscr{W} \cap I^+(p))$, или $\mathscr{W} = I^-(\mathscr{W} - I^+(p))$. Точка p не содержится в $I^-(\mathscr{W} - I^+(p))$, так что вторая возможность исключается.

Этот вывод можно переформулировать следующим образом: в $\mathscr{W}$ найдется точка, находящаяся в будущем относительно

любой данной пары точек из $\mathscr{W}$. Выберем теперь счетное плотное семейство точек $p_n \in \mathscr{W}$. Возьмем некоторую точку $q_0 \in \mathscr{W}$, находящуюся в будущем p_0. Поскольку q_0 и p_1 принадлежат $\mathscr{W}$, в будущем обеих этих точек можно подобрать точку $q_2 \in \mathscr{W}$ и т. д. Так как каждая точка q_n, полученная таким способом, лежит в прошлом последующей точки, в $\mathscr{W}$ можно построить времениподобную кривую, проходящую через все точки последовательности.

Для каждой точки $p \in \mathscr{W}$ $\mathscr{W} \cap I^+(p)$ — открытое и непустое множество и оно должно содержать по меньшей мере одну из p_n, поскольку эти точки плотны. Но для каждого k точка p_k лежит в прошлом q_k, поэтому сама p лежит в прошлом γ. Отсюда ясно, что каждая точка $\mathscr{W}$ лежит в прошлом γ, и, поскольку γ принадлежит открытому прошлому множеству $\mathscr{W}$, должно иметь место равенство $\mathscr{W} = I^-(\gamma)$. □

Обозначим через $\hat{\mathscr{M}}$ множество всех НП пространства $(\mathscr{M}, \mathbf{g})$. Тогда $\hat{\mathscr{M}}$ представляет все точки $\mathscr{M}$ вместе с точками будущей c-границы; аналогично, $\check{\mathscr{M}}$, множество всех НБ пространства $(\mathscr{M}, \mathbf{g})$, представляет $\mathscr{M}$ вместе с прошлой c-границей. Причинные отношения I, J и E можно распространить на $\hat{\mathscr{M}}$ и $\check{\mathscr{M}}$ следующим образом. О множествах $\mathscr{U}, \mathscr{V} \subset \hat{\mathscr{M}}$ будем говорить, что

$\mathscr{U} \in J^-(\mathscr{V}, \hat{\mathscr{M}})$ при $\mathscr{U} \subset \mathscr{V}$,

$\mathscr{U} \in I^-(\mathscr{V}, \hat{\mathscr{M}})$ при $\mathscr{U} \subset I^-(q)$ для некоторой точки $q \in \mathscr{V}$,

$\mathscr{U} \in E^-(\mathscr{V}, \hat{\mathscr{M}})$ при $\mathscr{U} \in J^-(\mathscr{V}, \hat{\mathscr{M}})$, но $\mathscr{U} \notin I^-(\mathscr{V}, \hat{\mathscr{M}})$.

С учетом этих отношений НП-пространство $\hat{\mathscr{M}}$ является причинным пространством [94]. Существует естественное инъективное отображение I^-: $\mathscr{M} \to \hat{\mathscr{M}}$, которое переводит точку $p \in \mathscr{M}$ в $I^-(p) \in \hat{\mathscr{M}}$. Это отображение является изоморфизмом причинного отношения J^-, поскольку $p \in J^-(q)$ тогда и только тогда, когда $I^-(p) \in J^-(I^-(q), \hat{\mathscr{M}})$. Причинное отношение I^- сохраняется только в одну сторону, т. е. $p \in I^-(q) \not\Leftrightarrow I^-(p) \in I^-(I^-(q), \mathscr{M})$. Подобным же образом можно определить причинные отношения на $\check{\mathscr{M}}$.

Теперь наша цель состоит в том, чтобы как-то составить из $\hat{\mathscr{M}}$ и $\check{\mathscr{M}}$ пространство $\mathscr{M}^*$, которое имело бы вид $\mathscr{M} \cup \Delta$, где Δ можно было бы назвать границей пространства $(\mathscr{M}, \mathbf{g})$. Для этого нам нужен метод отождествления соответствующих НП и НБ. Начнем с построения пространства $\mathscr{M}^\#$, представляющего собой объединение $\hat{\mathscr{M}}$ и $\check{\mathscr{M}}$, в котором каждое СНБ отождествлено с соответствующим СНП. Другими словами, $\mathscr{M}^\#$ соответствует точкам $\mathscr{M}$ вместе со всеми ГНП и ГНБ. Однако, как по-

казывает пример пространства де Ситтера 2-го рода, необходимо еще и отождествление некоторых ГНП с определенными ГНБ. Один способ такого построения состоит в наделении $\mathscr{M}^{\#}$ топологией и в последующем отождествлении некоторых точек $\mathscr{M}^{\#}$ таким образом, чтобы эта топология стала хаусдорфовой.

Как было упомянуто в разд. 6.4, базисом такой топологии в топологическом пространстве $\mathscr{M}$ могут служить множества вида $I^+(p) \cap I^-(q)$. К сожалению, подобный метод не может быть применен для задания базиса топологии $\mathscr{M}^{\#}$, поскольку в $\mathscr{M}^{\#}$ могут быть точки, которые не принадлежат хронологическому прошлому какой-либо точки $\mathscr{M}^{\#}$. Однако топологию $\mathscr{M}$ можно получить также исходя из подбазиса, который состоит из множеств вида $I^+(p)$, $I^-(p)$, $\mathscr{M} - \overline{I^+(p)}$ и $\mathscr{M} - \overline{I^-(p)}$. Следуя этой аналогии, Герок, Кронхеймер и Пенроуз показали, как можно задать топологию на $\mathscr{M}^{\#}$. Для НБ $\mathscr{A} \in \check{\mathscr{M}}$ определяем множества

$$\mathscr{A}^{\text{int}} \equiv \{\mathscr{V}: \mathscr{V} \in \hat{\mathscr{M}} \text{ и } \mathscr{V} \cap \mathscr{A} \neq \varnothing\},$$

и

$$\mathscr{A}^{\text{ext}} \equiv \{\mathscr{V}: \mathscr{V} \in \hat{\mathscr{M}} \text{ и } \mathscr{V} = I^-(\mathscr{W}) \Rightarrow I^+(\mathscr{W}) \not\subset \mathscr{A}\};$$

для НП $\mathscr{B} \in \hat{\mathscr{M}}$ множества $\mathscr{B}^{\text{int}}$ и $\mathscr{B}^{\text{exp}}$ определяются аналогично. Открытые множества пространства $\mathscr{M}^{\#}$ определяются тогда как объединения и конечные пересечения множеств вида $\mathscr{A}^{\text{int}}$, $\mathscr{A}^{\text{ext}}$, $\mathscr{B}^{\text{int}}$ и $\mathscr{B}^{\text{ext}}$. Множества $\mathscr{A}^{\text{int}}$ и $\mathscr{B}^{\text{int}}$ являются аналогами множеств $I^+(p)$ и $I^-(q)$ в $\mathscr{M}^{\#}$. Если, в частности, $\mathscr{A} = I^+(p)$ и $\mathscr{V} = I^-(q)$, то $\mathscr{V} \in \mathscr{A}^{\text{int}}$ тогда и только тогда, когда $q \in I^+(p)$. Эти определения позволяют, однако, включить ГНП в $\mathscr{A}^{\text{int}}$. Множества $\mathscr{A}^{\text{ext}}$ и $\mathscr{B}^{\text{ext}}$ являются аналогами $\mathscr{M} - \overline{I^+(p)}$ и $\mathscr{M} - \overline{I^-(q)}$.

Окончательно мы получаем пространство $\mathscr{M}^{\#}$, отождествив в $\mathscr{M}^{\#}$ наименьшее число точек, необходимое для превращения $\mathscr{M}^{\#}$ в хаусдорфово пространство. Точнее $\mathscr{M}^*$ — фактор-пространство $\mathscr{M}^{\#}/R_h$, где R_h — пересечение всех отношений эквивалентности $R \subset \mathscr{M}^{\#} \times \mathscr{M}^{\#}$, для которых $\mathscr{M}^{\#}/R$ хаусдорфово. Пространство $\mathscr{M}^*$ обладает топологией, индуцированной из $\mathscr{M}^{\#}$ и согласующейся с топологией многообразия $\mathscr{M}$ на подмножестве $\mathscr{M} \subset \mathscr{M}^*$. Вообще говоря, дифференцируемую структуру многообразия $\mathscr{M}$ нельзя распространить на c-границу Δ, хотя можно это сделать для части Δ в частном случае, который будет описан в следующем разделе.

6.9. Асимптотически простые пространства

При изучении таких конечных физических систем, как звезды, приходится рассматривать асимптотически плоские пространства, т. е. пространства, метрика которых стремится к метрике пространства Минковского на больших расстояниях от

рассматриваемой системы. Примерами пространств, имеющих асимптотически плоские области, служат решения Шварцшильда, Райсснера — Нордстрема и Керра. Как мы видели в гл. 5, конформная структура изотропной бесконечности этих пространств и пространства Минковского одинакова, что дало Пенроузу [123, 125, 128] основание использовать это обстоятельство в качестве своего рода определения асимптотически плоского пространства. Хотя Пенроуз не накладывал требования сильной причинности, мы будем рассматривать только сильно причинные пространства. Такое ограничение упрощает дело и не приводит к потере общности в тех ситуациях, которые мы хотим исследовать.

Ориентируемое во времени пространственно-ориентируемое сильно причинное пространство $(\mathscr{M}, \mathbf{g})$ назовем *асимптотически простым*, если существует пространство $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$ и вложение $\theta: \mathscr{M} \to \tilde{\mathscr{M}}$, при котором $\mathscr{M}$ становится многообразием с гладким краем $\partial\mathscr{M}$ в $\tilde{\mathscr{M}}$ и выполняются следующие условия:

1) на $\tilde{\mathscr{M}}$ существует гладкая (скажем, по меньшей мере класса C^3) функция Ω, такая, что на $\theta(\mathscr{M})$ функция Ω положительна и $\Omega^2\mathbf{g} = \theta_*(\tilde{\mathbf{g}})$ [т. е. $\tilde{\mathbf{g}}$ конформна $\mathbf{g}$ на $\theta(\mathscr{M})$];

2) на $\partial\mathscr{M}$ $\Omega = 0$ и $d\Omega \neq 0$;

3) каждая изотропная геодезическая в $\mathscr{M}$ имеет две конечные точки на $\partial\mathscr{M}$;

Мы будем обозначать $\mathscr{M} \cup \partial\mathscr{M} \equiv \bar{\mathscr{M}}$.

На самом деле приведенное определение — несколько более общее, чем нам необходимо: ему удовлетворяют и космологические модели, например пространство де Ситтера 1-го рода. Чтобы сузить его до асимптотически плоских пространств, мы будем говорить, что пространство $(\mathscr{M}, \mathbf{g})$ *асимптотически простое и пустое*, если в дополнение к условиям (1)—(3) оно удовлетворяет еще и условию

4) $R_{ab} = 0$ на открытой окрестности $\partial\mathscr{M}$ в $\bar{\mathscr{M}}$. (Это условие можно видоизменить так, чтобы вблизи $\partial\mathscr{M}$ разрешалось присутствие электромагнитного излучения.)

Край $\partial\mathscr{M}$ можно считать находящимся на бесконечности в том смысле, что в метрике $\mathbf{g}$ любой аффинный параметр на изотропной геодезической в $\mathscr{M}$ достигает неограниченно большого значения вблизи $\partial\mathscr{M}$. Это происходит потому, что аффинный параметр v в метрике $\mathbf{g}$ связан с аффинным параметром $\tilde{v}$ в метрике $\tilde{\mathbf{g}}$ соотношением $dv/d\tilde{v} = \Omega^{-2}$. Поскольку на $\partial\mathscr{M}$ $\Omega = 0, \int dv$ расходится.

Из условий (2) и (4) следует, что край $\partial\mathscr{M}$ является изотропной гиперповерхностью. Действительно, тензор Риччи $\tilde{R}_{ab}$ для метрики $\tilde{g}_{ab}$ связан с R_{ab} для g_{ab} равенством

$$\tilde{R}_a{}^b = \Omega^{-2} R_a{}^b - 2\Omega^{-1}(\Omega)_{|ac}\,\tilde{g}^{bc} + \{-\Omega^{-1}\Omega_{|cd} + 3\Omega^{-2}\Omega_{|c}\Omega_{|d}\}\,\tilde{g}^{cd}\delta_a{}^b,$$

где $|$ означает ковариантное дифференцирование относительно $\tilde{g}_{ab}$. Отсюда

$$\tilde{R} = \Omega^{-2}R - 6\Omega^{-1}\Omega_{|cd}\tilde{g}^{cd} + 3\Omega^{-1}\Omega_{|c}\Omega_{|d}\tilde{g}^{cd}.$$

Поскольку метрика $\tilde{g}_{ab}$ класса C^3, то на $\partial\mathscr{M}$, где $\Omega = 0$, $\tilde{R}$ принадлежит классу C^1. Это означает, что $\Omega_{|c}\Omega_{|d}\tilde{g}^{cd} = 0$. Но по условию (2) $\Omega_{|c} \neq 0$, т. е. $\Omega_{|c}\tilde{g}^{cd}$ — изотропный вектор, а поверхность $\partial\mathscr{M}$ $(\Omega = 0)$ — изотропная гиперповерхность.

В случае пространства Минковского $\partial\mathscr{M}$ состоит из двух изотропных поверхностей $\mathscr{I}^+$ и $\mathscr{I}^-$, каждая из которых имеет топологию $R^1 \times S^2$. (Заметим, что ей не принадлежат точки i^0, i^+, i^-, так как вблизи этих точек конформная граница не является гладким многообразием.) Мы покажем, что в действительности для любого асимптотически простого и пустого пространства $\partial\mathscr{M}$ имеет такую структуру.

Поскольку $\partial\mathscr{M}$ — изотропная поверхность, локально $\mathscr{M}$ лежит в ее прошлом или будущем. Из этого видно, что $\partial\mathscr{M}$ должна состоять из двух несвязных компонент $\mathscr{I}^+$ и $\mathscr{I}^-$, на которых расположены соответственно будущие и прошлые концы изотропных геодезических в $\mathscr{M}$. $\partial\mathscr{M}$ не может состоять более чем из двух компонент, ибо тогда нашлась бы некоторая точка $p \in \mathscr{M}$, из которой одни направленные в будущее изотропные геодезические шли бы к одной компоненте, а другие — к другой. Множества изотропных направлений в p, ведущие к каждой из компонент, были бы при этом открытыми, что невозможно ввиду связности множества изотропных направлений в будущее в точке.

Теперь мы установим одно важное свойство рассматриваемых пространств.

Лемма 6.9.1

Асимптотически простое и пустое пространство $(\mathscr{M}, \mathbf{g})$ причинно простое.

Пусть $\mathscr{W}$ — компактное множество $\mathscr{M}$. Нужно показать, что каждая изотропная геодезическая образующая поверхности $\dot{J}(\mathscr{W}, \bar{\mathscr{M}})$ имеет прошлый конец в $\mathscr{W}$. Допустим, что имеется такая изотропная геодезическая, которая не имеет конца в $\mathscr{W}$. Тогда у нее не может быть какого-либо конца в $\mathscr{M}$ и она должна пересечь $\mathscr{I}^-$, что невозможно.

Предложение 6.9.2

Асимптотически простое и пустое пространство $(\mathscr{M}, \mathbf{g})$ является глобально гиперболическим.

Доказательство аналогично доказательству предложения 6.6.7. На $\mathscr{M}$ вводится элемент объема так, чтобы полный объем $\mathscr{M}$ по этой мере был равен единице. Ввиду причинной простоты $(\mathscr{M}, \mathbf{g})$

функции $f^{\pm}(p)$, представляющие собой объемы областей $I^{\pm}(p)$, непрерывны на $\mathscr{M}$. Поскольку на $\mathscr{M}$ выполняется условие сильной причинности, $f^{+}(p)$ будет убывать вдоль каждой направленной в будущее непространственноподобной кривой. Пусть λ — непродолжимая в будущее времениподобная кривая. Допустим, что множество $\mathscr{F} = \bigcap\limits_{p \in \lambda} I^{+}(p)$ — не пустое. Тогда $\mathscr{F}$ должно быть будущим множеством и изотропные, порождающие границу $\mathscr{F}$ в $\mathscr{M}$, не могут иметь прошлых концов в $\mathscr{M}$. Таким образом, они должны пересечь $\mathscr{I}^{-}$, что снова ведет к противоречию. Отсюда видно, что $f^{+}(p)$ стремится к нулю, когда p устремляется в будущее по λ. Из этого следует, что каждая непродолжимая непространственноподобная кривая пересекает поверхность $\mathscr{H} \equiv \equiv \{p : f^{+}(p) = f^{-}(p)\}$, которая поэтому является поверхностью Коши для $\mathscr{M}$. □

Лемма 6.9.3

Пусть $\mathscr{W}$ — компактное множество асимптотического простого и пустого пространства $(\mathscr{M}, \mathbf{g})$. Тогда каждая из изотропных геодезических образующих поверхности $\mathscr{I}^{+}$ пересекает $\dot{J}^{+}(\mathscr{W}, \bar{\mathscr{M}})$ один раз; точка над J означает границу в $\bar{\mathscr{M}}$.

Пусть $p \in \lambda$, где λ — одна из изотропных геодезических образующих $\mathscr{I}^{+}$. Тогда прошлое множество (в $\mathscr{M}$) $J^{-}(p, \bar{\mathscr{M}}) \cap \mathscr{M}$ должно быть замкнуто в $\mathscr{M}$, ибо каждая из изотропных геодезических образующих границы этого множества, должна иметь будущую конечную точку в $\mathscr{I}^{+}$ в точке p. Поскольку на $\mathscr{M}$ выполняется условие сильной причинности, $\mathscr{M} - J^{-}(p, \bar{\mathscr{M}})$ будет непустым. Теперь предположим, что $\lambda \in J^{+}(\mathscr{W}, \bar{\mathscr{M}})$. Тогда прошлое множество $\bigcap\limits_{p \in \lambda} (J^{-}(p, \bar{\mathscr{M}}) \cap \mathscr{M})$ должно быть непустым. Однако это невозможно из-за того, что изотропные образующие границы этого множества пересекли бы при этом $\mathscr{I}^{+}$. Теперь допустим обратное, что λ не пересекает $J^{+}(\mathscr{W}, \bar{\mathscr{M}})$. Тогда $\mathscr{M} - \bigcup\limits_{p \in \lambda} (J^{-}(p, \bar{\mathscr{M}}) \cap \mathscr{M})$ будет непустым. Но это снова ведет к противоречию, потому что изотропные образующие границы прошлого множества $\bigcup\limits_{p \in \lambda} (J^{-}(p, \bar{\mathscr{M}}) \cap \mathscr{M})$ пересекли бы при этом $\mathscr{I}^{+}$. □

Следствие

Поверхность $\mathscr{I}^{+}$ топологически совпадает с $R^{1} \times \times (\dot{J}^{+}(\mathscr{W}, \bar{\mathscr{M}}) \cap \partial \mathscr{M})$.

Теперь мы покажем, что $\mathscr{I}^{\pm}$ и $\mathscr{M}$ обладают такой же топологией, как и в случае пространства Минковского.

Предложение 6.9.4 (*Герок* [59])
В асимптотически простом и пустом пространстве $(\mathscr{M}, \mathbf{g})$ поверхности $\mathscr{I}^+$ и $\mathscr{I}^-$ топологически эквивалентны $R^1 \times S^2$, а $\mathscr{M}$ топологически эквивалентно R^4.

Рассмотрим множество N всех изотропных геодезических в $\mathscr{M}$. Поскольку они все пересекают поверхность Коши $\mathscr{H}$, локальные координаты на N можно задать с помощью направлений этих геодезических на $\mathscr{H}$ и локальных координат из точек пересечения с $\mathscr{H}$. Этим мы превращаем N в расслоенное пространство направлений на $\mathscr{H}$ со слоем S^2. Однако каждая геодезическая пересекает также $\mathscr{I}^+$. Поэтому N является расслоенным пространством и над $\mathscr{I}^+$. В этом случае слоем является S^2 без одной точки, которая соответствует изотропной геодезической образующей $\mathscr{I}^+$, не принадлежащей $\mathscr{M}$. Другими словами, слоем является R^2. Следовательно, пространство N топологически эквивалентно $\mathscr{I}^+ \times R^2$. Однако поверхность $\mathscr{I}^+$ эквивалентна $R^1 \times (\dot{J}^+(\mathscr{W}, \bar{\mathscr{M}}) \cap \partial\mathscr{M})$. Это согласуется с тем, что $N \approx \mathscr{H} \times S^2$ лишь при условии, что $\mathscr{H} \approx R^3$ и $\mathscr{I}^+ \approx R^1 \times S^2$. □

Как показал Пенроуз [125], из этого результата следует, что тензор Вейля для метрики $\mathbf{g}$ равен нулю на $\mathscr{I}^+$ и $\mathscr{I}^-$. В качестве объяснения можно сказать, что компоненты тензора Вейля для метрики $\mathbf{g}$ вырождаются, т. е. вблизи $\mathscr{I}^+$ и $\mathscr{I}^-$ они убывают как различные степени аффинного параметра на изотропной геодезической. Далее, в работах [114, 122] были получены в виде интегралов по $\mathscr{I}^+$ законы сохранения для энергии-импульса, как бы измеренных с поверхности $\mathscr{I}^+$.

Изотропные поверхности $\mathscr{I}^+$ и $\mathscr{I}^-$ составляют почти всю c-границу Δ (см. определение в предыдущем разделе) пространства $(\mathscr{M}, \mathbf{g})$. Чтобы убедиться в этом, заметим, во-первых, что любая точка $p \in \mathscr{I}^+$ задает ГНП $I^-(p, \bar{\mathscr{M}}) \cap \mathscr{M}$. Допустим, что λ — непродолжимая в будущее кривая в $\mathscr{M}$. Если будущей конечной точкой λ является $p \in \mathscr{I}^+$, то ГНП $I^-(\lambda)$ совпадает с ГНП, которое определяется точкой p. Если у кривой λ нет конечной точки в будущем на $\mathscr{I}^+$, то $\mathscr{M} - I^-(\lambda)$ должно быть пустым, так как в противном случае изотропные геодезические образующие $I^-(\lambda)$ пересекали бы $\mathscr{I}^+$, что невозможно: λ не пересекает $\mathscr{I}^+$. Таким образом, на каждую точку из $\mathscr{I}^+$ приходится одно ГНП, и еще одно ГНП, связанное с i^+, есть само многообразие $\mathscr{M}$; других ГНП нет. Аналогично, множество всех ГНБ составлено теми ГНБ, которые соответствуют каждой точке $\mathscr{I}^-$ и одним ГНБ, соответствующим i^-, которое снова совпадает с самим $\mathscr{M}$.

Теперь требуется проверить, не отождествляем ли мы какие-нибудь ГНП или ГНБ, т. е. установить, что $\mathscr{M}^\#$ хаусдорфово. Ясно, что никакие два ГНП или ГНБ, соответствующие $\mathscr{I}^+$ или

$\mathscr{I}^-$, не являются неразделимыми по Хаусдорфу. Если $p \in \mathscr{I}^+$, то можно найти такую $q \in \mathscr{M}$, что $p \notin I^+(q, \bar{\mathscr{M}})$. Тогда $(I^+(q, \bar{\mathscr{M}}))^{\text{ext}}$ является окрестностью в $\mathscr{M}^{\#}$ для ГНП $I^-(p, \bar{\mathscr{M}}) \cap \mathscr{M}$, а $(I^+(q, \bar{\mathscr{M}}))^{\text{int}}$ будет непересекающейся окрестностью ГНП, соответствующего i^+. Таким образом, i^+ отделимо по Хаусдорфу от любой точки поверхности $\mathscr{I}^+$. Аналогично, i^- отделимо по Хаусдорфу от любой точки $\mathscr{I}^-$. Таким образом, c-граница любого асимптотически простого и пустого пространства $(\mathscr{M}, \mathbf{g})$ такая же, как у пространства Минковского, т. е. состоит из $\mathscr{I}^+$, $\mathscr{I}^-$ и двух точек i^+, i^-.

К числу асимптотически простых и пустых пространств относятся пространство Минковского и асимптотически плоские пространства, в которых есть только ограниченные объекты, например звезды, не испытывающие гравитационного коллапса. Однако решения Шварцшильда, Райсснера — Нордстрема или Керра не входят в этот класс, потому что в них имеются изотропные геодезические, у которых конечные точки не лежат на $\mathscr{I}^+$ или $\mathscr{I}^-$. Тем не менее у этих пространств имеются асимптотически плоские области, подобные тем, которые существуют у асимптотически плоских и простых пространств. Из этого видно, что следует ввести еще понятие *слабо асимптотически плоского и пустого пространства* $(\mathscr{M}, \mathbf{g})$ как пространства, для которого найдутся асимптотические простое и пустое пространство $(\mathscr{M}', \mathbf{g}')$ и окрестность $\mathscr{U}'$ края $\partial\mathscr{M}'$ многообразия $\mathscr{M}'$, такие, что $\mathscr{U}' \cap \mathscr{M}'$ изометрично некоторому открытому множеству $\mathscr{U} \subset \mathscr{M}$. Под это определение подходят все упомянутые выше случаи. В решениях Райсснера — Нордстрема и Керра существуют бесконечные последовательности из асимптотически плоских областей $\mathscr{U}$, которые изометричны окрестностям $\mathscr{U}'$ асимптотически простых и пустых пространств. Таким образом, в них изотропные бесконечности образуют бесконечные последовательности. Однако мы будем рассматривать в этих пространствах только одну асимптотически плоскую область. Тогда можно считать $(\mathscr{M}, \mathbf{g})$ конформно-вложенным в пространство $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$ так, чтобы окрестность $\mathscr{U}$ края $\partial\mathscr{M}$ в $\tilde{\mathscr{M}}$ была изометрична $\mathscr{U}'$. Край $\partial\mathscr{M}$ состоит только из пары изотропных поверхностей $\mathscr{I}^+$ и $\mathscr{I}^-$.

Обсуждение слабо асимптотически простых и пустых пространств будет продолжено в разд. 9.2 и 9.3.

Глава 7

Задача Коши в общей теории относительности

В этой главе вкратце рассмотрена задача Коши в общей теории относительности. Мы покажем, что при задании определенных данных на пространственноподобной 3-поверхности $\mathscr{S}$ существует единственная максимальная область Коши $D^+(\mathscr{S})$ и что метрика на множестве $\mathscr{U} \subset D^+(\mathscr{S})$ зависит только от начальных данных на $J^-(\mathscr{U}) \cap \mathscr{S}$. Мы установим также, что эта зависимость непрерывна, если $\mathscr{U}$ обладает компактным замыканием в $D^+(\mathscr{S})$. Обсуждение этого вопроса включено сюда как потому, что он интересен сам по себе, так и потому, что при этом используются некоторые результаты предыдущей главы, и, наконец, для того чтобы показать, что уравнения поля Эйнштейна действительно удовлетворяют постулату (а) разд. 3.2, согласно которому сигнал может быть передан из одной точки в другую, только когда эти точки можно соединить непространственноподобной кривой. Однако в последующих трех главах результаты данной главы фактически не используются, и поэтому читатель, больше интересующийся сингулярностями, может ее пропустить.

В разд. 7.1 обсуждается специфика рассматриваемой задачи и дана ее точная формулировка. Для того чтобы связь между тензором Риччи и метрикой, установленную в области покрытой одной координатной системой, представить в форме, справедливой на всем многообразии, в разд. 7.2 вводится глобальная фоновая метрика $\mathbf{g}$. На ковариантные производные физической метрики $\mathbf{g}$ относительно фоновой метрики $\hat{\mathbf{g}}$ мы налагаем четыре калибровочных условия. При этом исключаются четыре степени свободы, соответствующие возможности диффеоморфизмов решений уравнений Эйнштейна, и мы приходим к приведенным гиперболическим уравнениям Эйнштейна второго порядка для метрики $\mathbf{g}$ относительно фоновой метрики $\hat{\mathbf{g}}$. В силу уравнений сохранения эти калибровочные условия выполняются во все моменты времени, если они и их первые производные удовлетворяются в начальный момент.

В разд. 7.3 будет показано, что существенную часть начальных данных для $\mathbf{g}$ на трехмерном многообразии $\mathscr{S}$ можно выразить через два трехмерных тензорных поля h^{ab}, χ^{ab} на $\mathscr{S}$. При этом трехмерное многообразие $\mathscr{S}$ вкладывается в четырехмерное многообразие $\mathscr{M}$, а метрика $\mathbf{g}$ определяется на $\mathscr{S}$ так, что h^{ab}

и χ^{ab} становятся соответственно первой и второй фундаментальными формами на $\mathscr{S}$ в метрике $\mathbf{g}$. Это можно сделать так, чтобы на $\mathscr{S}$ выполнялись калибровочные соотношения. В разд. 7.4 мы установим некоторые основные неравенства для гиперболических уравнений второго порядка. Они связывают интегралы от квадратичных выражений по первым производным решений с начальными данными для этих решений. В разд. 7.5 мы докажем существование и единственность решений приведенных уравнений Эйнштейна с $T_{\alpha\beta} = 0$ (пустое пространство) в виде малых возмущений какого-либо заданного решения в пустом пространстве. После этого существование и единственность локальных решений для произвольных начальных данных доказываются путем деления начальной гиперповерхности на малые почти плоские области и последующего соединения полученных решений.

В разд. 7.6 будет показано, что при фиксированных начальных данных существует единственное максимальное решение в пустом пространстве и что его зависимость от начальных данных в определенном смысле непрерывна. Наконец, в разд. 7.7 мы наметим путь для распространения этих результатов на решения при наличии материи.

7.1. Характер задачи

Задача Коши для гравитационного поля в некоторых важных аспектах отличается от той же задачи для других физических полей.

1. Уравнения Эйнштейна *нелинейны*. Правда, в этом отношении они не слишком сильно отличаются от других полей: хотя электромагнитное, скалярное и прочие поля в заданном пространстве-времени *сами по себе* подчиняются линейным уравнениям, они образуют нелинейную систему при учете их взаимодействий. Отличительным свойством гравитационного поля является то, что оно представляет собой поле с *самодействием*: оно нелинейно даже в отсутствие других полей. Именно поэтому гравитационное поле характеризует пространство-время, в котором оно распространяется. Для получения решений нелинейных уравнений обычно пользуются методом последовательных приближений такими линейными уравнениями, относительно решений которых известно, что они сходятся к решению нелинейного уравнения в определенной окрестности начальной поверхности.

2. Две метрики $\mathbf{g}_1$ и $\mathbf{g}_2$ на многообразии $\mathscr{M}$ физически эквивалентны, если существует диффеоморфизм $\phi: \mathscr{M} \to \mathscr{M}$, который переводит $\mathbf{g}_1$ и $\mathbf{g}_2$ ($\phi_* \mathbf{g}_1 = \mathbf{g}_2$); ясно, что $\mathbf{g}_1$ удовлетворяет уравнениям поля, если и только если им удовлетворяет $\mathbf{g}_2$. Поэтому решения уравнений поля могут быть единственны лишь с точностью до диффеоморфизма. Для того чтобы из этого класса

эквивалентности метрик выделить определенный член, описывающий пространство-время, вводят фиксированную «фоновую» метрику и накладывают четыре «калибровочных условия» на ковариантные производные физической метрики относительно фоновой метрики. Эти условия устраняют четыре степени свободы, соответствующие указанным диффеоморфизмам, и обеспечивают единственное решение для компонент метрики. Все это аналогично наложению условий Лоренца на электромагнитное поле для устранения свободы в выборе калибровки.

3. Поскольку метрика определяет структуру пространства-времени, мы не знаем заранее, какова область Коши начальной поверхности, а следовательно, не знаем и области, в которой должно быть определено решение. Дается только трехмерное многообразие $\mathscr{S}$ с некоторыми начальными данными $\boldsymbol{\omega}$ на нем и требуется найти четырехмерное многообразие $\mathscr{M}$, вложение $\theta: \mathscr{S} \to \mathscr{M}$ и метрику $\mathbf{g}$ на $\mathscr{M}$, удовлетворяющую уравнениям Эйнштейна, начальным данным на $\theta(\mathscr{S})$ и требованию, чтобы вложение $\theta(\mathscr{S})$ было поверхностью Коши в $\mathscr{M}$. Мы будем говорить, что $(\mathscr{M}, \theta, \mathbf{g})$, или просто $\mathscr{M}$, есть *развитие* пары $(\mathscr{S}, \boldsymbol{\omega})$. Другое развитие $(\mathscr{M}', \theta', \mathbf{g}')$ той же пары $(\mathscr{S}, \boldsymbol{\omega})$ назовем *расширением развития* $\mathscr{M}$, если существует диффеоморфизм α многообразия $\mathscr{M}$ на $\mathscr{M}'$, который оставляет образ $\mathscr{S}$ точечно неподвижным и переводит $\mathbf{g}'$ в $\mathbf{g}$ (т. е. $\theta^{-1}\alpha^{-1}\theta' = \mathrm{id}$ на $\mathscr{S}$, и $\alpha_* \mathbf{g}' = \mathbf{g}$). Мы покажем, что при выполнении на $\mathscr{S}$ определенных *уравнений связи* развития $(\mathscr{S}, \boldsymbol{\omega})$ существуют и, далее, что существует развитие, которое максимально в том смысле, что является расширением любого другого развития пары $(\mathscr{S}, \boldsymbol{\omega})$. Заметим, при такой формулировке задачи Коши учтена возможность производить диффеоморфизмы, поскольку всякое развитие представляет собой расширение любого своего диффеоморфизма, сохраняющего неподвижными точки образа поверхности $\mathscr{S}$.

7.2. Приведенные уравнения Эйнштейна

В гл. 2 тензор Риччи был выражен через частные производные компонент метрического тензора по координатам. Для целей этой главы удобно получить выражение, применимое во всем многообразии $\mathscr{M}$, а не только по отдельности в каждой координатной окрестности. Для этой цели введем наряду с физической метрикой $\mathbf{g}$ фоновую метрику $\hat{\mathbf{g}}$. При двух метриках требуется осторожность в обращении с ковариантными и контравариантными индексами. (Во избежание путаницы мы откажемся на время от обычного правила поднятия и опускания индексов.) Ковариантные и контравариантные формы $\mathbf{g}$ и $\hat{\mathbf{g}}$ связаны соотношением

$$g^{ab} g_{bc} = \delta^a{}_c = \hat{g}^{ab} \hat{g}_{bc}. \tag{7.1}$$

Удобно принять контравариантную форму метрики g^{ab} за основу и считать, что ковариантная форма выводится из (7.1). Используя альтернирующий тензор $\hat{\eta}_{abcd}$, определенный фоновой метрикой, можно представить связь между g_{ab} и g^{ab} в следующем явном виде:

$$g_{ab} = \frac{1}{3!} g^{cd} g^{ef} g^{ij} (\det \mathbf{g})\, \hat{\eta}_{acei} \hat{\eta}_{bdfj}, \tag{7.2}$$

где

$$(\det \mathbf{g})^{-1} \equiv \frac{1}{4!} g^{ab} g^{cd} g^{ef} g^{ij} \hat{\eta}_{acei} \hat{\eta}_{bdfj}$$

есть определитель из компонент $\mathbf{g}^{ab}$ в базисе, ортонормированном относительно метрики $\hat{\mathbf{g}}$.

Разность связности $\boldsymbol{\Gamma}$, определяемой метрикой $\mathbf{g}$, и связности $\hat{\boldsymbol{\Gamma}}$, определяемой $\hat{\mathbf{g}}$, является тензором и может быть выражена через ковариантную производную $\mathbf{g}$ относительно $\hat{\boldsymbol{\Gamma}}$ (ср. с разд. 3.3):

$$\delta\Gamma^a{}_{bc} \equiv \Gamma^a{}_{bc} - \hat{\Gamma}^a{}_{bc} = \frac{1}{2} g^{ij}{}_{|k} \left(g_{bi} g_{cj} g^{ak} - g_{bi} \delta^k{}_c \delta^a{}_j - g_{ci} \delta^k{}_b \delta^a{}_j\right), \tag{7.3}$$

где вертикальной чертой обозначена ковариантная производная относительно $\hat{\boldsymbol{\Gamma}}$, а символом δ — разность величин определенных относительно $\mathbf{g}$ и $\hat{\mathbf{g}}$. Тогда из (2.20) следует, что

$$\delta R_{ab} = \delta\Gamma^d{}_{ab\,|d} - \delta\Gamma^d{}_{ad\,|b} + \delta\Gamma^d{}_{ab}\, \delta\Gamma^e{}_{de} - \delta\Gamma^d{}_{ae}\, \delta\Gamma^e{}_{bd}. \tag{7.4}$$

Отсюда

$$\begin{aligned}
\delta\left(R^{ab} - \frac{1}{2} g^{ab} R\right) &= g^{ai} g^{bj}\, \delta R_{ij} + 2\, \delta g^{i(a} g^{b)j} \hat{R}_{ij} - \delta g^{ai}\, \delta g^{bj} \hat{R}_{ij} - \\
&\quad - \frac{1}{2} \delta g^{ab} \hat{R} - \frac{1}{2} g^{ab} (\delta g^{ij} \hat{R}_{ij} + g^{ij}\, \delta R_{ij}) = \\
&= \frac{1}{2} g^{ij}\, \delta g^{ab}{}_{|ij} - g^{i(a} \psi^{b)}{}_{|i} + \frac{1}{2} g^{ab} \left(\psi^i{}_{|i} - g_{cd} g^{ij}\, \delta^{cd}{}_{|ij}\right) + \\
&\quad + (\text{члены, содержащие } \delta g^{cd}{}_{|i} \text{ и } \delta g^{ef}),
\end{aligned} \tag{7.5}$$

$$\begin{aligned}
\psi^b \equiv g^{bc}{}_{|c} - \frac{1}{2} g^{bc} g_{de} g^{de}{}_{|c} &= (\det \mathbf{g})^{-1} \left((\det \mathbf{g})\, g^{bc}\right)_{|c} = \\
&= (\det \mathbf{g})^{-1}\, \phi^{bc}{}_{|c}
\end{aligned} \tag{7.6}$$

и

$$\phi^{bc} \equiv (\det \mathbf{g})\, \delta g^{bc}.$$

Наш план состоит в следующем. Выберем подходящую фоновую метрику $\hat{\mathbf{g}}$ и представим уравнения Эйнштейна в виде

$$R^{ab} - \frac{1}{2} R g^{ab} = \delta\left(R^{ab} - \frac{1}{2} R g^{ab}\right) + \hat{R}^{ab} - \frac{1}{2} \hat{g}^{ab} \hat{R} = 8\pi T^{ab}. \tag{7.7}$$

Это уравнение будем рассматривать как нелинейное дифференциальное уравнение второго порядка для определения $\mathbf{g}$ через его значения и значения его первых производных на некоторой начальной поверхности. Разумеется, чтобы эта система была полна, нужно задать уравнения, которым подчиняются физические поля, входящие в тензор энергии-импульса T^{ab}. Однако даже тогда, когда это сделано, мы еще не имеем системы, которая однозначно определяет развитие во времени через начальные данные. Причина этого уже упоминалась: решения уравнений Эйнштейна могут быть единственны лишь с точностью до диффеоморфизма. Для получения определенного решения этот произвол устраняется наложением четырех *калибровочных условий* на ковариантные производные $\mathbf{g}$ относительно фоновой метрики $\hat{\mathbf{g}}$. Мы будем пользоваться так называемыми «гармоническими» условиями:

$$\psi^b = \phi^{bc}{}_{|c} = 0,$$

которые аналогичны калибровочному условию Лоренца $A^i{}_{;i} = 0$ в электродинамике. Вместе с этим условием получим *приведенные уравнения Эйнштейна*

$$g^{ij}\phi^{ab}{}_{|ij} + (\text{члены с } \phi^{cd}{}_{|e} \text{ и } \phi^{ab}) = 16\pi T^{ab} - 2\hat{R}^{ab} + \hat{g}^{ab}\hat{R}. \quad (7.8)$$

Мы обозначим левую часть уравнения (7.8) через $E^{ab}{}_{cd}(\phi^{cd})$, где $E^{ab}{}_{cd}$ — *оператор Эйнштейна*. Таким образом, при подходящем выборе тензора энергии-импульса мы получаем гиперболические уравнения второго порядка и в разд. 7.5 продемонстрируем существование и единственность его решений. Но до этого нужно еще проверить совместность гармонических условий с уравнениями Эйнштейна. Поясним это: мы получили (7.8) из уравнений Эйнштейна, положив $\phi^{bc}{}_{|c} = 0$. Теперь мы должны убедиться, что решение, к которому приводит (7.8), действительно обладает таким свойством. Для этого продифференцируем (7.8) и свернем. Получим уравнение вида

$$g^{ij}\psi^b{}_{|ij} + B_c{}^{bi}\psi^c{}_{|i} + C_c{}^b\psi^c = 16\pi T^{ab}{}_{;a}, \quad (7.9)$$

где точкой с запятой обозначена производная относительно $\mathbf{g}$, а тензоры $B_c{}^{bi}$ и $C_c{}^b$ зависят от $\hat{g}^{ab}$, $\hat{R}^a{}_{bcd}$, g^{ab} и $g^{ab}{}_{|c}$. Уравнение (7.9) можно рассматривать как линейное гиперболическое уравнение второго порядка для ψ^b. Тогда, поскольку правая часть обращается в нуль, можно воспользоваться теоремой единственности решения для таких уравнений (предложение 7.4.5) и показать, что $\psi^b = 0$ всюду, если значения этой величины и ее первых производных равны нулю на начальной поверхности.

В следующем разделе мы увидим, что подходящим диффеоморфизмом начальные данные для ψ_b можно обратить в нуль.

Остается еще показать, что единственное решение, полученное наложением гармонического калибровочного условия, связано диффеоморфизмом с любым другим решением уравнений Эйнштейна с теми же начальными данными. Это будет сделано в разд. 7.4 путем специального выбора фоновой метрики.

7.3. Начальные данные

Поскольку (7.8) является гиперболической системой второго порядка, может показаться, что для получения решения следует задать значения g^{ab} и $g^{ab}{}_{|c}\,u^c$ на начальной гиперповерхности $\theta(\mathscr{S})$, где u^c — некоторое векторное поле, касательное к $\theta(\mathscr{S})$. Однако эти 20 компонент не все существенны или независимы; при произвольном изменении некоторых из них решение самое большее претерпевает диффеоморфизм, в то же время другие должны удовлетворять определенным условиям совместности.

Рассмотрим диффеоморфизм $\mu:\mathscr{M}\to\mathscr{M}$, оставляющий $\theta(\mathscr{S})$ точечно неподвижным. Он индуцирует отображение μ_*, переводящее g^{ab} в точке $p\in\theta(\mathscr{S})$ в новый тензор $\mu_* g^{ab}$ в той же точке. Если вектор $n_a\in T^*_p$ ортогонален к $\theta(\mathscr{S})$ [т. е. $n_a V^a=0$ для любого $V^a\in T_p$ касательного к $\theta(\mathscr{S})$] и нормирован так, что $n_a\hat{g}^{ab}n_b=-1$, то подходящим выбором μ вектор $n_a\mu_* g^{ab}$ можно сделать равным любому не касательному к $\theta(\mathscr{S})$ вектору в точке p. Следовательно, компоненты $n_a g^{ab}$ не существенны. С другой стороны, поскольку μ оставляет $\theta(\mathscr{S})$ точечно неподвижным, метрика $h_{ab}=\theta^* g_{ab}$, индуцированная на $\mathscr{S}$, будет оставаться неизменной. Поэтому только та часть $\mathbf{g}$, которая лежит на $\theta(\mathscr{S})$, необходима для получения решения. Остальные компоненты $n_a g^{ab}$ можно задавать произвольно; решение при этом самое большее претерпит диффеоморфизм. Чтобы прийти к этому выводу иным путем, вспомним, что мы сформулировали задачу Коши для начальных данных, определенных на абстрактном трехмерном многообразии $\mathscr{S}$, и затем искали вложение его в некоторое четырехмерное многообразие. На самой поверхности $\mathscr{S}$ нельзя задать четырехмерное тензорное поле вроде $\mathbf{g}$, а можно ввести только трехмерную метрику $\mathbf{h}$, которую мы будем считать положительно определенной. Контравариантная и ковариантная компоненты $\mathbf{h}$ связаны соотношением

$$h^{ab}h_{bc}=\delta^a{}_c, \tag{7.10}$$

где теперь $\delta^a{}_c$ — трехмерный тензор на $\mathscr{S}$. Вложение θ преобразует h_{ab} в контравариантное тензорное поле $\theta_* h^{ab}$ на $\theta(\mathscr{S})$, обладающее свойством

$$n_a\theta_* h^{ab}=0. \tag{7.11}$$

Ввиду произвольности $n_a g^{ab}$ мы теперь можем на $\theta(\mathscr{S})$ задать тензорное поле $\mathbf{g}$:

$$g^{ab} = \theta_* h^{ab} - u^a u^b, \tag{7.12}$$

где u^a — произвольное, не касательное к $\theta(\mathscr{S})$ и всюду отличное от нуля векторное поле на $\theta(\mathscr{S})$. Определяя g_{ab} из (7.1), имеем

$$h_{ab} = \theta^* g_{ab}, \quad n_a g^{ab} = -n_a u^a u^b, \quad g_{ab} u^a u^b = -1. \tag{7.13}$$

Отсюда h_{ab} — метрика, индуцированная на $\mathscr{S}$ метрикой $\mathbf{g}$, а u^a — единичный вектор, ортогональный к $\theta(\mathscr{S})$ в метрике $\mathbf{g}$.

Положение с первыми производными $g^{ab}{}_{|c}\, u^c$ аналогичное: $n_a g^{ab}{}_{|c}\, u^c$ можно привести соответствующим диффеоморфизмом к любым наперед заданным значениям. Однако здесь имеется дополнительная сложность, состоящая в том, что $g^{ab}{}_{|c}$ зависит не только от $\mathbf{g}$, но и от фоновой метрики $\hat{\mathbf{g}}$. Чтобы выразить существенную часть первой производной $\mathbf{g}$ только через тензорные поля, заданные на $\mathscr{S}$, мы поступим следующим образом. Введем на $\mathscr{S}$ симметричное контравариантное тензорное поле χ^{ab}. При вложении χ^{ab} отображается в тензорное поле $\theta_*\chi^{ab}$ на $\theta(\mathscr{S})$. Потребуем, чтобы $\theta_*\chi^{ab}$ было равно второй фундаментальной форме (см. разд. 2.7) подмногообразия $\theta(\mathscr{S})$ в метрике $\mathbf{g}$. Тогда получим

$$\begin{aligned}\theta_*\chi^{ab} &= \theta_* h^{ac}\theta_* h^{bd}(u^e g_{ec})_{;d} = \\ &= \theta_* h^{ac}\theta_* h^{bd}\left((u^e g_{ec})_{|d} - \delta\Gamma^f{}_{cd} u^e g_{ef}\right).\end{aligned} \tag{7.14}$$

Используя (7.3), находим

$$\theta_*\chi^{ab} = \frac{1}{2}\theta_* h^{ac}\theta_* h^{bd}\left(-g_{ci}g_{dj}g^{ij}{}_{|k}u^k + g_{bi}u^i{}_{|c} + g_{ci}u^i{}_{|b}\right). \tag{7.15}$$

Обращением этого равенства можно получить выражение для $g^{ab}{}_{|c}u^c$ через $\theta_*\chi^{ab}$:

$$\frac{1}{2}g^{ab}{}_{|c}u^c = -\theta_*\chi^{ab} + \theta_* h^{ac}\theta_* h^{bd} g_{i(c}u^i{}_{|d)} + u^{(a}W^{b)}, \tag{7.16}$$

где W^b — некоторое векторное поле на $\theta(\mathscr{S})$. Подходящим диффеоморфизмом μ ему можно придать любое нужное значение.

Тензорные поля h^{ab} и χ^{ab} нельзя задать на $\mathscr{S}$ совершенно независимо, поскольку свертка уравнений Эйнштейна (7.7) с n_a дает четыре уравнения, которые не содержат $g^{ab}{}_{|cd}u^c u^d$, т. е. вторых производных $\mathbf{g}$, имеющих ненулевые проекции на нормаль к $\theta(\mathscr{S})$. Таким образом, должны существовать четыре связи между g^{ab}, $g^{ab}{}_{|c}u^c$ и $n_a T^{ab}$. Используя (2.36) и (2.35), их можно

представить в виде уравнений на трехмерном многообразии $\mathscr{S}$:

$$\chi^{cd}{}_{\|d}h_{ce} - \chi^{cd}{}_{\|e}h_{cd} = 8\pi\theta^*(T_{de}u^d), \tag{7.17}$$

$$\frac{1}{2}(R' + (\chi^{dc}h_{dc})^2 - \chi^{ab}\chi^{cd}h_{ac}h_{bd}) = 8\pi\theta^*(T_{de}u^d u^e), \tag{7.18}$$

где двойной вертикальной чертой $\|$ обозначена ковариантная производная на $\mathscr{S}$ относительно метрики $\mathbf{h}$, а R' есть скалярная кривизна, полученная из $\mathbf{h}$.

Таким образом, данные $\boldsymbol{\omega}$ на $\mathscr{S}$, необходимые для получения решения, состоят из начальных данных для материальных полей (например, в случае скалярного поля ϕ они состояли бы из двух функций на $\mathscr{S}$ — значения ϕ и ее нормальной производной) и двух тензорных полей h^{ab} и χ^{ab} на $\mathscr{S}$, которые подчиняются *уравнениям связи* (7.17), (7.18). Последние представляют собой эллиптические уравнения на поверхности $\mathscr{S}$ и накладывают четыре связи на 12 независимых компонент (h^{ab}, χ^{ab}). При этих условиях можно показать, что восемь из этих компонент задаются независимо, а оставшиеся четыре находятся решением уравнений связи (см., например, [17]). Пару ($\mathscr{S}$, $\boldsymbol{\omega}$), удовлетворяющую этим требованиям, будем называть *комплектом начальных данных*. Следующим нашим шагом будет вложение $\mathscr{S}$ в подходящее четырехмерное многообразие $\mathscr{M}$ с метрикой $\hat{\mathbf{g}}$ и задание g_{ab} на $\theta(\mathscr{S})$ соотношением (7.12) при некотором выборе вектора u^a. Положим $u^a = \hat{g}^{ab}n_b$. Тогда u^a будет ортогонален к $\theta(\mathscr{S})$ в обеих метриках: $\mathbf{g}$ и $\hat{\mathbf{g}}$. Воспользуемся также свободой выбора W^a в выражении $g^{ab}{}_{|c}u^c$ из (7.16) для того, чтобы обратить ψ^b в нуль на $\theta(\mathscr{S})$. Это приведет к требованию

$$W^b = -g^{bc}{}_{|d}g_{ce}\theta_* h^{ed} + \frac{1}{2}g_{cd}g^{cd}{}_{|e}\theta_* h^{eb} + \\ + u^b(g_{cd}\theta_*\chi^{cd} - g_{lc}u^l{}_{|d}\theta_* g^{cd}). \tag{7.19}$$

[Заметим, что все производные в (7.19) касательны к $\theta(\mathscr{S})$, как и требуется тем обстоятельством, что рассматриваемые поля должны быть заданы на одной только поверхности $\theta(\mathscr{S})$.] Для выполнения всюду в $\mathscr{M}$ условия $\psi^b = 0$ нужно, кроме того, чтобы вектор $\psi^b{}_{|c}\,u^c$ обращался в нуль на $\theta(\mathscr{S})$. Но это сразу следует из уравнений связи, если на $\theta(\mathscr{S})$ удовлетворяются приведенные уравнения Эйнштейна (7.8). Следовательно, мы можем приступать к решению уравнений (7.8) как нелинейной гиперболической системы второго порядка на многообразии $\mathscr{M}$ с метрикой $\hat{\mathbf{g}}$.

[Заметим, что для ϕ^{bc} мы имеем 10 уравнений; при доказательстве существования решений этих 10 уравнений мы не будем делить их на систему уравнений связи и систему уравнений, описывающих развитие во времени, и, следовательно, не возникает вопроса об удовлетворении уравнений связи вне $\theta(\mathscr{S})$.]

7.4. Гиперболические уравнения второго порядка

В этом разделе мы воспроизведем некоторые результаты относительно гиперболических уравнений второго порядка, приведенные в [40]. При этом мы обобщим их так, чтобы они были применимы ко всему многообразию, а не только к одной координатной окрестности. Эти результаты будут использованы в последующих разделах для доказательства существования и единственности развитий комплекта начальных данных $(\mathscr{S}, \omega)$.

Введем сначала ряд определений. Мы будем пользоваться прописными латинскими буквами для обозначения совокупностей контравариантных или ковариантных индексов; так, тензор типа (r, s) будем записывать как $K^I{}_J$, а число индексов, которые представляют собирательный индекс I, обозначим через $|I|$, т. е. $|I| = r$. Введем на $\mathscr{M}$ положительно определенную метрику e_{ab} и

$$e_{IJ} = \underbrace{e_{ab}e_{cd} \ldots e_{pq}}_{r \text{ сомножителей}}, \quad e^{IJ} = \underbrace{e^{ab}e^{cd} \ldots e^{pq}}_{r \text{ сомножителей}},$$

где $|I| = |J| = r$. Затем определим $|K^I{}_J|$ (или просто $|\mathbf{K}|$), норму тензора $K^I{}_J$: $|K^I{}_J| = (K^I{}_J K^L{}_M e_{IL} e^{JM})^{1/2}$, где под повторяющимися собирательными индексами подразумевается свертка по всем индексам, которые они представляют. Через $|D^m K^I{}_J|$ (или просто $|D^m \mathbf{K}|$) будем обозначать $|K^I{}_{J|L}|$, где $|L| = m$, а черта $|$ означает, как и прежде, ковариантное дифференцирование относительно $\hat{\mathbf{g}}$.

Пусть $\mathscr{N}$ — некоторое вложенное подмногообразие с компактным замыканием в $\mathscr{M}$. Тогда $\| K^I{}_J, \mathscr{N} \|_m$ определяется как

$$\left\{ \sum_{p=0}^{m} \int_{\mathscr{N}} (|D^p K^I{}_J|)^2 \, d\sigma \right\}^{1/2},$$

где $d\sigma$ — объемный элемент на $\mathscr{N}$, индуцированный метрикой e. Зададим еще $\| \mathbf{K}, \mathscr{N} \|_m$ в виде такого же выражения, но с производными, взятыми в касательном к $\mathscr{N}$ направлении. Очевидно, $\| \mathbf{K}, \mathscr{N} \|_m \geqslant \| \mathbf{K}, \mathscr{N} \|_m$.

Тогда векторные пространства тензорных полей $K^I{}_J$ типа (r, s) являются по определению *пространствами Соболева* $W^m(r, s, \mathscr{N})$, если их значения и значения их производных (в смысле обобщенных функций) определены почти всюду на $\mathscr{N}$ (т. е. всюду, исключая, быть может, множество меры нуль; в оставшейся части этого раздела почти всюду утверждения следует понимать в смысле «почти всюду») и если $\| K^I{}_J, \mathscr{N} \|_m$ — конечная величина. Пространство Соболева с нормой $\| \ , \mathscr{N} \|_m$ является банаховым

пространством, в котором тензорные C^m-поля типа (r, s) образуют плотное подмножество. Если $\mathbf{e}'$ — другая непрерывная положительно определенная метрика на $\mathscr{M}$, то найдутся такие положительные постоянные C_1 и C_2, что

$$C_1 |K^I{}_J| \leqslant |K^I{}_J|' \leqslant C_2 |K^I{}_J| \quad \text{на} \quad \mathscr{N}$$

и

$$C_1 \| K^I{}_J, \mathscr{N} \|_m \leqslant \| K^I{}_J, \mathscr{N} \|'_m \leqslant C_2 \| K^I{}_J, \mathscr{N} \|_m.$$

Следовательно, $\| \ , \mathscr{N} \|'_m$ будет эквивалентной нормой. Аналогично другая фоновая C^m-метрика $\hat{\mathbf{g}}'$ будет давать эквивалентную норму. В действительности из двух лемм, приведенных ниже, следует, что для $\hat{\mathbf{g}}'' \in W^m(\mathscr{N})$ и размерности $\mathscr{N}$, меньшей $2m$, норма, полученная с помощью ковариантных производных относительно $\hat{\mathbf{g}}''$, снова эквивалентна исходной.

Теперь мы приведем три фундаментальных результата для пространств Соболева. Их доказательства можно получить, пользуясь книгой Соболева [159]. При этом потребуются слабые ограничения на вид $\mathscr{N}$. Будет достаточно потребовать, чтобы в каждой точке p края $\partial\mathscr{N}$ можно было вложить в $\bar{\mathscr{N}}$ половину n-мерного конуса с вершиной в p, где n — размерность $\bar{\mathscr{N}}$. Это требование выполняется, в частности, когда край $\partial\mathscr{N}$ гладкий.

Лемма 7.4.1

Существует положительная постоянная P_1 (зависящая от $\mathscr{N}$, $\mathbf{e}$ и $\hat{\mathbf{g}}$), такая, что $|\mathbf{K}| \leqslant P_1 \| \mathbf{K}, \mathscr{N} \|_m$ на $\mathscr{N}$ для любого $K^I{}_J \in W^m(\mathscr{N})$ с $2m > n$, где n — размерность $\mathscr{N}$.

Если $K^I{}_J \in W^m(\mathscr{N})$ и $2m > n$, то отсюда и из того, что векторное пространство всех непрерывных полей $K^I{}_J$ на $\mathscr{N}$ с нормой $\sup_{\mathscr{N}} |\mathbf{K}|$ является банаховым пространством, следует непрерывность $K^I{}_J$ на $\mathscr{N}$. Аналогично, если $K^I{}_J \in W^{m+p}$, то $K^I{}_J$ будет C^p-полем на $\mathscr{N}$.

Лемма 7.4.2

Существует положительная постоянная P_2 (зависящая от $\mathscr{N}$, $\mathbf{e}$ и $\mathbf{g}$), такая, что для любых полей $K^I{}_J$, $L^P{}_Q \in W^m(\mathscr{N})$ с $4m \geqslant n$ выполняется неравенство

$$\| K^I{}_J L^P{}_Q, \mathscr{N} \|_0 \leqslant P_2 \| \mathbf{K}, \mathscr{N} \|_m \| \mathbf{L}, \mathscr{N} \|_m.$$

Если $n \leqslant 4$ и $2m > n$, то из этой и предыдущей лемм следует, что для любых двух полей $K^I{}_J$, $L^P{}_Q \in W^m(\mathscr{N})$ их произведение $K^I{}_J L^P{}_Q$ также принадлежит $W^m(\mathscr{N})$.

Лемма 7.4.3

Если $\mathcal{N}'$ есть $(n-1)$-мерное подмногообразие, гладко вложенное в $\mathcal{N}$, то существует положительная постоянная P_3 (зависящая от $\mathcal{N}$, $\mathcal{N}'$, $\mathbf{e}$ и $\hat{\mathbf{g}}$), такая, что для любого поля $K^I{}_J \in W^{m+1}(\mathcal{N})$ выполняется неравенство

$$\| \mathbf{K}, \mathcal{N}' \|_m \leqslant P_3 \| \mathbf{K}, \mathcal{N} \|_{m+1}.$$

Мы докажем существование и единственность развитий комплекта $(\mathcal{S}, \boldsymbol{\omega})$, когда $h^{ab} \in W^{4+a}(\mathcal{S})$ и $\chi^{ab} \in W^{3+a}(\mathcal{S})$, где a — любая неотрицательная целая. [Если поверхность $\mathcal{S}$ не компактна, то $h^{ab} \in W^m(\mathcal{S})$ означает, что $h^{ab} \in W^m(\mathcal{N})$ при каком-либо открытом подмножестве $\mathcal{N} \subset \mathcal{S}$ с компактным замыканием.] Для этого достаточно, чтобы на $\mathcal{S}$ h^{ab} принадлежала классу C^{4+a} и χ^{ab} — классу C^{3+a}; необходимым условием, согласно лемме 7.4.1, является принадлежность h^{ab} классу C^{2+a} и χ^{ab} — классу C^{1+a}. Решение для g^{ab}, которое при этом получается, будет принадлежать пространству $W^{4+a}(\mathcal{H})$ для каждой гладкой пространственноподобной поверхности $\mathcal{H}$ и, следовательно, производные до $(2+a)$-го порядка включительно будут ограничены, т. е. g^{ab} будет класса $C^{(2+a)-}$ на $\mathcal{M}$.

Эти условия дифференцируемости можно ослабить для случаев, подобных ударным волнам, когда решение на вполне «хороших» гиперповерхностях уже не принадлежит пространству W^4 (см. [29, 83, 120, 127]). Однако не найдено доказательства теоремы существования и единственности для общего случая, когда решение не принадлежит W^4. Условие принадлежности решения к классу W^4 уже само по себе является улучшением результатов предшествующих работ [29], и тем не менее оно несколько сильнее, чем хотелось бы, ибо для определения уравнений Эйнштейна в смысле обобщенных функций достаточно непрерывности метрики и локальной квадратичной интегрируемости ее обобщенных производных (т. е. при $\mathbf{g}$ класса C^0 и принадлежащем W^1). С другой стороны, всякое условие принадлежности классу W^p при $p < 4$ не гарантирует единственности геодезических, а при $p < 3$ — их существования. Сами мы стоим на той точке зрения, что эти различия в условиях дифференцируемости не существенны, поскольку, как было объяснено в разд. 3.1, для модели пространства-времени можно принять, что метрика принадлежит классу C^∞.

Чтобы доказать существование и единственность развитий, мы установим сейчас ряд фундаментальных неравенств (леммы 7.4.4 и 7.4.6) для гиперболических уравнений второго порядка, используя метод, подобный доказательству теорем сохранения в разд. 4.3.

Рассмотрим многообразие $\hat{\mathcal{M}}$ вида $\mathcal{H} \times R^1$, где $\mathcal{H}$ — трехмерное многообразие. Пусть $\mathcal{U}$ — открытое множество $\hat{\mathcal{M}}$ с компактным замыканием, у которого имеется граница $\partial\mathcal{U}$ и которое пересекает $\mathcal{H}(0)$, причем через $\mathcal{H}(t)$ обозначена поверхность $\mathcal{H} \times \{t\}$, $t \in R^1$. Пусть $\mathcal{U}_+$ и $\mathcal{U}(t')$ — те части $\mathcal{U}$, для которых

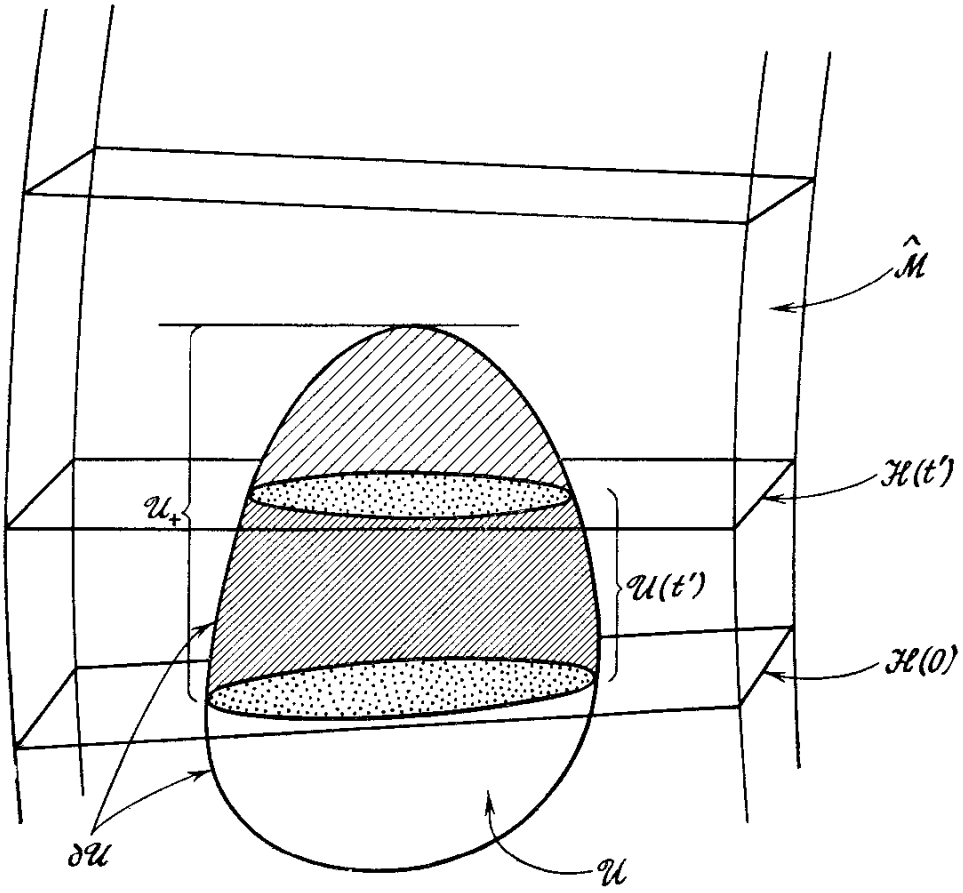


Рис. 48. $\mathcal{U}$ — открытое множество с компактным замыканием в многообразии $\hat{\mathcal{M}} = \mathcal{H} \times R^1$. $\mathcal{U}_+$ — область множества $\mathcal{U}$, для которой $t \geqslant 0$, и $\mathcal{U}(t')$ — область $\mathcal{U}$ между $t = 0$ и $t = t' > 0$.

соответственно $t \geqslant 0$ и $t' \geqslant t \geqslant 0$ (рис. 48). Пусть $\hat{\mathbf{g}}$ — фоновая C^2-метрика на $\mathcal{U}_+$, а $\mathbf{e}$ — положительно определенная C^1-метрика. Рассмотрим тензорные поля $K^I{}_J$, подчиняющиеся гиперболическому уравнению второго порядка вида

$$L(K) \equiv A^{ab} K^I{}_{J|ab} + B^{aPI}{}_{QJ} K^Q{}_{P|a} + C^{PI}{}_{QJ} K^Q{}_P = F^I{}_J, \qquad (7.20)$$

где $\mathbf{A}$ — лоренцева метрика на $\mathcal{U}_+$ (т. е. симметричное тензорное поле сигнатуры $+2$), $\mathbf{B}$, $\mathbf{C}$ и $\mathbf{F}$ — тензорные поля типов, обозначенных их индексами, и вертикальная черта означает ковариантное дифференцирование относительно метрики $\hat{\mathbf{g}}$.

Лемма 7.4.4
Если

1) $\partial\mathcal{U} \cap \bar{\mathcal{U}}_+$ ахронально относительно $\mathbf{A}$,

2) существует некоторая постоянная $Q_1 > 0$, для которой на $\bar{\mathcal{U}}_+$

$$A^{ab}t_{|a}t_{|b} \leqslant -Q_1$$

и

$$A^{ab}W_aW_b \geqslant Q_1 e^{ab}W_aW_b$$

для любой 1-формы $\mathbf{W}$, удовлетворяющей равенству $A^{ab}t_{|a}W_b = 0$,

3) существует некоторая постоянная Q_2, такая, что на $\bar{\mathcal{U}}_+$

$$|\mathbf{A}| \leqslant Q_2, \quad |\mathbf{DA}| \leqslant Q_2, \quad |\mathbf{B}| \leqslant Q_2, \quad |\mathbf{C}| \leqslant Q_2,$$

то найдется некоторая положительная постоянная P_4 (зависящая от $\mathcal{U}$, $\mathbf{e}$, $\hat{\mathbf{g}}$, Q_1 и Q_2), такая, что для любого решения $K^I{}_J$ уравнения (7.20) справедливо неравенство

$$\|\mathbf{K}, \mathcal{H}(t) \cap \mathcal{U}_+\|_1 \leqslant P_4\{\|\mathbf{K}, \mathcal{H}(0) \cap \mathcal{U}_+\|_1 + \|\mathbf{F}, \mathcal{U}(t)\|_0\}.$$

Для тензорного поля $K^I{}_J$ по аналогии с тензором энергии-импульса для скалярного поля единичной массы (разд. 3.2) образуем «тензор энергии» S^{ab}:

$$S^{ab} = \left\{\left(A^{ac}A^{bd} - \frac{1}{2}A^{ab}A^{cd}\right)K^I{}_{J|c}K^P{}_{Q|d} - \frac{1}{2}A^{ab}K^I{}_JK^P{}_Q\right\}e^{JQ}e_{IP}. \tag{7.21}$$

Тензор S^{ab} подчиняется условию энергодоминантности (разд. 4.3) относительно метрики $\mathbf{A}$ (иными словами, если вектор W_a времениподобен относительно $\mathbf{A}$, то $S^{ab}W_aW_b \geqslant 0$, а вектор $S^{ab}W_a$ непространственноподобен относительно $\mathbf{A}$). Более того, в силу условий (2) и (3) будут существовать такие положительные постоянные Q_3 и Q_4, что

$$Q_3(|\mathbf{K}|^2 + |\mathbf{DK}|^2) \leqslant S^{ab}t_{|a}t_{|b} \leqslant Q_4(|\mathbf{K}|^2 + |\mathbf{DK}|^2). \tag{7.22}$$

Теперь мы применим лемму 4.3.1 к S^{ab}, выбрав в качестве $\mathcal{U}_+$ компактную область $\mathcal{F}$ и пользуясь объемным элементом $\mathrm{d}\hat{v}$ и ковариантным дифференцированием, которые определяются метрикой $\hat{\mathbf{g}}$:

$$\int_{\mathcal{H}(t)\cap\bar{\mathcal{U}}_+} S^{ab}t_{|a}\,\mathrm{d}\hat{\sigma}_b \leqslant \int_{\mathcal{H}(0)\cap\bar{\mathcal{U}}_+} S^{ab}t_{|a}\,\mathrm{d}\hat{\sigma}_b + \int_0^t \left\{\int_{\mathcal{H}(t')\cap\bar{\mathcal{U}}_+} (PS^{ab}t_{|a} + S^{ab}{}_{|a})\,\mathrm{d}\hat{\sigma}_b\right\}\mathrm{d}t', \tag{7.23}$$

где P — положительная постоянная, не зависящая от S^{ab}. (В первом слагаемом правой части по сравнению с леммой 4.3.1

изменен знак из-за того, что элемент $\mathrm{d}\hat{\sigma}_b$ поверхности $\mathscr{H}(t)$ взят с той же ориентацией, что и $t_{|b}$, т. е. $\mathrm{d}\hat{\sigma}_b = t_{|b}\,\mathrm{d}\tilde{\sigma}$, где $\mathrm{d}\tilde{\sigma}$ — положительно определенная мера на $\mathscr{H}(t)$.) Ввиду непрерывности $\mathbf{e}$ и $\hat{\mathbf{g}}$ найдутся такие положительные постоянные Q_5 и Q_6, что на $\overline{\mathscr{U}}_+$ будет выполняться неравенство

$$Q_5\,\mathrm{d}\sigma \leqslant \mathrm{d}\tilde{\sigma} \leqslant Q_6\,\mathrm{d}\sigma, \tag{7.24}$$

где $\mathrm{d}\sigma$ — элемент площади на $\mathscr{H}(t)$, индуцированный метрикой $\mathbf{e}$. Таким образом, в силу (7.22) и (7.23) существует некоторая постоянная Q_7, такая, что

$$\| \mathbf{K},\ \mathscr{H}(t) \cap \mathscr{U}_+ \|_1^2 \leqslant Q_7 \Big\{ \| \mathbf{K},\ \mathscr{H}(0) \cap \mathscr{U}_+ \|_1^2 + \\ + \int_0^t \| \mathbf{K},\ \mathscr{H}(t') \cap \mathscr{U}_+ \|_1^2\,\mathrm{d}t' + \int_0^t (S^{ab}{}_{|b} t_{|a}\,\mathrm{d}\sigma)\,\mathrm{d}t' \Big\}. \tag{7.25}$$

Из (7.20) следует, что

$S^{ab}{}_{|b} = A^{ac} K^I{}_{J|c} F^P{}_Q e^{JQ} e_{IP}$ + члены, квадратичные по $K^I{}_J$ и $K^P{}_{Q|c}$, с коэффициентами, содержащими A^{cd}, $A^{cd}{}_{|e}$, $\hat{R}^c{}_{def}$, $B^{cPI}{}_{QJ}$ и $C^{PI}{}_{QJ}$. (7.26)

Поскольку все коэффициенты на $\mathscr{U}_+$ ограничены, существует такая постоянная Q_8, что

$$S^{ab}{}_{|b} t_{|a} \leqslant Q_8 \{ |\mathbf{F}|^2 + |\mathbf{K}|^2 + |\mathbf{DK}|^2 \}. \tag{7.27}$$

Следовательно, найдется постоянная Q_9, при которой в силу (7.25) и (7.27) будет выполняться неравенство

$$\| \mathbf{K},\ \mathscr{H}(t) \cap \mathscr{U}_+ \|_1 \leqslant Q_9 \Big\{ \| \mathbf{K},\ \mathscr{H}(0) \cap \mathscr{U}_+ \|_1^2 + \\ + \int_0^t \| \mathbf{K},\ \mathscr{H}(t') \cap \mathscr{U}_+ \|_1\,\mathrm{d}t' + \| \mathbf{F},\ \mathscr{U}(t) \|_0^2 \Big\};$$

это неравенство можно представить в виде

$$\frac{\mathrm{d}x}{\mathrm{d}t} \leqslant Q_9 \{x + y\}, \tag{7.28}$$

где

$$x(t) = \int_0^t \| \mathbf{K},\ \mathscr{H}(t') \cap \mathscr{U}_+ \|_1^2\,\mathrm{d}t'.$$

Отсюда

$$x \leqslant e^{Q_9 t} \int_0^t e^{-Q_9 t'} y(t')\,\mathrm{d}t'. \tag{7.29}$$

Поскольку y — монотонно растущая функция t и поскольку t ограничено в $\bar{\mathcal{U}}_+$, существует такая постоянная Q_{10}, которая обеспечивает выполнение неравенства

$$x \leqslant Q_{10} y.$$

Итак,

$$\| \mathbf{K}, \mathcal{H}(t) \cap \mathcal{U}_+ \|_1 \leqslant P_4 \{ \| \mathbf{K}, \mathcal{H}(0) \cap \mathcal{U}_+ \|_1 + \| \mathbf{F}, \mathcal{U}(t) \|_0 \},$$

где

$$P_4 = (Q_9 + Q_{10})^{1/2}. \qquad \square$$

С помощью этого неравенства мы немедленно можем доказать единственность решений линейных гиперболических уравнений 2-го порядка, т. е. уравнений, для которых **A**, **B**, **C** и **F** не зависят от **K**. В самом деле, пусть $K^1{}_J$ и $K^2{}_J$ будут решениями уравнения $L(\mathbf{K}) = \mathbf{F}$, причем начальные значения как их самих, так и их первых производных на $\mathcal{H}(0) \cap \mathcal{U}$ совпадают. Тогда предыдущий результат можно применить к уравнению $L(\mathbf{K}^1 - \mathbf{K}^2) = 0$ и получить, что

$$\| \mathbf{K}^1 - \mathbf{K}^2, \mathcal{H}(t) \cap \mathcal{U}_+ \|_1 = 0.$$

Отсюда $\mathbf{K}^1 = \mathbf{K}^2$ на $\mathcal{U}_+$. Таким образом, мы имеем

Предложение 7.4.5

Пусть **A** — лоренцева C^1-метрика на $\hat{\mathcal{M}}$ и пусть **B**, **C** и **F** локально ограничены. Пусть $\mathcal{H} \subset \hat{\mathcal{M}}$ — некоторая пространственноподобная и непричинная относительно **A** 3-поверхность. Тогда решение линейного уравнения (7.20) однозначно определено на множестве $\mathcal{V} \in D^+(\mathcal{H}, \mathbf{A})$ его значениями и значениями его первых производных на $\mathcal{H} \cap J^-(\mathcal{V}, \mathbf{A})$.

Согласно предложению 6.6.7, $D^+(\mathcal{H}, \mathbf{A})$ имеет вид $\mathcal{H} \times R^1$. Если $q \in \mathcal{V}$, то по предложению 6.6.6 множество $J^-(q) \cap J^+(\mathcal{H})$ компактно и может быть взято в качестве $\bar{\mathcal{U}}_+$. $\square$

Итак, физическое поле, подчиняющееся линейному уравнению вида (7.20), будет удовлетворять постулату причинности (а) разд. 3.2 при условии, что изотропный конус в метрике **A** совпадает с изотропным конусом пространственновременной метрики **g** или лежит внутри него.

Чтобы доказать существование решений уравнений (7.20), нам понадобятся неравенства для производных **K** более высокого порядка. Мы примем теперь, что фоновая метрика $\hat{\mathbf{g}}$ по крайней мере класса C^{5+a}, где a — неотрицательное целое число, а множество $\mathcal{U}$ возьмем таким, чтобы область $\mathcal{H}(0) \cap \bar{\mathcal{U}}$ имела

гладкую границу и чтобы при этом существовал диффеоморфизм

$$\lambda:(\mathscr{H}(0)\cap\bar{\mathscr{U}})\times[0,\ t_1]\to\bar{\mathscr{U}}_+,$$

обладающий тем свойством, что для всякого $t\in[0, t_1]$

$$\lambda\{(\mathscr{H}(0)\cap\bar{\mathscr{U}}),\ t\}=\mathscr{H}(t)\cap\bar{\mathscr{U}}_+.$$

Сделаем этот выбор так, чтобы на поверхности $\mathscr{H}(t)\cap\mathscr{U}_+$ существовали верхние границы $\tilde{P}_1$, $\tilde{P}_2$, $\tilde{P}_3$ для постоянных P_1, P_2, P_3 в леммах 7.4.1—7.4.3.

Лемма 7.4.6
Если выполнены условия (1) и (2) леммы 7.4.4 и если
4) существует такая постоянная Q_3, при которой

$$\|\mathbf{A},\ \mathscr{U}_+\|_{4+a}\leqslant Q_3,\quad \|\mathbf{B},\ \mathscr{U}_+\|_{3+a}<Q_3,\quad \|\mathbf{C},\ \mathscr{U}_+\|_{3+a}<Q_3$$

[в силу леммы 7.4.1 здесь содержится и условие (3)], то существуют такие положительные постоянные $P_{5,a}$ (зависящие от $\mathscr{U}$, $\mathbf{e}$, $\hat{\mathbf{g}}$, a, Q_1 и Q_3), что выполняется неравенство

$$\|\mathbf{K},\ \mathscr{H}(t)\cap\mathscr{U}_+\|_{4+a}\leqslant P_{5,a}\{\|\mathbf{K},\ \mathscr{H}(0)\cap\mathscr{U}_+\|_{4+a}+\|\mathbf{F},\ \mathscr{U}(t)\|_{3+a}\}. \tag{7.30}$$

Из леммы 7.4.4 мы имеем неравенство для

$$\|\mathbf{K},\ \mathscr{H}(t)\cap\mathscr{U}_+\|_1.$$

Чтобы получить неравенство для

$$\|\mathbf{K},\ \mathscr{H}(t)\cap\mathscr{U}_+\|_2,$$

образуем тензор «энергии» S^{ab} для первых производных $K^I{}_{J|c}$ и поступим, как прежде. Дивергенцию $S^{ab}{}_{|b}$ можно при этом вычислить дифференцированием уравнений (7.20):

$$S^{ab}{}_{|b}=A^{ad}K^I{}_{J|cd}F^P{}_{Q|e}e^{ec}e^{JQ}e_{IP}+(\text{члены, квадратичные по } K^I{}_{J|c} \text{ и } K^I{}_{J|cd} \text{ с коэффициентами, зависящими от } A^{cd},\ A^{cd}{}_{|e},\ \hat{R}^c{}_{def},\ \hat{R}^c{}_{def|g},\ B^{cPI}{}_{QJ},\ B^{cPI}{}_{QJ|d},\ C^{PI}{}_{QJ},\ C^{PI}{}_{QJ|d}); \tag{7.31}$$

Все эти коэффициенты (за исключением, быть может, $B^{cPI}{}_{QJ|d}$ и $C^{PI}{}_{QJ|d}$) ограничены на $\bar{\mathscr{U}}_+$ в случае $a=0$. При интегрировании по поверхности $\mathscr{H}(t')\cap\mathscr{U}_+$ в (7.31) член, включающий $B^{cPI}{}_{QJ|d}$, имеет вид

$$-\int_{\mathscr{H}(t')\cap\mathscr{U}_+}A^{ab}K^I{}_{J|cb}B^{dPR}{}_{QS|e}K^S{}_{R|d}e^{ce}e^{QJ}e_{PI}\,\mathrm{d}\hat{\sigma}_a. \tag{7.32}$$

Существует некоторая постоянная Q_4 для всех t', при которой интеграл (7.32) меньше или равен величине

$$Q_4 \int\limits_{\mathcal{H}(t')\cap\mathcal{U}_+} |\,\mathrm{DB}\,||\,\mathrm{DK}\,||\,\mathrm{D}^2\mathbf{K}\,|\,\mathrm{d}\sigma \leqslant$$

$$\leqslant \frac{1}{2} Q_4 \int\limits_{\mathcal{H}(t')\cap\mathcal{U}_+} |\,\mathrm{D}^2\mathbf{K}\,|^2 + |\,\mathrm{DB}\,|^2\,|\,\mathrm{DK}\,|^2\,\mathrm{d}\sigma. \qquad (7.33)$$

Согласно лемме 7.4.2,

$$\int\limits_{\mathcal{H}(t')\cap\mathcal{U}_+} |\,\mathrm{DB}\,|^2\,|\,\mathrm{DK}\,|^2\,\mathrm{d}\sigma \leqslant$$

$$\leqslant \tilde{P}_2{}^2 \|\,\mathbf{B},\ \mathcal{H}(t')\cap\mathcal{U}_+\,\|_2{}^2\,\|\,\mathbf{K},\ \mathcal{H}(t')\cap\mathcal{U}_+\,\|_2{}^2,$$

где по условию (4) и лемме 7.4.3

$$\|\,\mathbf{B},\ \mathcal{H}(t')\cap\mathcal{U}_+\,\|_2 < \tilde{P}_3 Q_3.$$

Ограничение на член, включающий $C^{PI}{}_{QJd}$, можно получить подобным же образом. Таким образом, согласно лемме 4.3.1, найдется постоянная Q_5, такая, что выполняется неравенство

$$\int\limits_{\mathcal{H}(t)\cap\mathcal{U}_+} (|\,\mathrm{D}^2\mathbf{K}\,| + |\,\mathrm{DK}\,|^2)\,\mathrm{d}\sigma \leqslant Q_5 \left\{ \int\limits_{\mathcal{H}(0)\cap\mathcal{U}_+} (|\,\mathrm{D}^2\mathbf{K}\,|^2 + |\,\mathrm{DK}\,|^2)\,\mathrm{d}\sigma + \right.$$

$$\left. + \int\limits_0^t \|\,\mathbf{K},\ \mathcal{H}(t')\cap\mathcal{U}_+\,\|_2{}^2\,\mathrm{d}t' + \int\limits_{\mathcal{U}(t)} |\,\mathrm{DF}\,|^2\,\mathrm{d}\sigma \right\}. \qquad (7.34)$$

Из леммы 7.4.4 следует, что

$$\int\limits_{\mathcal{H}(t)\cap\mathcal{U}_+} |\,\mathbf{K}\,|^2\,\mathrm{d}\sigma \leqslant \|\,\mathbf{K},\ \mathcal{H}(t)\cap\mathcal{U}_+\,\|_1{}^2 \leqslant$$

$$\leqslant 2P_4{}^2\{\|\,\mathbf{K},\ \mathcal{H}(0)\cap\mathcal{U}\,\|_1{}^2 + \|\,\mathbf{F},\ \mathcal{U}(t)\,\|_3{}^2\}. \qquad (7.35)$$

Складывая это неравенство с (7.34), получаем

$$\|\,\mathbf{K},\ \mathcal{H}(t)\cap\mathcal{U}_+\,\|_2{}^2 \leqslant Q_6 \left\{ \|\,\mathbf{K},\ \mathcal{H}(0)\cap\mathcal{U}\,\|_2{}^2 + \right.$$

$$\left. + \int\limits_0^t \|\,\mathbf{K},\ \mathcal{H}(t')\cap\mathcal{U}_+\,\|_2{}^2\,\mathrm{d}t' + \|\,\mathbf{F},\ \mathcal{U}(t)\,\|_1{}^2 \right\}, \qquad (7.36)$$

где $Q_6 = Q_5 + 2P_4$. Из рассуждений, аналогичных ходу доказательства леммы 7.4.4, следует существование некоторой постоянной Q_7, при которой

$$\|\,\mathbf{K},\ \mathcal{H}(t)\cap\mathcal{U}_+\,\|_2 \leqslant Q_7\{\|\,\mathbf{K},\ \mathcal{H}(0)\,\mathcal{U}\,\|_2 + \|\,\mathbf{F},\ \mathcal{U}(t)\,\|_1\}. \qquad (7.37)$$

Тогда из леммы 7.4.1 следует, что на $\mathcal{U}_+$ справедливо неравенство

$$|\mathbf{K}| \leqslant \tilde{P}_1 Q_7 \{\|\mathbf{K}, \mathcal{H}(0) \cap \mathcal{U}\|_2 + \|\mathbf{F}, \mathcal{U}(t)\|_0\}. \tag{7.38}$$

Используя это неравенство, можно подобным же методом получить неравенство для $\|\mathbf{K}, \mathcal{H}(t) \cap \mathcal{U}_+\|_3$. Дивергенция тензора «энергии» приводит к выражению вида

$$Q_8 \int\limits_{\mathcal{H}(t') \cap \mathcal{U}_+} (|\mathrm{D}^3\mathbf{K}|^2 + |\mathrm{D}^2\mathbf{B}|^2 |\mathrm{D}\mathbf{K}|^2)\, \mathrm{d}\sigma. \tag{7.39}$$

Согласно лемме 7.4.2, второе слагаемое ограничено величиной

$$Q_8 \tilde{P}_2^2 \|\mathbf{B}, \mathcal{H}(t') \cap \mathcal{U}_+\|_3^2 \|\mathbf{K}, \mathcal{H}(t') \cap \mathcal{U}_+\|_2^2,$$

где по условию (4) множитель $\|\mathbf{B}, \mathcal{H}(t) \cap \mathcal{U}_+\|_3$ определен почти всюду и квадратично-интегрируем по t'. Следовательно, мы можем получить неравенство для $\|\mathbf{K}, \mathcal{H}(t) \cap \mathcal{U}_+\|_3$ тем же способом, что и для $\|\mathbf{K}, \mathcal{H}(t) \cap \mathcal{U}_+\|_2$. Процедура получения неравенств для производных более высокого порядка аналогична. □

Следствие

Существуют постоянные $P_{6,a}$ и $P_{7,a}$, при которых

$$\|\mathbf{K}, \mathcal{H}(t) \cap \mathcal{U}_+\|_{4+a} \leqslant P_{6,a} \{\|\mathbf{K}, \mathcal{H}(0) \cap \mathcal{U}\|_{4+a} + \\ + \|K^I{}_{J|a} u^a, \mathcal{H}(0) \cap \mathcal{U}\|_{3+a} + \|\mathbf{F}, \mathcal{U}_+\|_{3+a}\}$$

и

$$\|\mathbf{K}, \mathcal{U}_+\|_{4+a} \leqslant P_{7,a} \text{ (тот же множитель)},$$

где u^a — некоторое векторное C^{3+a}-поле на $\mathcal{H}(0)$, причем оно нигде не касательно к $\mathcal{H}(0)$.

Согласно (7.20), производные $\mathbf{K}$ второго и более высоких порядков могут быть выражены через поле $\mathbf{F}$, его производные в направлении u^a, $K^I{}_{J|a}\, u^a$ и производные $\mathbf{K}$ в направлениях, касательных к поверхности $\mathcal{H}(0)$. Из леммы 7.4.3 следует, что

$$\begin{aligned} &\|\mathbf{A}, \mathcal{H}(0) \cap \mathcal{U}\|_{3+a} < \tilde{P}_3 Q_3, \\ &\|\mathbf{B}, \mathcal{H}(0) \cap \mathcal{U}\|_{2+a} < \tilde{P}_3 Q_3, \\ &\|\mathbf{C}, \mathcal{H}(0) \cap \mathcal{U}\|_{2+a} < \tilde{P}_3 Q_3, \\ &\|\mathbf{F}, \mathcal{H}(0) \cap \mathcal{U}\|_{2+a} < \tilde{P}_3 \|\mathbf{F}, \mathcal{U}_+\|_{3+a}. \end{aligned} \tag{7.40}$$

Итак, должна существовать некоторая постоянная Q_4, такая, что

$$\|\mathbf{K}, \mathcal{H}(0) \cap \mathcal{U}\|_{4+a} \leqslant Q_4 \{\|\mathbf{K}, \mathcal{H}(0) \cap \mathcal{U}\|_{4+a} + \\ + \|K^I{}_{J|a} u^a, \mathcal{H}(0) \cap \mathcal{U}\|_{3+a} + \|\mathbf{F}, \mathcal{U}_+\|_{3+a}\}. \tag{7.41}$$

Второй результат непосредственно следует отсюда, так как координата t ограничена на $\mathscr{U}_+$. □

Теперь мы можем приступить к доказательству существования решений линейных уравнений вида (7.20). Сначала предположим, что компоненты **A**, **B**, **C**, **F**, **u** и $\hat{\mathbf{g}}$ являются аналитическими функциями локальных координат x^1, x^2, x^3 и x^4 ($x^4 = t$) в некоторой координатной окрестности $\mathscr{V}$, а в качестве начальных данных $K^I{}_J = {}_0K^I{}_J$ и $K^I{}_{J|a}u^a = {}_1K^I{}_J$ возьмем аналитические функции координат x^1, x^2, x^3 на $\mathscr{H}(0) \cap \mathscr{V}$. Тогда из (7.20) можно вычислить частные производные компонент тензора **K** в направлении, не касательном к $\mathscr{H}(0)$: $\partial^2(K^I{}_J)/\partial t^2$, $\partial^3(K^I{}_J)/\partial t^2\,\partial x^i$, $\partial^3(K^I{}_J)/\partial t^3$ и т. д. через ${}_0K^I{}_J$ и ${}_1K^I{}_J$ на $\mathscr{H}(0)$. Таким образом, $K^I{}_J$ можно представить в виде формального степенного ряда по x^1, x^2, x^3 и t вокруг начала координат p. По теореме Коши — Ковалевской [36, p. 39] этот ряд будет сходиться в некотором шаре $\mathscr{V}(r)$ координатного радиуса r и даст решение уравнений (7.20) с заданными начальными условиями. После этого мы выберем аналитический атлас из C^∞-атласа многообразия $\mathscr{M}$, покроем $\mathscr{H}(0) \cap \bar{\mathscr{U}}$ координатными окрестностями вида $\mathscr{V}(r)$ из этого атласа и в каждой координатной окрестности построим решение так, как это было сделано выше. Таким образом мы получаем решение в области $\mathscr{U}(t_2)$ для некоторого $t_2 > 0$. Затем повторяем этот процесс, используя $\mathscr{H}(t_2)$. По теореме Коши — Ковалевской отношение последовательных интервалов t, для которых степенной ряд сходится, не зависит от начальных данных, и, следовательно, решение можно продолжить на всю область $\mathscr{U}_+$ за конечное число шагов. Этим доказывается существование решений линейных уравнений вида (7.20), когда коэффициенты, правая часть и начальные данные, аналитичны. Теперь мы снимем требование аналитичности.

Предложение 7.4.7

Если выполнены условия (1), (2), (4) и если

5) $\mathbf{F} \in W^{3+a}(\mathscr{U}_+)$,

6) ${}_0\mathbf{K} \in W^{4+a}(\mathscr{H}(0) \cap \bar{\mathscr{U}})$, ${}_1\mathbf{K} \in W^{3+a}(\mathscr{H}(0) \cap \bar{\mathscr{U}})$,

то уравнение (7.20) имеет единственное решение $\mathbf{K} \in W^{4+a}(\mathscr{U}_+)$, для которого на $\mathscr{H}(0)$ $K^I{}_J = {}_0K^I{}_J$ и $K^I{}_{J|a}u^a = {}_1K^I{}_J$.

Для доказательства этого утверждения мы аппроксимируем коэффициенты и начальные данные аналитическими функциями и покажем, что получающиеся при этом аналитические решения сходятся к полю, которое является решением данного уравнения с заданными начальными условиями. Пусть $\mathbf{A}_n$ ($n = 1, 2, 3$) — последовательность аналитических полей на $\bar{\mathscr{U}}_+$, сильно сходящаяся в $W^{4+a}(\mathscr{U}_+)$ к **A**. (Говорят, что $\mathbf{A}_n$ сильно сходится к **A**

в W^m, если $\| \mathbf{A}_n - \mathbf{A} \|_m$ сходится к нулю.) Пусть $\mathbf{B}_n$, $\mathbf{C}_n$ и $\mathbf{F}_n$ — аналитические поля на $\bar{\mathcal{U}}_+$, на которые сильно сходятся соответственно к $\mathbf{B}$, $\mathbf{C}$ и $\mathbf{F}$ в пространстве $W^{3+a}(\mathcal{U}_+)$, и пусть ${}_0\mathbf{K}_n$ и ${}_1\mathbf{K}_n$ — аналитические поля на $\mathscr{H}(0) \cap \bar{\mathcal{U}}$, сильно сходящиеся к ${}_0\mathbf{K}$ и ${}_1\mathbf{K}$ соответственно в $W^{4+a}(\mathscr{H}(0) \cap \mathcal{U})$ и $W^{3+a}(\mathscr{H}(0) \cap \mathcal{U})$. При каждом значении n будет существовать аналитическое решение $\mathbf{K}_n$ уравнения (7.20) с начальными значениями $K_n{}^I{}_J = {}_0K_n{}^I{}_J$, $K_n{}^I{}_{J|a}u^a = {}_1K_n{}^I{}_J$. Согласно следствию леммы 7.4.6, норма $\| \mathbf{K}_n, \mathcal{U}_+ \|_{4+a}$ ограничена при $n \to \infty$. Тогда по теореме Рисса [140] существует поле $\mathbf{K} \in W^{4+a}(\mathcal{U}_+)$ и подпоследовательность $K_{n'}$ последовательности $\mathbf{K}_n$, такая, что для каждого b, $0 \leqslant b \leqslant 4 + a$, $D^b\mathbf{K}_{n'}$ слабо сходится к $D^b\mathbf{K}$. (Последовательность полей $I_n{}^I{}_J$ на $\mathscr{N}$ называется слабо сходящейся к $I^I{}_J$, если

$$\int_{\mathscr{N}} I_n{}^I{}_J J^J{}_I \, d\sigma \to \int_{\mathscr{N}} I^I{}_J J^J{}_I \, d\sigma$$

для любого C^∞-поля $J^I{}_J$.)

Поскольку $\mathbf{A}_n$, $\mathbf{B}_n$ и $\mathbf{C}_n$ сильно сходятся в $W^3(\mathcal{U}_+)$ к $\mathbf{A}$, $\mathbf{B}$ и $\mathbf{C}$, $\sup|\mathbf{A} - \mathbf{A}_n|$, $\sup|\mathbf{B} - \mathbf{B}_n|$ и $\sup|\mathbf{C} - \mathbf{C}_n|$ будут сходиться к нулю. Следовательно, $L_{n'}(\mathbf{K}_{n'})$ будет слабо сходиться к $L(\mathbf{K})$. Но $L_{n'}(\mathbf{K}_{n'}) = \mathbf{F}_{n'}$ и $\mathbf{F}_{n'}$ сильно сходится к $\mathbf{F}$. Отсюда $L(\mathbf{K}) = \mathbf{F}$. На $\mathscr{H}(\ ) \cap \bar{\mathcal{U}}$ $K_{n'}{}^I{}_J$ и $K_{n'}{}^I{}_{J|a}u^a$ будут слабо сходиться к $K^I{}_J$ и $K^I{}_{J|a}u^a$, которые поэтому должны быть равны соответственно ${}_0K^I{}_J$ и ${}_1K^I{}_J$. Таким образом, $\mathbf{K}$ является решением данного уравнения с заданными начальными условиями. Согласно предложению 7.4.5, оно единственно. Поскольку каждый тензор $\mathbf{K}_n$ удовлетворяет неравенству из леммы 7.4.6, ему будет удовлетворять и $\mathbf{K}$. □

7.5. Существование и единственность решения задачи Коши для уравнений Эйнштейна для пустого пространства

Результаты предыдущего раздела мы используем теперь в задаче Коши в общей теории относительности. Сначала рассмотрим уравнения Эйнштейна для пустого пространства ($T^{ab} = 0$), а влияние материи исследуем в разд. 7.7.

Приведенные уравнения Эйнштейна

$$E^{ab}{}_{cd}(\phi^{cd}) = 8\pi T^{ab} - \left(\hat{R}^{ab} - \frac{1}{2}\hat{R}\hat{g}^{ab}\right) \tag{7.42}$$

представляют собой квазилинейные гиперболические уравнения второго порядка, т. е. имеют вид (7.20), причем коэффициенты $\mathbf{A}$, $\mathbf{B}$ и $\mathbf{C}$ являются функциями $\mathbf{K}$ и $D\mathbf{K}$ (в нашем случае $A^{ab} = g^{ab}$ является на самом деле функцией ϕ^{ab} и не зависит от $\phi^{ab}{}_{|c}$). Чтобы доказать существование решений этих уравнений, мы

поступим следующим образом. Возьмем некоторое подходящее пробное поле ϕ'^{ab} и используем его для определения значений коэффициентов **A**, **B** и **C** в операторе E. Затем будем решать (7.42) при этих значениях уже как линейное уравнение с начальными данными, определяемыми полем ϕ', и получим новое поле ϕ''^{ab}. Тем самым мы получим отображение α, переводящее $\boldsymbol{\phi}'$ в $\boldsymbol{\phi}''$, и покажем, что при определенных условиях это отображение имеет неподвижную точку (т. е. существует ϕ, для которого $\alpha(\boldsymbol{\phi}) = \boldsymbol{\phi}$). Эта неподвижная точка и будет искомым решением нашего квазилинейного уравнения.

Возьмем в качестве фоновой метрики решение уравнения Эйнштейна в пустом пространстве и выберем поверхности $\mathscr{H}(t) \cap \bar{\mathscr{U}}_+$ и $\partial\mathscr{U} \cap \bar{\mathscr{U}}_+$ так, чтобы они были пространственноподобными в метрике $\hat{\mathbf{g}}$. Тогда по лемме 7.4.1 найдется такая положительная постоянная $\tilde{Q}_a$, что коэффициенты **A**′, **B**′, **C**′, определяемые функцией ϕ', будут удовлетворять условиям (1), (2) и (4) леммы 7.4.6 с заданными значениями Q_1 и Q_3, если только

$$\| \boldsymbol{\phi}', \mathscr{U}_+ \|_{4+a} < \tilde{Q}_a. \tag{7.43}$$

В этом случае из (7.41) имеем

$$\| \boldsymbol{\phi}'', \mathscr{U}_+ \|_{4+a} \leqslant P_{7,a} \{ \| {}_0\boldsymbol{\phi}, \mathscr{H}(0) \cap \bar{\mathscr{U}} \|_{4+a} + \| {}_1\boldsymbol{\phi}, \mathscr{H}(0) \cap \bar{\mathscr{U}} \|_{3+a} \}.$$

Следовательно, отображение $\alpha: W^{4+a}(\mathscr{U}_+) \to W^{4+a}(\mathscr{U}_+)$ преобразует замкнутый шар $W(r)$ радиуса r ($r < \tilde{Q}_a$) в $W^{4+a}(\mathscr{U}_+)$ в себя при условии, что

$$\| {}_0\boldsymbol{\phi}, \mathscr{H}(0) \cap \bar{\mathscr{U}} \|_{4+a} \leqslant \tfrac{1}{2} r P_{7,a}^{-1}$$

и

$$\| {}_1\boldsymbol{\phi}, \mathscr{H}(0) \cap \bar{\mathscr{U}} \|_{3+a} \leqslant \tfrac{1}{2} r P_{7,a}^{-1}. \tag{7.44}$$

Покажем, что α имеет неподвижную точку, если выполняется (7.44) и радиус r достаточно мал.

Допустим, что $\boldsymbol{\phi}_1'$ и $\boldsymbol{\phi}_2'$ принадлежат пространству $W(r)$. Поля $\boldsymbol{\phi}_1'' = \alpha(\boldsymbol{\phi}_1')$ и $\boldsymbol{\phi}_2'' = \alpha(\boldsymbol{\phi}_2')$ удовлетворяют равенствам $E_1'(\boldsymbol{\phi}_1'') = 0$, $E_2'(\boldsymbol{\phi}_2'') = 0$, где E_1' — оператор Эйнштейна с коэффициентами $\mathbf{A}_1'$, $\mathbf{B}_1'$ и $\mathbf{C}_1'$, которые определяются функцией $\boldsymbol{\phi}_1'$. Следовательно,

$$E_1'(\boldsymbol{\phi}_1'' - \boldsymbol{\phi}_2'') = (E_1' - E_2')(\boldsymbol{\phi}_2''). \tag{7.45}$$

Поскольку коэффициенты $\mathbf{A}_1'$, $\boldsymbol{B}_1'$ и $\mathbf{C}_1'$ дифференцируемым образом зависят от $\boldsymbol{\phi}_1'$ и $\mathrm{D}\boldsymbol{\phi}_1'$ при $\boldsymbol{\phi}_1'$ из $W(r)$, то существует

такая постоянная Q_4, что на $\bar{\mathcal{U}}_+$

$$\begin{aligned}
&|\mathbf{A}'_1 - \mathbf{A}'_2| \leqslant Q_4 |\boldsymbol{\phi}'_1 - \boldsymbol{\phi}'_2|,\\
&|\mathbf{B}'_1 - \mathbf{B}'_2| \leqslant Q_4 (|\boldsymbol{\phi}'_1 - \boldsymbol{\phi}'_2| + |\mathrm{D}\boldsymbol{\phi}'_1 - \mathrm{D}\boldsymbol{\phi}'_2|),\\
&|\mathbf{C}'_1 - \mathbf{C}'_2| \leqslant Q_4 (|\boldsymbol{\phi}'_1 - \boldsymbol{\phi}'_2| + |\mathrm{D}\boldsymbol{\phi}'_1 - \mathrm{D}\boldsymbol{\phi}'_2|).
\end{aligned} \tag{7.46}$$

Поэтому, в силу лемм 7.4.1 и 7.4.6,

$$|(E'_1 - E'_2)(\boldsymbol{\phi}''_2)| \leqslant 3rQ_4 \tilde{P}_1 P^{-1}_{7,a} P_{6,a} (|\boldsymbol{\phi}'_1 - \boldsymbol{\phi}'_2| + |\mathrm{D}\boldsymbol{\phi}'_1 - \mathrm{D}\boldsymbol{\phi}'_2|).$$

Теперь применим лемму 7.4.4 к 7.4.5, чтобы получить неравенство

$$\|\boldsymbol{\phi}''_1 - \boldsymbol{\phi}''_2,\ \bar{\mathcal{U}}_+\|_1 \leqslant rQ_5 \|\boldsymbol{\phi}'_1 - \boldsymbol{\phi}'_2,\ \bar{\mathcal{U}}_+\|_1, \tag{7.47}$$

где Q_5 — некоторая постоянная, не зависящая от r. Таким образом, для достаточно малых r отображение α будет стягивающим по норме $\|\ \|_1$ (т. е. $\|\alpha(\boldsymbol{\phi}_1) - \alpha(\boldsymbol{\phi}_2)\|_1 < \|\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2\|_1$) и последовательность $\alpha^n(\boldsymbol{\phi}'_1)$ будет сильно сходиться в $W^1(\bar{\mathcal{U}}_+)$ к некоторому полю $\boldsymbol{\phi}$. Но по теореме Рисса из $\alpha^n(\boldsymbol{\phi}'_1)$ можно выбрать подпоследовательность, слабо сходящуюся к некоторому полю $\tilde{\boldsymbol{\phi}} \in W(r)$. Следовательно, поле $\boldsymbol{\phi}$ должно совпадать с $\tilde{\boldsymbol{\phi}}$ и принадлежать пространству $W(r)$. Поэтому отображение $\alpha(\boldsymbol{\phi})$ будет определенным и для него

$$\|\alpha(\boldsymbol{\phi}) - \alpha^{n+1}(\phi'_1),\ \bar{\mathcal{U}}_+\|_1 \leqslant rQ_5 \|\boldsymbol{\phi} - \alpha^n(\phi'_1),\ \bar{\mathcal{U}}_+\|_1.$$

При $n \to \infty$ правая часть стремится к нулю. Из этого следует, что $\|\alpha(\boldsymbol{\phi}) - \boldsymbol{\phi},\ \bar{\mathcal{U}}_+\|_1 = 0$ и потому $\alpha(\boldsymbol{\phi}) = \boldsymbol{\phi}$. Поскольку α — стягивающее отображение, эта неподвижная точка единственна в $W(r)$. Следовательно, мы доказали

Предложение 7.5.1

Если $\hat{\mathbf{g}}$ — решение уравнений Эйнштейна в пустом пространстве, то приведенные уравнения Эйнштейна в пустом пространстве имеют решение $\boldsymbol{\phi} \in W^{4+a}(\bar{\mathcal{U}}_+)$ при достаточно малых $\|_0\boldsymbol{\phi},\ \mathscr{H}(0) \cap \bar{\mathcal{U}}\|_{4+a}$ и $\|_1\boldsymbol{\phi},\ \mathscr{H}(0) \cap \bar{\mathcal{U}}\|_{3+a}$. Норма $\|\boldsymbol{\phi},\ \mathscr{H}(0) \cap \bar{\mathcal{U}}_+\|_{4+a}$ ограничена, и, следовательно, поле $\boldsymbol{\phi}$ по крайней мере класса $C^{(2+a)-}$. □

Локально это решение единственно даже среди решений, которые не принадлежат пространству $W^4(\bar{\mathcal{U}}_+)$.

Предложение 7.5.2

Пусть $\tilde{\boldsymbol{\phi}}$ является C^{1-}-решением приведенных уравнений Эйнштейна в пустом пространстве с теми же начальными данными на открытом множестве $\mathscr{V} \subset \mathscr{H}(0) \cap \mathcal{U}$. Тогда $\tilde{\boldsymbol{\phi}} = \boldsymbol{\phi}$ в некоторой окрестности множества $\mathscr{V}$ в $\bar{\mathcal{U}}_+$.

Ввиду непрерывности $\tilde{\boldsymbol{\phi}}$ в $\mathscr{U}$ найдется окрестность $\mathscr{U}'$ множества $\mathscr{V}$, в которой для **A**, **B** и **C** выполняются условия леммы 7.4.4. Как прежде, имеем

$$\tilde{E}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}) = -(\tilde{E} - E)(\boldsymbol{\phi}). \tag{7.48}$$

Аналогично, будет существовать постоянная Q_6, обеспечивающая выполнение неравенства

$$\| (\tilde{E} - E)(\boldsymbol{\phi}),\ \mathscr{H}(t) \cap \mathscr{U}'_+ \|_0 \leqslant Q_6 \| \tilde{\boldsymbol{\phi}} - \boldsymbol{\phi},\ \mathscr{H}(t) \cap \mathscr{U}'_+ \|_1.$$

Применяя к (7.48) лемму 7.4.4, получим неравенство вида

$$\frac{dx}{dt} \leqslant Q_7 x,$$

где

$$x = \int_0^t \| \tilde{\boldsymbol{\phi}} - \boldsymbol{\phi},\ \mathscr{H}(t') \cap \mathscr{U}'_+ \|_1 \, dt'.$$

Отсюда $\tilde{\boldsymbol{\phi}} = \boldsymbol{\phi}$ на $\bar{\mathscr{U}}'_+$. □

Из предложения 7.5.1 видно, что при наложении достаточно малых возмущений на начальные данные для решения уравнений Эйнштейна в пустом пространстве мы снова получаем некоторое решение в области $\mathscr{U}_+$. Однако в действительности требуется доказать, что решение существует для любых начальных данных h^{ab} и χ^{ab}, которые удовлетворяют уравнениям связи на трехмерном многообразии $\mathscr{S}$. Для этого мы поступим следующим образом. Будем считать, что многообразием $\mathscr{M}$ является R^4, что **e** — евклидова метрика, а $\hat{\mathbf{g}}$ — плоская метрика, т. е. метрика Минковского (она является решением уравнений Эйнштейна в пустом пространстве). Область $\mathscr{U}$ выберем так, чтобы в обычных координатах Минковского x^1, x^2, x^3 и x^4 ($x^4 = t$) поверхность $\partial\mathscr{U} \cap \bar{\mathscr{U}}_+$ была пространственноподобна, а область $\mathscr{H}(0) \cap \bar{\mathscr{U}}$ состояла из точек, для которых $(x^1)^2 + (x^2)^2 + (x^3)^2 \leqslant 1$, $x^4 = 0$. Теперь мы будем руководствоваться идеей, что любая метрика выглядит почти плоской, если ее рассматривать в достаточно малых масштабах. Поэтому при отображении достаточно малой области поверхности $\mathscr{S}$ на $\mathscr{H}(0) \cap \bar{\mathscr{U}}$ можно учесть предложение 7.5.1 и получить решение в $\mathscr{U}_+$. Затем можно повторить эту процедуру для других частей $\mathscr{S}$ и объединить полученные решения так, чтобы образовалось многообразие $\mathscr{M}$ с метрикой **g**, которое и будет развитием комплекта начальных данных $(\mathscr{S}, \boldsymbol{\omega})$.

Пусть $\mathscr{V}_1$ — координатная окрестность в $\mathscr{S}$ с координатами y^1, y^2, y^3, в которой в начале координат p координатные компоненты h^{ab} равны δ^{ab}. Пусть $\mathscr{V}_1(f_1)$ — открытый шар координатного радиуса f_1 с центром в p. Зададим отображение θ: $\mathscr{V}_1(f_1) \to \mathscr{U}$

соотношениями $x^i = f_1^{-1} y^i$ $(i = 1, 2, 3)$, $x^4 = 0$. По обычному закону преобразования базиса компоненты $\theta_* h^{ab}$ и $\theta_* \chi^{ab}$ в координатах $\{x\}$ равны компонентам h^{ab} и χ^{ab} в координатах $\{y\}$, умноженным на f_1^{-2}. Введем в $\mathscr{V}_1$ новые поля h'^{ab} и χ'^{ab}: $h'^{ab} = f_1^2 h^{ab}$ и $\chi'^{ab} = f_1^3 \chi^{ab}$. Тогда, поскольку h на $\mathscr{S}$ непрерывна (фактически принадлежит классу C^{2+a}), выбором достаточно малого f_1 можно величины $g'^{ab} - \hat{g}^{ab}$ и $g'^{ab}{}_{|c} u^c$ сделать сколь угодно малыми на $\mathscr{H}(0) \cap \mathscr{U}$, где g'^{ab} и $g'^{ab}{}_{|c} u_c$ получаются из h'^{ab} и χ'^{ab}, как в разд. 7.3. При этом производные g'^{ab} и $g'^{ab}{}_{|c} u^c$ на поверхности $\mathscr{H}(0)$ тоже убывают по мере убывания f_1. Следовательно, $\| {}_0\boldsymbol{\phi}', \mathscr{H}(0) \cap \bar{\mathscr{U}} \|_{4+a}$ и $\| {}_1\boldsymbol{\phi}', \mathscr{H}(0) \cap \bar{\mathscr{U}} \|_{3+a}$ могут быть приведены к настолько малым значениям, что можно будет использовать предложение 7.5.1 и решение для $\boldsymbol{\phi}'$, полученное в $\mathscr{U}_+$. Тогда $g_1^{ab} = f_1^{-2} g'^{ab}$ будет решением приведенных уравнений Эйнштейна с начальными данными, определяемыми тензорами h^{ab} и χ^{ab}. Аналогично можно получить решение $\mathscr{U}_-$, той части $\mathscr{U}$, в которой $t \leqslant 0$.

Теперь мы можем покрыть $\mathscr{S}$ координатными окрестностями $\mathscr{V}_\alpha(f_\alpha)$ вида $\mathscr{V}_1(f_1)$, отобразив их вложениями θ_α на окрестности $\mathscr{U}_\alpha$ вида $\mathscr{U}$, и получить решения ${}_\alpha g^{ab}$ в $\mathscr{U}_\alpha$. Тогда задача сведется к тому, чтобы отождествить соответствующие точки в пересечениях окрестностей и тем самым превратить набор окрестностей $\mathscr{U}_\alpha$ в многообразие $\mathscr{M}$ с метрикой $\mathbf{g}$. Это можно сделать, используя гармонические калибровочные условия

$$\phi^{bc}{}_{|c} = g^{bc}{}_{|c} - \frac{1}{2} g^{bc} g_{de} g^{dc}{}_{|c} = 0. \tag{7.49}$$

По определению $\delta\Gamma^a{}_{bc}$ (7.3) оно эквивалентно условию $g^{de}\delta\Gamma^b{}_{de}=0$. Поэтому для любой функции z

$$z_{;ab} g^{ab} = z_{|ab} g^{ab} - \delta\Gamma^c{}_{ab} z_{|c} g^{ab} = z_{|ab} g^{ab}. \tag{7.50}$$

Если фоновой метрикой является метрика Минковского, а z — одна из координат Минковского x^1, x^2, x^3, x^4, то правая часть (7.50) обращается в нуль. Допустим теперь, что на некотором многообразии $\mathscr{M}$ мы имеем произвольную лоренцеву W^{4+a}-метрику $\mathbf{g}$. В некоторой окрестности $\mathscr{Y} \subset \mathscr{M}$ мы можем получить такие четыре решения z^1, z^2, z^3, z^4 линейного уравнения

$$z_{;ab} g^{ab} = 0, \tag{7.51}$$

градиенты которых линейно независимы в каждой точке $\mathscr{Y}$. Тогда мы можем задать диффеоморфизм $\mu: \mathscr{Y} \to \hat{\mathscr{M}}$ вида $x^a = z^a$ $(a = 1, 2, 3, 4)$. Этот диффеоморфизм будет обладать тем свойством, что метрика $\mu_* g^{ab}$ на $\hat{\mathscr{M}}$ будет удовлетворять гармо-

ническому калибровочному условию относительно метрики Минковского $\hat{\mathbf{g}}$ на $\hat{\mathscr{M}}$. Таким образом, если метрика $\mathbf{g}$ является решением уравнений Эйнштейна на $\mathscr{M}$, метрика $\mu_*\mathbf{g}$ будет решением приведенных уравнений Эйнштейна с фоновой метрикой $\hat{\mathbf{g}}$.

Процедура отождествления точек в пересечении двух окрестностей $\mathscr{U}_\alpha$ и $\mathscr{U}_\beta$ сводится таким образом к решению (7.51) в $\mathscr{U}_\alpha$ для координат $x_\beta{}^1$, $x_\beta{}^2$, $x_\beta{}^3$, $x_\beta{}^4$ с использованием начальных данных для $x_\beta{}^a$ и $x_\beta{}^a{}_{|b}u^b$, определяемых пересечением координатных окрестностей $\mathscr{V}_\alpha$ и $\mathscr{V}_\beta$ на $\mathscr{S}$. В действительности $x_\beta{}^i{}_{|a}u^a = 0$ $(i = 1, 2, 3)$ и $x_\beta{}^4{}_{|a}u^a = 1$, где $u^a = \partial/\partial x_\alpha{}^a$ — единичный вектор в $\mathscr{U}_\alpha$, ортогональный к $\mathscr{H}(0)$ в метрике $\hat{\mathbf{g}}$. Таким образом, $x_\beta{}^4 = x_\alpha{}^4$, хотя $x_\beta{}^i$ вообще не будет совпадать с $x_\alpha{}^i$. По предложению 7.4.7 координаты $x_\beta{}^a$ будут в $\mathscr{U}_\alpha$ $C^{(2+a)}$-функциями. [В предложении 7.4.7 фоновая метрика, относительно которой берутся ковариантные производные, должна была принадлежать классу $C^{(5+a)-}$. Поэтому предложение 7.4.7 нельзя было применить к (7.51), где ковариантные производные берутся относительно метрики $\mathbf{g}$, принадлежащей только пространству W^{4+a}. Однако мы можем ввести фоновую C^{5+a} метрику $\tilde{\mathbf{g}}$ и представить (7.51) в виде

$$z_{\|ab}g^{ab} + z_{\|a}B^a = 0,$$

где $\|$ означает ковариантное дифференцирование относительно $\tilde{\mathbf{g}}$. После этого можно воспользоваться предложением (7.4.7).]

Поскольку градиенты $x_\beta{}^a$ на $\mathscr{H}(0) \cap \mathscr{U}_\alpha$ линейно независимы, они будут линейно независимы и в некоторой окрестности $\mathscr{U}''_\alpha$ поверхности $\mathscr{H}(0)$ в $\mathscr{U}_\alpha$. Метрика $\mu_* g_\alpha{}^{ab}$ на $\mu(\mathscr{U}''_\alpha)$ в $\mathscr{U}_\beta$ будет по меньшей мере класса C^{1-}. Ввиду того что она подчиняется приведенным уравнениям Эйнштейна и имеет те же начальные данные на $\theta(\mathscr{V}_\alpha \cap \mathscr{V}_\beta)$, она должна совпадать с $\mathbf{g}_\beta$ в некоторой окрестности $\mathscr{U}_\beta' \subset \mathscr{U}$ области $\theta_\beta(\mathscr{V}_\alpha \cap \mathscr{V}_\beta)$. Отсюда видно, что можно соединить $\mathscr{U}'''_\alpha$ и $\mathscr{U}'_\beta$ и получить развитие области $\mathscr{V}_\alpha \cup \mathscr{V}_\beta$. Выбрав покрытие $\{\mathscr{V}_\alpha\}$ поверхности $\mathscr{S}$ локально конечным, мы можем поступить подобным же образом, соединить подмножества других окрестностей $\{\mathscr{U}_\alpha\}$ и получить развитие $\mathscr{S}$, т. е. построить такие многообразия $\mathscr{M}$ с метрикой $\mathbf{g}$ и вложение $\theta: \mathscr{S} \to \mathscr{M}$, что метрика $\mathbf{g}$ удовлетворяет уравнениям Эйнштейна в пустом пространстве и согласуется с имеющимися начальными данными $\boldsymbol{\omega}$ на поверхности $\theta(\mathscr{S})$, которая является поверхностью Коши для $\mathscr{M}$. Если $(\mathscr{M}', \mathbf{g}')$ — другое развитие комплекта $(\mathscr{S}, \boldsymbol{\omega})$, то аналогичная процедура устанавливает диффеоморфизм μ между некоторой окрестностью поверхности $\theta'(\mathscr{S}')$

в $\mathscr{M}'$ и некоторой окрестностью поверхности $\theta(\mathscr{S})$ в $\mathscr{M}$, для которого $\mu_* g'^{ab} = g^{ab}$. Следовательно, нами доказана

Теорема о локальном решении задачи Коши

Если $h^{ab} \in W^{4+a}(\mathscr{S})$ и $\chi^{ab} \in W^{3+a}(\mathscr{S})$ удовлетворяют уравнениям связи для пустого пространства, то существуют решения $(\mathscr{M}, \mathbf{g})$ уравнений Эйнштейна в пустом пространстве, такие, что $\mathbf{g} \in W^{4+a}(\mathscr{M})$ и $\mathbf{g} \in W^{4+a}$ на любой гладкой пространственноподобной поверхности $\mathscr{H}$. Эти решения локально единственны в следующем смысле: если $(\mathscr{M}', \mathbf{g}')$ — другое развитие того же комплекта начальных данных $(\mathscr{S}, \boldsymbol{\omega})$, принадлежащее пространству W^{4+a}, оба развития $(\mathscr{M}, \mathbf{g})$ и $(\mathscr{M}', \mathbf{g}')$ являются расширениями некоторого общего развития комплекта $(\mathscr{S}, \boldsymbol{\omega})$.

Утверждение, что $\mathbf{g} \in W^{4+a}(\mathscr{H})$, следует из леммы 7.4.6 и из того, что поверхности постоянного f можно выбрать произвольно. □

7.6. Максимальное развитие и устойчивость

Итак, мы показали следующее: если начальные данные удовлетворяют уравнениям связи для пустого пространства и при этом можно найти некоторое развитие начальных данных (т. е. можно распространить решение на некоторое расстояние в будущее и прошлое от начальной поверхности, если эти данные удовлетворяют уравнениям связи), то это развитие, вообще говоря, можно продолжить дальше в будущее и прошлое, что даст более протяженное развитие комплекта начальных данных $(\mathscr{S}, \boldsymbol{\omega})$. Однако, следуя ходу рассуждений в [33], мы докажем существование единственного (с точностью до диффеоморфизма) развития $(\mathscr{M}, \mathbf{g})$ комплекта $(\mathscr{S}, \boldsymbol{\omega})$, являющегося расширением любого другого развития этого комплекта.

Напомним, что $(\mathscr{M}_1, \mathbf{g}_1)$ — это расширение пространства $(\mathscr{M}_2, \mathbf{g}_2)$, если существует вложение $\mu: \mathscr{M}_2 \to \mathscr{M}_1$, при котором $\mu_* \mathbf{g}_2 = \mathbf{g}_1$ и $\theta_1^{-1} \mu \theta_2$ суть тождественное отображение на $\mathscr{S}$. Если даны точка $q \in \mathscr{S}$ и расстояние s, мы можем однозначно задать точки $p_1 \in \mathscr{M}_1$ и $p_2 \in \mathscr{M}_2$ путем смещения q на расстояние s вдоль геодезических, ортогональных к $\theta_1(\mathscr{S})$ и $\theta_2(\mathscr{S})$ и проходящих через $\theta_1(q)$ и $\theta_2(q)$ соответственно. Поскольку $\mu(p_2)$ должно совпадать с p_1, вложение μ будет единственным. Поэтому можно частично упорядочить множество всех развитий комплекта $(\mathscr{S}, \boldsymbol{\omega})$, положив $(\mathscr{M}_2, \mathbf{g}_2) \leqslant (\mathscr{M}_1, \mathbf{g}_1)$, если $(\mathscr{M}_1, \mathbf{g}_1)$ является расширением $(\mathscr{M}_2, \mathbf{g}_2)$. Пусть теперь $\{\mathscr{M}_\alpha, \mathbf{g}_\alpha\}$ — вполне упорядоченное множество (множество $\mathscr{A}$ называется вполне упорядоченным, если для каждой пары различных элементов $a, b \in \mathscr{A}$ или $a \leqslant b$ или $b \leqslant a$) развитий комплекта $(\mathscr{S}, \boldsymbol{\omega})$; тогда можно построить многообразие $\mathscr{M}'$ как объединение всех $\mathscr{M}_\alpha$, причем при

$(\mathscr{M}_\alpha, \mathbf{g}_\alpha) \leqslant (\mathscr{M}_\beta, \mathbf{g}_\beta)$ каждая точка $p_\alpha \in \mathscr{M}_\alpha$ отождествляется с $\mu_{\alpha\beta}(p_\alpha) \in \mathscr{M}_\beta$, где $\mu_{\alpha\beta}: \mathscr{M}_\beta \to \mathscr{M}_\beta$ — вложение. Многообразие $\mathscr{M}'$ будет иметь индуцированную метрику $\mathbf{g}'$, равную $\mu_{\alpha *}\mathbf{g}_\alpha$ на каждой $\mu_\alpha(\mathscr{M}_\alpha)$, где $\mu_\alpha: \mathscr{M}_\alpha \to \mathscr{M}'$ — естественное вложение. Ясно, что $(\mathscr{M}', \mathbf{g}')$ тоже будет развитием комплекта $(\mathscr{S}, \boldsymbol{\omega})$; поэтому любое вполне упорядоченное множество имеет верхнюю грань, и, таким образом, по лемме Цорна (см., например, [88], p. 33) существует максимальное развитие $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$ комплекта $(\mathscr{S}, \boldsymbol{\omega})$, расширением которого является только оно само.

Теперь мы покажем, что $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$ является расширением любого развития комплекта $(\mathscr{S}, \boldsymbol{\omega})$. Допустим, что $(\mathscr{M}', \mathbf{g}')$ — некоторое другое развитие $(\mathscr{S}, \boldsymbol{\omega})$. Согласно локальной теореме Коши, существуют развития $(\mathscr{S}, \boldsymbol{\omega})$, для которых как $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$, так и $(\mathscr{M}', \mathbf{g}')$ являются расширениями. Множество всех общих для них развитий также частично упорядочено, и снова по лемме Цорна найдется максимальное расширение $(\mathscr{M}'', \mathbf{g}'')$ с вложениями $\tilde{\mu}: \mathscr{M}'' \to \tilde{\mathscr{M}}$ и $\mu': \mathscr{M}'' \to \mathscr{M}'$ и т. д. Пусть $\mathscr{M}^+$ будет объединением $\tilde{\mathscr{M}}$, $\mathscr{M}'$ и $\mathscr{M}''$, в котором каждая точка $p'' \in \mathscr{M}''$ отождествляется с $\tilde{\mu}(p'') \in \tilde{\mathscr{M}}$ и с $\mu'(p'') \in \mathscr{M}'$. Если мы сможем показать, что многообразие $\mathscr{M}^+$ хаусдорфово, то тем самым докажем, что пара $(\mathscr{M}^+, \mathbf{g}^+)$ является развитием комплекта $(\mathscr{S}, \boldsymbol{\omega})$. Она будет расширением как $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$, так и $(\mathscr{M}', g')$. Однако единственным расширением пространства $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$ является оно само и, таким образом, $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$ должно совпасть с $(\mathscr{M}^+, \mathbf{g}^+)$ и быть расширением пространства $(\mathscr{M}', \mathbf{g}')$.

Допустим, что $\mathscr{M}^+$ — не хаусдорфово многообразие. Тогда существуют точки $\tilde{p} \in (\tilde{\mu}(\mathscr{M}''))^\bullet \subset \tilde{\mathscr{M}}$ и $p' \in (\mu'(\mathscr{M}''))^\bullet \subset \mathscr{M}'$, для которых каждая окрестность $\mathscr{U}$ точки $\tilde{p}$ обладает тем свойством, что $\overline{\mu'(\tilde{\mu}^{-1}(\mathscr{U}))}$ содержит p'. Далее, поскольку пространство $(\mathscr{M}'', \mathbf{g}'')$ является развитием, оно, как и его образ $\tilde{\mu}(\mathscr{M}'')$ в $\tilde{\mathscr{M}}$, будет глобально гиперболическим. Поэтому граница множества $\tilde{\mu}(\mathscr{M}'')$ в $\tilde{\mathscr{M}}$ должна быть ахрональной. Пусть γ — времениподобная кривая в $\tilde{\mathscr{M}}$ с конечной точкой будущего p'. Тогда p' должна быть предельной точкой в $\mathscr{M}'$ кривой $\mu'\tilde{\mu}^{-1}(\gamma)$. На самом деле она должна быть конечной точкой будущего, так как в $(\mathscr{M}', \mathbf{g}')$ выполняется сильная причинность. Следовательно, для данной точки $\tilde{p}$ точка p' единственна. Далее, в силу непрерывности векторы в p' можно однозначно связать с векторами в $\tilde{p}$. Таким образом мы можем подобрать такие нормальные координатные окрестности $\tilde{\mathscr{U}}$ для $\tilde{p}$ в $\tilde{\mathscr{M}}$ и $\mathscr{U}'$ для p' в $\mathscr{M}'$, точки $\tilde{\mathscr{U}} \cap \tilde{\mu}(\mathscr{M}'')$ которых при отображении $\mu'(\tilde{\mu}^{-1})$ будут отображаться в точки $\mathscr{U}' \cap \mu'(\mathscr{M}'')$ с теми же значениями координат.

Отсюда следует, что множество $\mathcal{F}$ всех „нехаусдорфовых“ точек из $(\tilde{\mu}(\mathcal{M}''))$ открыто в $(\tilde{\mu}(\mathcal{M}''))^{\bullet}$. Мы предположим теперь, что $\mathcal{F}$ — непустое множество, и придем затем к противоречию.

Пусть $\tilde{\lambda}$ — направленная в прошлое изотропная геодезическая в $\tilde{\mathcal{M}}$, проходящая через $\tilde{p} \in \mathcal{F}$; тогда, используя возможность связать направления в $\tilde{p}$ с направлениями в p', можно провести в $\mathcal{M}'$ в соответствующем направлении изотропную геодезическую в прошлое λ', проходящую через p'. Каждой точке множества $\tilde{\lambda} \cap (\tilde{\mu}(\mathcal{M}''))^{\bullet}$ будет соответствовать точка из $\lambda' \cap (\mu'(\mathcal{M}''))^{\bullet}$, и, следовательно, каждая точка, принадлежащая $\tilde{\lambda} \cap (\tilde{\mu}(\mathcal{M}''))^{\bullet}$, будет точкой из $\mathcal{F}$. Поскольку $\tilde{\theta}(\mathcal{S})$ является поверхностью Коши для $\tilde{\mathcal{M}}$, кривая $\tilde{\lambda}$ должна покинуть $(\tilde{\mu}(\mathcal{M}''))^{\bullet}$ в некоторой точке $\tilde{q}$. В некоторой окрестности $\tilde{q}$ найдется точка $\tilde{r} \in \mathcal{F}$, через которую можно провести пространственноподобную поверхность $\tilde{\mathcal{H}}$, обладающую тем свойством, что $(\tilde{\mathcal{H}} - \tilde{r}) \subset \tilde{\mu}(\mathcal{M}'')$. В $\mathcal{M}'$ существует соответствующая пространственноподобная поверхность $\mathcal{H}' = (\mu' \tilde{\mu}^{-1}(\tilde{\mathcal{H}} - \tilde{r})) \cup r'$, проходящая через r'. Поверхности $\tilde{\mathcal{H}}$ и $\mathcal{H}'$ можно рассматривать как образы трехмерного многообразия $\mathcal{H}$ при вложениях $\tilde{\psi}: \mathcal{H} \to \tilde{\mathcal{M}}$ и $\psi: \mathcal{H} \to \mathcal{M}'$, для которых $\tilde{\psi}^{-1} \tilde{\mu} \mu'^{-1} \psi'$ является тождественным отображением на $\mathcal{H} - \tilde{\psi}^{-1}(p)$. Индуцированные на $\mathcal{H}$ метрики $\tilde{\psi}_*(\tilde{\mathbf{g}})$ и $\psi'_*(\mathbf{g}')$ согласуются одна с другой, поскольку $\tilde{\mathcal{H}} - \tilde{p}$ и $\mathcal{H}' - p'$ изометричны. Согласно локальной теореме Коши, они принадлежат пространству $W^{4+a}(\mathcal{H})$. Аналогично, вторые фундаментальные формы также согласуются и принадлежат $W^{3+a}(\mathcal{H})$. Окрестности поверхностей $\tilde{\mathcal{H}}$ в $\tilde{\mathcal{M}}$ и $\mathcal{H}'$ в $\mathcal{M}'$ будут W^{4+a}-развитиями $\mathcal{H}$. По локальной теореме Коши они должны быть расширениями одного и того же развития $(\mathcal{M}^*, \mathbf{g}^*)$. Присоединяя $(\mathcal{M}^*, \mathbf{g}^*)$ к $(\mathcal{M}'', \mathbf{g}'')$, мы получили бы большее развитие комплекта начальных данных $(\mathcal{S}, \boldsymbol{\omega})$, причем $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ и $(\mathcal{M}', \mathbf{g}')$ были бы расширениями этого развития. Однако это невозможно, ибо $(\mathcal{M}'', \mathbf{g}'')$ уже является наибольшим общим развитием. Из этого следует, что многообразие $\mathcal{M}^+$ должно быть хаусдорфовым, а следовательно, и пространство $(\tilde{\mathcal{M}}, \mathbf{g})$ должно быть расширением пространства $(\mathcal{M}', \mathbf{g}')$.

Итак, нами доказана

Теорема о глобальном решении задачи Коши

Если $h^{ab} \in W^{4+a}(\mathcal{S})$ и $\chi^{ab} \in W^{3+a}(\mathcal{S})$ удовлетворяют уравнениям связи для пустого пространства, то для уравнений Эйнштейна в пустом пространстве существует максимальное развитие $(\mathcal{M}, \mathbf{g})$ с $g \in W^{4+a}(\mathcal{M})$ и $\mathbf{g} \in W^{4+a}(\mathcal{H})$ комплекта начальных

данных для любой гладкой пространственноподобной поверхности $\mathscr{H}$. Это развитие является расширением любого другого такого развития.

Пока мы доказали лишь то, что полученное нами развитие — максимальное среди развитий, принадлежащих пространству W^{4+a} (W^{4+a}-развитий). Если $a>0$, то существуют также W^{4+a-1}-, W^{4+a-2}-, ..., W^4-развития, которые являются расширениями полученного развития. Однако Шоке-Брюа показал [30], что все эти расширения должны совпасть с W^4-развитием. Это происходит вследствие того, что мы можем дифференцировать приведенные уравнения Эйнштейна и затем рассматривать их как *линейные* уравнения на W^4-развитии для первых производных g^{ab}. Затем, пользуясь предложением 7.4.7, можно показать, что $g^{ab} \in W^5$ на W^4-развитии, если начальные данные принадлежат пространству W^5. Следуя этому рецепту, мы можем показать, что при начальных данных класса C^∞ будет существовать C^∞-развитие, которое фактически совпадает с W^4-развитием.

Мы доказали существование и единственность максимальных развитий только для случая, когда метрика не «хуже», чем класса W^4. В действительности можно доказать существование развитий для начальных данных из W^3, но в этом случае мы не сможем доказать единственность. Не исключено, что максимальное W^4-развитие может быть расширено или так, чтобы метрика перестала принадлежать пространству W^4, или так, чтобы $\theta(\mathscr{S})$ перестала быть поверхностью Коши. В последнем случае появится горизонт Коши; примеры этого были даны в гл. 6. С другой стороны, возможно присутствие сингулярности того или иного рода, и тогда развитие нельзя расширить так, чтобы порядок дифференцируемости метрики был достаточен для ее физической интерпретации. Действительно, как мы увидим в следующей главе (теорема 4), если множество $\mathscr{S}$ компактно и $\chi^{ab}h_{ab}<0$ всюду на $\mathscr{S}$, развитие нельзя расширить так, чтобы оно было геодезически полным и обладало C^{2-}-метрикой (т. е. локально ограниченной кривизной).

Мы показали, что существует отображение из пространства пар тензоров (h^{ab}, χ^{ab}) на $\mathscr{S}$, удовлетворяющих уравнениям связи, в пространство классов эквивалентности метрики $\mathbf{g}$ на многообразии $\mathscr{M}$, которое, согласно предложению 6.6.8, диффеоморфно $\mathscr{S} \times R^1$. Если две пары (h^{ab}, χ^{ab}) и (h'^{ab}, χ'^{ab}) эквивалентны относительно диффеоморфизма $\lambda: \mathscr{S} \to \mathscr{S}$ (т. е. $\lambda_* h^{ab} = h'^{ab}$ и $\lambda_* \chi^{ab} = \chi'^{ab}$), то они порождают эквивалентные метрики $\mathbf{g}$. Таким образом, мы имеем отображение классов эквивалентности пар (h^{ab}, χ^{ab}) в классы эквивалентности метрик g. h^{ab} и χ^{ab} вместе имеют 12 независимых компонент. Уравнения связи накладывают на них четыре соотношения, а эквивалентность

относительно диффеоморфизмов можно использовать для исключения еще трех произвольных функций; остается пять независимых функций. Можно считать, что одна из этих функций характеризует положение поверхности $\theta(\mathscr{S})$ в развитии $(\mathscr{M}, \mathbf{g})$. Следовательно, максимальные развития, удовлетворяющие уравнениям Эйнштейна, в пустом пространстве определяются четырьмя функциями трех переменных.

Было бы желательно показать, что отображение из классов эквивалентности пар (h^{ab}, χ^{ab}) в классы эквивалентности метрик $\mathbf{g}$ в каком-то смысле непрерывно. Подходящей для этого топологией на классах эквивалентности является компактная открытая W^r-топология (ср. с разд. 6.4). Пусть $\hat{\mathbf{g}}$ — лоренцева C^r-метрика на $\mathscr{M}$, а $\mathscr{U}$ — открытое множество с компактным замыканием. Пусть V — открытое множество в $W^r(\mathscr{U})$, и пусть $O(\mathscr{U}, V)$ — множество всех таких лоренцевых метрик на $\mathscr{M}$, сужения которых на $\mathscr{U}$ принадлежат V. Открытые множества компактной открытой W^r-топологии в пространстве $\mathscr{L}_r(\mathscr{M})$ всех лоренцевых W^r-метрик на $\mathscr{M}$ определяются как объединения и конечные пересечения множеств вида $O(U, V)$. Тогда топология пространства $\mathscr{L}_r^*(\mathscr{M})$ классов эквивалентности W^r-метрик на $\mathscr{M}$ индуцируется проекцией

$$\pi: \mathscr{L}_r(\mathscr{M}) \to \mathscr{L}_r^*(\mathscr{M}),$$

которая ставит в соответствие метрике ее класс эквивалентности [т. е. открытые множества в $\mathscr{L}_r^*(\mathscr{M})$ имеют вид $\pi(Q)$, где Q — открытое множество в $\mathscr{L}_r(\mathscr{M})$]. Аналогично компактная открытая W^r-топология в пространстве $\Omega_r(\mathscr{S})$ всех пар (h^{ab}, χ^{ab}), удовлетворяющих уравнениям связи, определяется множествами вида $O(\mathscr{U}, V, V')$, состоящими из пар, для которых $h^{ab} \in V$ и $\chi^{ab} \in V'$, где V и V' — открытые множества соответственно в $W^r(\mathscr{S})$ и $W^{r-1}(\mathscr{S})$. C^∞-метрики на $\mathscr{M}$ образуют подпространство $\mathscr{L}_\infty(\mathscr{M})$ пространства $\mathscr{L}(\mathscr{M})$ всех лоренцевых метрик на $\mathscr{M}$. Так как C^∞-метрика принадлежит пространству W^r при любом r, $\mathscr{L}_\infty(\mathscr{M})$ обладает рассматриваемой W^r-топологией. Мы можем ввести в $\mathscr{L}_\infty(\mathscr{M})$ топологию, задавая ее всеми открытыми множествами в $\mathscr{L}_\infty(\mathscr{M})$ в W^r-топологиях при каждом r. Аналогично определяется C^∞-топология на $\mathscr{L}_\infty^*(\mathscr{M})$ и на $\Omega_\infty(\mathscr{S})$.

Желательно также показать, что отображение Δ_r из пространства $\Omega_r^*(\mathscr{S})$ классов эквивалентности пар (h^{ab}, χ^{ab}) в пространство $\mathscr{L}_r^*(\mathscr{M})$ классов эквивалентности метрик непрерывно по открытой компактной W^r-топологии в обоих пространствах. Другими словами, предположим, что имеются начальные данные $h^{ab} \in W^r(\mathscr{S})$ и $\chi^{ab} \in W^{r-1}(\mathscr{S})$, которые на $\mathscr{M}$ дают решение $\mathbf{g} \in W^r(\mathscr{M})$. Желательно показать, что для $\varepsilon > 0$ и для области

$\mathscr{V} \subset \mathscr{M}$ с компактным замыканием существуют $\delta > 0$ и некоторая область $\mathscr{Y} \subset \mathscr{S}$ с компактным замыканием, такие, что $\| \mathbf{g}' - \mathbf{g},\ \mathscr{V} \|_r < \varepsilon$ при всех начальных данных $(h'^{ab},\ \chi'^{ab})$, удовлетворяющих условиям $\| \mathbf{h}' - \mathbf{h},\ \mathscr{Y} \|_r < \frac{1}{2}\delta$ и $\| \chi' - \chi,\ \mathscr{Y} \|_{r-1} < \frac{1}{2}\delta$. Это утверждение, возможно, справедливо, но доказать мы его не в состоянии. Что мы действительно можем доказать — это справедливость данного утверждения для метрики, принадлежащей классу $C^{(r+1)-}$. Оно сразу следует из предложения 7.5.1, если взять $\mathbf{g}$ в качестве фоновой метрики и некоторую подходящую окрестность из $J^-(\overline{\mathscr{V}}) \cap J^+(\theta(\mathscr{S}))$ в качестве $\mathscr{U}$. Если мы обратимся к лемме 7.4.6, то обнаружим, что в действительности можно ослабить требования к фоновой метрике: вместо $C^{(r+1)-}$-метрики можно взять W^{r+1}-метрику, но W^r-метрика уже недопустима ввиду присутствия производных $r-1$ порядка тензора кривизны для фоновой метрики. (Под словами „фоновая W^{r+1}-метрика“ мы имеем в виду принадлежность этой метрики пространству W^{r+1} относительно какой-то другой фоновой C^{r+1}-метрики.) Итак, отображение $\Delta_r : \Omega_r^*(\mathscr{S}) \to \mathscr{L}_r^*(\mathscr{M})$ из классов эквивалентности начальных данных в классы эквивалентности метрик будет непрерывным в открытой компактной W^r-топологии для всякой W^{r+1}-метрики. Хотя W^{r+1}-метрики образуют среди W^r-метрик плотное множество, возможно, что отображение Δ_r не будет непрерывно на W^r-метрике, если она не является одновременно и W^{r+1}-метрикой. Однако $\infty + 1 = \infty$ и соответственно отображение $\Delta_\infty : \Omega_\infty^*(\mathscr{S}) \to \mathscr{L}_\infty^*(\mathscr{M})$ будет непрерывным по C^∞-топологии в обоих пространствах.

Этот результат может быть сформулирован так:

Теорема об устойчивости решения задачи Коши

Пусть $(\mathscr{M}, g)$ — максимальное W^{5+a}-развитие начальных данных $\mathbf{h} \in W^{5+a}(\mathscr{S})$ и $\boldsymbol{\chi} \in W^{4+a}(\mathscr{S})$, $0 \leqslant a \leqslant \infty$, $\mathscr{V}$ — область из $J^+(\theta(\mathscr{S}))$ с компактным замыканием, Z — окрестность $\mathbf{g}$ в $\mathscr{L}_{5+a}(\mathscr{V})$ и $\mathscr{U} \subset \theta(\mathscr{S})$ — открытая окрестность $J^+(\overline{\mathscr{V}}) \cap \theta(\mathscr{S})$ с компактным замыканием. Тогда в $\Omega_{5+a}(\mathscr{U})$ существует окрестность Y пары $(\mathbf{h}, \boldsymbol{\chi})$, в которой для всех $(\mathbf{h}', \boldsymbol{\chi}') \in Y$, удовлетворяющих уравнениям связи, существует диффеоморфизм $\mu : \mathscr{M}' \to \mathscr{M}$ со свойствами:

1) отображение $\theta^{-1}\mu\theta'$ является тождественным на $\theta^{-1}(\mathscr{U})$,

2) $\mu_* \mathbf{g}' \in Z$,

где $(\mathscr{M}', \mathbf{g}')$ — максимальное развитие пары $(\mathbf{h}', \boldsymbol{\chi}')$. □

Утверждение этой теоремы сводится примерно к следующему: если возмущение начальных данных на поверхности Коши $\theta(\mathscr{S})$

мало в области $J^-(\bar{\mathscr{V}}) \cup \theta(\mathscr{S})$, то мы получим новое решение, близкое к исходному в $\mathscr{V}$. В действительности возмущение начальных данных должно быть малым на несколько большей области поверхности Коши, чем $J^-(\bar{\mathscr{V}}) \cap \theta(\mathscr{S})$, так как изотропные конусы в новом решении будут немного иными, чем прежде, и может оказаться, что $\mathscr{V}$ не лежит в области Коши множества $J^-(\bar{\mathscr{V}}) \cap \theta(\mathscr{S})$.

7.7. Уравнения Эйнштейна при наличии материи

До сих пор мы для простоты рассматривали уравнения Эйнштейна только для пустого пространства. Однако аналогичные результаты получаются и при наличии материи, если уравнения, которым подчиняются материальные поля $\Psi_{(i)}{}^{I}{}_{J}$, удовлетворяют определенным, физическим разумным условиям. Схема получения этих результатов такова. Решаем уравнения движения материи с заданными начальными условиями в заданном пространстве-времени с метрикой $\mathbf{g}'$; затем решаем приведенные уравнения Эйнштейна (7.42) как линейные уравнения с коэффициентами, которые определяются метрикой $\mathbf{g}'$, и с правой частью T'^{ab}, которая определяется $\mathbf{g}'$ и решением для материальных полей. При этом получаем новую метрику $\mathbf{g}''$ и повторяем все снова, взяв $\mathbf{g}''$ вместо $\mathbf{g}'$. Для того чтобы показать, что такая процедура сходится к решению системы из уравнений Эйнштейна и уравнений движения материи, на последние необходимо наложить определенные условия. Мы их сформулируем следующим образом:

а) если $\{{}_0\Psi_{(i)}\} \in W^{4+a}(\mathscr{H})$ и $\{{}_1\Psi_{(i)}\} \in W^{3+a}(\mathscr{H})$ — начальные данные на ахрональной пространственноподобной поверхности $\mathscr{H}$ в W^{4+a}-метрике $\mathbf{g}$, то существует единственное решение материальных уравнений в некоторой окрестности $\mathscr{H}$ в области $D^+(\mathscr{H})$, причем $\{\Psi_{(i)}\} \in W^{4+a}(\mathscr{H}')$ для любой гладкой пространственноподобной поверхности $\mathscr{H}'$ и

$$\Psi_{(i)} = {}_0\Psi_{(i)}, \quad \Psi_{(i)}{}^{I}{}_{J\,|\,a} = {}_1\Psi_{(i)}{}^{I}{}_{J} \quad \text{на} \quad \mathscr{H};$$

б) если $\{\Psi_{(i)}\}$ — W^{5+a}-решение в области $\mathscr{U}_+$ в W^{5+a}-метрике $\mathbf{g}$, существуют такие положительные постоянные $\tilde{Q}_1$ и $\tilde{Q}_2$, что для любого W^{4+a}-решения $\{\Psi'_{(i)}\}$ в метрике $\mathbf{g}'$, для которой

$$\| \mathbf{g}' - \mathbf{g},\ \mathscr{U}_+ \|_{4+a} < \tilde{Q}_1,$$

и

$$\sum_{(i)} \{ \| {}_0\Psi'_{(i)} - {}_0\Psi_{(i)},\ \mathscr{H}(0) \cap \mathscr{U} \|_{4+a} + \| {}_1\Psi'_{(i)} - {}_1\Psi_{(i)},\ \mathscr{H}(0) \cap \mathscr{U} \|_{3+a} \} < \tilde{Q}_1,$$

выполняется неравенство

$$\sum_{(i)} \| \Psi'_{(i)} - \Psi_{(i)}, \mathcal{U}_+ \|_{4+a} \leqslant \tilde{Q}_2 \{ \| \mathbf{g}' - \mathbf{g}, \mathcal{U}_+ \|_{4+a} + \\ + \sum_{(i)} \| \Psi'_{(i)} - {}_0\Psi_{(i)}, \mathcal{H}(0) \cap \mathcal{U} \|_{4+a} + \\ + \sum_{(i)} \| {}_1\Psi'_{(i)} - {}_1\Psi_{(i)}, \mathcal{H}(0) \cap \mathcal{U} \|_{3+a} \};$$

в) тензор энергии-импульса T_{ab} является полиномом по

$$\Psi_{(i)}{}^I{}_J, \quad \Psi_{(i)}{}^I{}_{J;a} \quad \text{и} \quad g^{ab}.$$

Условие (а) представляет собой локальную теорему Коши для материальных полей в заданной пространственно-временной метрике. Условие (б) представляет теорему об устойчивости решения задачи Коши для материальных полей при вариациях начальных данных и пространственно-временной метрики $\mathbf{g}$. Если уравнения движения материи — это квазилинейные гиперболические уравнения второго порядка и их изотропный конус совпадает с изотропным конусом метрики $\mathbf{g}$ или лежит внутри него, то выполнение этих условий можно установить так же, как это было сделано для приведенных уравнений Эйнштейна.

В случае скалярного поля или электромагнитного потенциала, удовлетворяющих линейным уравнениям, эти условия выполняются в силу предложения 7.4.7. Мы можем рассмотреть также скалярное поле, взаимодействующее с электромагнитным потенциалом: зададим метрику и электромагнитный потенциал, решим уравнение скалярного поля как линейное уравнение в этой метрике и с этим потенциалом и затем решим уравнения электромагнитного поля в заданной метрике со скалярным полем в качестве источника. Можно показать, что итерации этой процедуры при достаточно малых начальных данных сходятся на множестве вида $\mathcal{U}_+$ к решению уравнений скалярного и электромагнитного полей с их взаимодействием в заданной метрике; затем, изменяя масштабы метрики и полей, покажем, что при достаточной малости (по измерениям в пространственно-временной метрике $\mathbf{g}$) $\mathcal{U}_+$ можно получить решение для любых допустимых начальных данных. Эта же процедура действует при любом числе связанных квазилинейных гиперболических уравнений второго порядка, если связь не содержит производные выше первого порядка.

Уравнения движения идеальной жидкости не являются гиперболическими уравнениями второго порядка, а образуют квазилинейную систему первого порядка. (Определение гиперболической системы первого порядка см. в [36], р. 577.) Для такой системы можно получить аналогичные результаты при условии, что конус лучей совпадает с изотропным конусом пространства-времени

с метрикой $\mathbf{g}$ или лежит внутри него. Требование, чтобы уравнения движения материи были гиперболическими уравнениями второго порядка или составляли гиперболическую систему первого порядка и чтобы их конусы совпали с конусом пространственно-временной метрики $\mathbf{g}$ или лежали внутри него, можно считать более строгой формулировкой постулата локальной причинности из гл. 3.

При выполнении условий (а), (б) и (в) можно доказать предположения 7.5.1 и 7.5.2 для системы приведенных уравнений Эйнштейна и уравнений движения материи; из этих предложений следуют локальная и глобальная теоремы о решениях задачи Коши и теорема об устойчивости.

Глава 8

Сингулярности пространства-времени

В этой главе мы используем результаты гл. 4 и 6 для того, чтобы установить основные свойства пространственно-временных сингулярностей. Астрофизические и космологические следствия будут рассмотрены в последующих главах.

В разд. 8.1 мы обсудим вопрос об определении самого понятия сингулярности в пространстве-времени. Мы введем понятие b-неполноты — развитие представления о геодезической неполноте — как признак того, что из пространства-времени вырезаны сингулярные точки; при этом мы намечаем два возможных способа связать b-неполноту с определенным типом сингулярности кривизны. В разд. 8.2 даны четыре теоремы, которые говорят о наличии неполноты в самых разнообразных ситуациях. В разд. 8.3 мы излагаем предложенный Шмидтом метод построения b-границы, изображающей сингулярные точки пространства-времени. В разд. 8.4 доказывается, что сингулярности, предсказываемые по крайней мере одной из упомянутых теорем, не могут в точности совпадать с разрывами тензора кривизны. Мы покажем также, что неполные геодезические не бывают изолированными, а составляют трехпараметрическое семейство. Ситуация, в которой неполные кривые полностью или частично заключены в компактной области пространства-времени, обсуждается в разд. 8.5. Мы показываем, что это связано с нехаусдорфовым поведением b-границы и что в характерном пространстве-времени наблюдатель, движущийся по одной из этих неполных кривых, испытывал бы действие бесконечных сил, вызванных кривизной. Мы показываем, также, что свойства, подобные тем, которыми обладает пространство Тауба — НУТ, не могут возникнуть у пространства-времени, содержащего какую-либо материю.

8.1. Определение сингулярностей

Аналогия с электродинамикой могла бы привести к мысли, что пространственно-временную сингулярность разумно определить как точку, в которой метрический тензор не определен или недифференцируем нужное число раз. Однако при этом мы натолкнемся на ту неприятность, что эти точки можно просто вырезать и сказать, что оставшееся многообразие представляет все

пространство-время, которое, согласно предлагаемому определению, будет уже несингулярным. Конечно, рассматривать такие сингулярные точки, как часть пространства-времени, было бы неразумно, ибо в них не удовлетворялись бы обычные уравнения физики и нельзя было бы производить какие-либо измерения. Поэтому в разд. 3.1 мы определили пространство-время как пару $(\mathscr{M}, \mathbf{g})$, в которой метрика $\mathbf{g}$ лоренцева и нужное число раз дифференцируема, и, потребовав, что $(\mathscr{M}, \mathbf{g})$ нельзя расширить с сохранением нужной степени дифференцируемости, гарантировали, что из пространства-времени вместе с сингулярными точками не была выброшена ни одна несингулярная точка.

Вопрос о наличии сингулярностей в пространстве-времени превращается в вопрос о том, были ли вырезаны из пространства-времени какие-либо сингулярные точки. Есть надежда, что это можно установить по факту неполноты (в том или ином смысле) пространства-времени.

В случае многообразия $\mathscr{M}$ с положительно определенной метрикой $\mathbf{g}$ мы можем ввести функцию расстояния $\rho(x, y)$, которая равна точной нижней грани длин кривых, соединяющих x и y. Функция $\rho(x, y)$ является метрикой в топологическом смысле, т. е. базисом для открытых множеств многообразия $\mathscr{M}$ служат множества $\mathscr{B}(x, r)$, состоящие из всех точек $y \in \mathscr{M}$, для которых $\rho(x, y) < r$. Пара $(\mathscr{M}, \mathbf{g})$ называется *метрически полной* (m-полной), если любая фундаментальная по функции расстояния ρ последовательность сходится к точке, лежащей в $\mathscr{M}$. (*Фундаментальной последовательностью* называется бесконечная последовательность, для которой при любом $\varepsilon > 0$ существует число N, такое, что $\rho(x_n, x_m) < \varepsilon$ для любых n и m, бóльших N.) Имеется альтернативная формулировка: $(\mathscr{M}, \mathbf{g})$ является m-полной, если каждая C^1-кривая конечной длины имеет конечную точку в смысле разд. 6.2 (заметим, что эта кривая не обязательно должна быть класса C^1 на конце). Отсюда следует, что m-полнота обеспечивает *геодезическую полноту* (*g-полноту*), т. е. возможность продолжить каждую геодезическую до произвольных значений ее аффинного параметра. Действительно, было доказано [92], что для положительно-определенной метрики g-полнота и m-полнота эквивалентны.

Лоренцева метрика не определяет какую-либо топологическую метрику, поэтому у нас остается только g-полнота. Можно различать три типа g-неполноты: времениподобную, изотропную и пространственноподобную. Если вырезать из пространства-времени регулярную точку, то получившееся многообразие будет неполно во всех трех смыслах и можно было бы надеяться, что полнота одного из упомянутых типов свидетельствует о полноте других двух типов. К сожалению, это не обязательно так [97], что видно из следующего примера, найденного Героком [54]. Рас-

смотрим двумерное пространство-время с координатами x, t и метрикой g_{ab}. Введем новую метрику $\hat{g}_{ab} = \Omega^2 g_{ab}$, где положительная функция Ω обладает следующими свойствами:

1) $\Omega = 1$ вне области между вертикальными линиями $x = -1$ и $x = +1$.

2) Ω симметрична относительно оси t, т. е. $\Omega(t, x) = \Omega(t, -x)$.

3) На оси t $t^2\Omega \to 0$ при $t \to \infty$.

В силу свойства (2) ось t — времениподобная геодезическая, которая, согласно (3), неполна при $t \to \infty$. Однако каждая изотропная или пространственноподобная геодезическая покидает область между $x = -1$ и $x = +1$ и не возвращается в нее. Следовательно, в силу свойства (1) данное пространство изотропно и пространственноподобно полно. Можно построить примеры, которые неполны в любом из трех указанных выше смыслов и полны в оставшихся двух.

Времениподобная неполнота имеет непосредственное физическое истолкование: из нее следует возможность существования свободно движущихся наблюдателей, мировые линии которых имеют начало или конец на конечном интервале собственного времени. Такое поведение мировой линии может вызвать даже более сильное возражение, чем бесконечная кривизна, и, по-видимому, пространство, в котором имеются такие линии, следует рассматривать как сингулярное. Хотя аффинный параметр на изотропной геодезической не совпадает по своему физическому смыслу с собственным временем на времениподобной геодезической, тем не менее изотропно геодезически неполные пространства, вроятно, тоже нужно считать сингулярными, как из-за того, что изотропные геодезические представляют собой истории частиц с нулевой массой покоя, так и по причине существования примеров (например, решение Райсснера — Нордстрема, разд. 5.5), которые, надо думать, сингулярны и при этом времениподобно полны, но изотропно геодезически неполны. Поскольку по пространственноподобным кривым ничто не движется, смысл пространственноподобной геодезической неполноты не столь ясен. Поэтому мы будем придерживаться той точки зрения, что *времениподобная и изотропная геодезическая полнота является минимальным условием, при котором пространство-время можно считать свободным от сингулярностей.* Отсюда, если пространство-время времениподобно или изотропно геодезически неполно, мы скажем, что в нем есть сингулярность.

Преимущество принятия времениподобной и (или) изотропной неполноты в качестве индикатора сингулярности состоит в том, что, исходя из этого, мы можем доказать ряд теорем об условиях, при которых сингулярность появляется. Однако класс времениподобно и (или) изотропно неполных пространств не включает в себя все пространства, которые было бы желательно

считать в том или ином смысле сингулярными. Например, Герок [54] построил пространство-время, которое геодезически полно, но содержит непродолжимую времениподобную кривую (не геодезическую) конечной длины с ограниченным ускорением. Такую кривую мог бы пролететь наблюдатель на каком-нибудь космическом корабле, обладая конечным запасом топлива. По истечении конечного промежутка времени его нельзя было бы изображать точкой этого пространственно-временного многообразия. Если мы собираемся утверждать, что пространство-время, в котором жизнь свободно падающего наблюдателя преждевременно обрывается, содержит сингулярность, то, очевидно, мы должны быть готовы к тому же выводу и в случае наблюдателя на космическом корабле. Для этого нам требуется какое-то обобщение понятия аффинного параметра на случай любых C^1-кривых — геодезических и негеодезических. Тогда можно было бы ввести понятие полноты, требуя, чтобы в полном пространстве любая C^1-кривая конечной длины, в смысле измерения по такому параметру, имела бы конечную точку.

Идея, которой мы собираемся воспользоваться, была, по-видимому, впервые высказана Эресманом [46] и сформулирована в более изящной форме Шмидтом [148].

Пусть $\lambda(t)$ — C^1-кривая, проходящая через $p \in \mathscr{M}$, и $\{\mathbf{E}_i\}$ $(i = 1, 2, 3)$ — базис в T_p. Мы можем параллельно перенести $\{\mathbf{E}_i\}$ вдоль $\lambda(t)$ и получить базис в $T_{\lambda(t)}$ при каждом значении t. Тогда касательный вектор $\mathbf{V} = (\partial/\partial t)_{\lambda(t)}$ можно будет выразить через этот базис: $\mathbf{V} = \mathbf{V}^i(t)\,\mathbf{E}_i$ — и мы сможем посредством формулы

$$u = \int_p \left(\sum_i V^i V^i\right)^{1/2} \mathrm{d}t$$

ввести *обобщенный аффинный параметр* и на λ. Параметр u зависит от точки p и от базиса $\{\mathbf{E}_i\}$ в p. Если $\{\mathbf{E}_{i'}\}$ — другой базис в p, то существует некоторая несингулярная матрица $A_{i'}^{\ j}$, такая, что

$$\mathbf{E}_i = \sum_{j'} A_i^{\ j'} \mathbf{E}_{j'}.$$

Поскольку $\{\mathbf{E}_{i'}\}$ и $\{\mathbf{E}_i\}$ переносятся вдоль $\lambda(t)$ параллельно, это соотношение выполняется при постоянных $A_i^{\ j'}$. Таким образом,

$$V^{i'}(t) = \sum_j A_j^{\ i'} V^j(t).$$

Ввиду несингулярности матрицы $A_i^{\ j'}$ найдется постоянная $C > 0$, для которой

$$C \sum_i V^i V^i \leqslant \sum_{i'} V^{i'} V^{i'} \leqslant C^{-1} \sum_i V^i V^i.$$

Следовательно, длина кривой λ по параметру u конечна тогда и только тогда, когда она конечна по параметру u'. Если λ — геодезическая кривая, то u — аффинный параметр на λ; красота же этого определения состоит в том, что u можно задать на любой C^1-кривой. Мы будем говорить, что пространство-время $(\mathscr{M}, \mathbf{g})$ *b-полное* (сокращение от *bundle-complete*, см. разд. 8.3), если у каждой C^1-кривой конечной длины, измеренной по обобщенному аффинному параметру, существует конечная точка. Если эта длина конечна при одном обобщенном аффинном параметре, то она будет конечна при всех таких параметрах, так что мы ничего не теряем, если ограничимся лишь ортонормированными базисами. Когда метрика $\mathbf{g}$ положительно определена, обобщенный аффинный параметр, определяемый ортонормированным базисом, является длиной дуги и потому b-полнота совпадает с m-полнотой. Однако понятие b-полноты можно ввести, даже если метрика не является положительно-определенной; фактически для ее определения достаточно наличия связности на $\mathscr{M}$. Ясно, что из b-полноты следует g-полнота, но приведенный выше пример показывает, что обратное утверждение неверно.

Итак, мы будем называть пространство-время свободным от сингулярностей, если оно b-полное. Это определение согласуется с высказанной выше точкой зрения, что времениподобная и изотропная геодезическая полнота является минимальным условием для того, чтобы пространство-время можно было считать свободным от сингулярностей. Может появиться желание слегка ослабить это условие и называть пространство-время свободным от сингулярностей только тогда, когда оно непространственноподобно b-полное, т. е. если все непространственноподобные C^1-кривые конечной длины, измеренной по обобщенному аффинному параметру, имеют конечную точку. Однако это определение выглядело бы довольно громоздким при формулировке b-полноты через расслоенные пространства; такую формулировку мы дадим в разд. 8.3. Фактически в каждой из теорем, доказываемых нами в разд. 8.2, подразумевается, что пространство-время $(\mathscr{M}, \mathbf{g})$ времениподобно или изотропно g-неполное и, следовательно, содержит сингулярность по обоим из вышеприведенных определений.

Интуитивно чувствуется, что с сингулярностью должна быть связана неограниченно большая кривизна вблизи сингулярной точки. Но мы исключили сингулярные точки из определенного нами пространства-времени, и поэтому возникает трудность с тем, что такое «вблизи» и что такое «неограниченно большая». Можно сказать, что точки b-неполной кривой находятся вблизи сингулярности, если им соответствуют значения обобщенного аффинного параметра, близкие к верхней грани этого параметра. Труднее придать точный смысл «неограниченно большой кривизне»,

поскольку численные значения компонент тензора кривизны зависят от базиса, в котором они измеряются. Один из подходов состоит в том, чтобы рассматривать скалярные полиномы по g_{ab}, η_{abcd} и R_{abcd}. Будем говорить, что b-неполная кривая соответствует *сингулярности скалярных полиномов кривизны*, если любой из этих скалярных полиномов неограничен на данной неполной кривой. Однако, когда метрика лоренцева, эти полиномы не определяют полностью тензор кривизны: Пенроуз заметил, что для решений в виде плоской волны все скалярные полиномы равны нулю, тогда как тензор Римана отличен от нуля (подобно тому как ненулевой вектор может иметь нулевую длину). Таким образом, кривизна может в определенном смысле быть очень большой, даже если скалярные полиномы малы. С другой стороны, можно измерить компоненты тензора кривизны в базисе, параллельно перенесенном вдоль кривой. Будем говорить, что b-неполная кривая соответствует *сингулярности кривизны в параллельно перенесенном базисе*, если любая из компонент тензора кривизны на этой кривой не ограничена. Ясно, что существование сингулярности скалярных полиномов означает также существование сингулярности в параллельно перенесенном базисе.

Можно было бы надеяться, что в физически реалистических решениях b-неполнота кривой соответствует обоим видам сингулярности кривизны. Однако пространство Тауба — НУТ (разд. 5.8) служит примером решения, где, по-видимому, эта надежда не оправдывается. В этом пространстве неполные геодезические полностью заключены в некоторой компактной окрестности горизонта. Метрика в этой компактной окрестности вполне регулярна, и потому скалярные полиномы тензора кривизны конечны. Ввиду особого характера этого решения компоненты тензора кривизны в базисе, параллельно перенесенном вдоль заключенной геодезической, остаются ограниченными. Поскольку заключенная геодезическая содержится в компактном множестве, многообразие $\mathscr{M}$ нельзя расширить до такого большого четырехмерного хаусдорфова паракомпактного многообразия $\mathscr{M}'$, в котором неполные геодезические можно было бы продолжить. Следовательно, эта неполнота возникла не от удаления сингулярных точек. Тем не менее движение по одной из этих неполных времениподобных геодезических доставило бы наблюдателю мало удовольствия: хотя его мировая линия не имеет конца, она кружится и кружится внутри компактного множества, и наблюдатель никогда не перейдет через определенный момент своей жизни. Поэтому представляется разумным считать такого рода пространство-время сингулярным, хотя в нем и нет сингулярности скалярных полиномов кривизны или сингулярности кривизны в параллельно перенесенном базисе. Согласно лемме 6.4.8, не-

полнота, состоящая в полном заключении, может возникнуть лишь при нарушении условия сильной причинности. В § 8.5 мы покажем, что в типовом пространстве-времени частично или полностью заключенная b-неполная кривая соответствует сингулярности кривизны в параллельно перенесенном базисе. Мы покажем также, что полностью заключенная неполнота такого рода, как в пространстве Тауба — НУТ, не появляется, если в пространстве-времени имеется хоть какая-либо материя.

8.2. Теоремы о сингулярностях

В разд. 5.4 было показано, что при определенных разумных условиях в пространственно-однородных решениях будут сингулярности. Аналогичные теоремы можно доказать для ряда других классов точных симметрий. Тем не менее вовсе не обязательно было бы придавать этим результатам какой-либо физический смысл, поскольку они все-таки обусловлены ограничением, чтобы симметрия была точной, чего, конечно, не будет ни в какой реальной физической ситуации. Поэтому многие авторы полагали, что сингулярности являются лишь следствием симметрии и не появятся в общем решении. Эта точка зрения была поддержана Е. М. Лифшицем, И. М. Халатниковым и их сотрудниками, которые показали, что определенные классы решений с пространственноподобными сингулярностями не содержат того числа произвольных функций, которое предполагается у общего решения уравнений поля (эта работа подытожена в [101]). Как считали авторы [101], это свидетельствует о том, что данные Коши, приводящие к таким сингулярностям, имеют меру нуль в множестве всех возможных данных Коши и, значит, не должны иметь отношения к реальной Вселенной. Однако впоследствии В. А. Белинский, И. М. Халатников и Е. М. Лифшиц [5] нашли иные классы решений, которые, по-видимому, содержали полное число произвольных функций, причем в этих решениях были сингулярности. Поэтому они отказались от утверждения, что в общих решениях нет сингулярностей. Их методы интересны тем, что проливают свет на возможную структуру сингулярностей, но не ясно, сходятся ли используемые ими степенные ряды. Никто еще на нашел общего вида условий, при которых возникновение сингулярности неизбежно. Тем не менее мы рассматриваем результаты Белинского, Лифшица и Халатникова как свидетельство в пользу нашей точки зрения, согласно которой сингулярности, предсказываемые теоремами этого раздела, вообще говоря, связаны с бесконечностью кривизны.

Первая теорема о сингулярностях, не содержащая каких-либо предположений о симметриях, была доказана Пенроузом [126]. Он ставил целью доказать, что в звезде, коллапсирующей внутрь

своего шварцшильдова радиуса, появится сингулярность. Если коллапс строго сферически-симметричен, решение можно получить прямым интегрированием, и сингулярность всегда возникает. Однако далеко не очевидно, что так будет и тогда, когда существует неоднородность или хотя бы небольшой момент количества движения. В самом деле, в ньютоновской теории самый незначительный момент количества движения может воспрепятствовать появлению бесконечной плотности и вызвать у звезды переход от сжатия к расширению. Однако Пенроуз показал, что в общей теории относительности ситуация совсем иная: если коллапсирующая звезда ушла внутрь своей шварцшильдовой поверхности (поверхность $r = 2m$), то она уже не сможет из-под нее выйти. Шварцшильдова поверхность в действительности определена лишь для строго сферически-симметричного решения, но Пенроуз использовал более общий критерий, который эквивалентен в случае решения Шварцшильда существованию шварцшильдовой поверхности и приложим к решениям, не обладающим строгой симметрией. Этим критерием является существование *замкнутой ловушечной поверхности* $\mathcal{T}$, которая представляет собой такую замкнутую (т. е. компактную, без края) пространственноподобную 2-поверхность класса C^2 (обычно S^2), что два семейства ортогональных к $\mathcal{T}$ изотропных геодезических у поверхности $\mathcal{T}$ сходятся (т. е. ${}_1\hat{\chi}_{ab}g^{ab}$ и ${}_2\hat{\chi}_{ab}g^{ab}$ отрицательны, где ${}_1\hat{\chi}_{ab}$ и ${}_2\hat{\chi}_{ab}$ — две изотропные вторые фундаментальные формы на $\mathcal{T}$). (В последующих главах мы обсудим обстоятельства, при которых появляется такая поверхность.) Это можно представить себе следующим образом: $\mathcal{T}$ находится в столь сильном гравитационном поле, что даже расходящиеся лучи света затягиваются обратно и превращаются в сходящиеся. Поскольку ничто не может двигаться быстрее света, материя, заключенная в $\mathcal{T}$, последовательно запирается внутри 2-поверхностей все меньшей и меньшей площади, и все выглядит так, словно должно случиться что-то «плохое». Что это так и есть, строго доказывает теорема Пенроуза.

Теорема 1

Пространство-время $(\mathcal{M}, \mathbf{g})$ не может быть изотропно геодезически полным, если

1) $R_{ab}K^aK^b \geqslant 0$ для всех изотропных векторов (ср. с разд. 4.3);

2) в $\mathcal{M}$ существует некомпактная поверхность Коши $\mathcal{H}$;

3) в $\mathcal{M}$ существует замкнутая ловушечная поверхность $\mathcal{T}$.

Замечание. Идея доказательства такова. Нужно показать, что граница будущего поверхности $\mathcal{T}$ будет компактной, если $\mathcal{M}$ — изотропно геодезически полное многообразие. Затем доказывается, что это несовместимо с некомпактностью $\mathcal{H}$.

Доказательство. Из существования поверхности Коши следует, что $\mathcal{M}$ — глобально-гиперболическое (предложение 6.6.3) и потому причинно простое (предложение 6.6.1) многообразие. Это означает, что границей множества $J^+(\mathcal{T})$ будет поверхность $E^+(\mathcal{T})$ и ее образующими будут ортогональные $\mathcal{T}$ геодезические

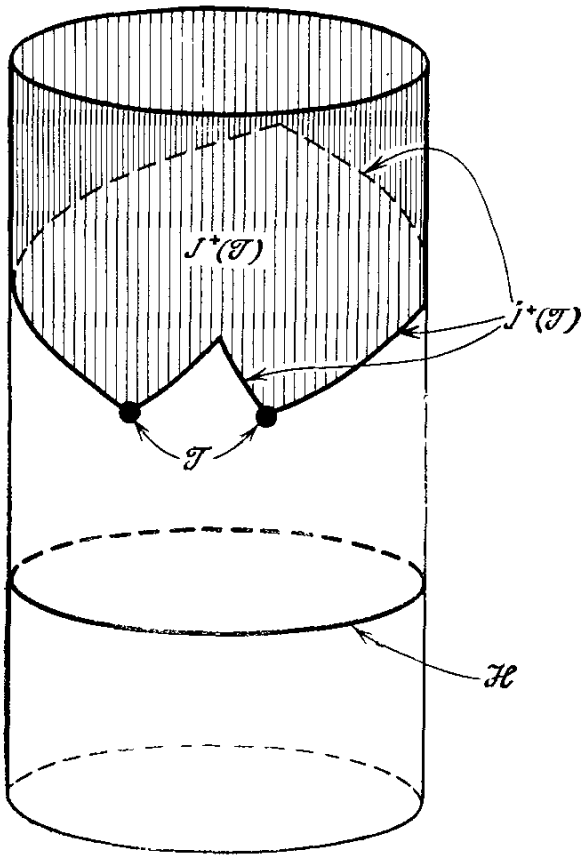


Рис. 49. Двумерное сечение геодезически полного пространства с компактной поверхностью Коши $\mathcal{H}$. Будущее $J^+(\mathcal{T})$ 2-сферы $\mathcal{T}$ имеет компактную границу $\dot{J}^+(\mathcal{T})$, поскольку расходящиеся изотропные геодезические встречаются, обойдя цилиндр кругом.

сегменты с конечными точками в прошлом на $\mathcal{T}$. Допустим, что $\mathcal{M}$ — изотропно геодезически полное многообразие. Тогда, согласно условиям (1) и (3) и предложению 4.4.6 вдоль любой направленной в будущее изотропной геодезической, ортогональной к $\mathcal{T}$, в пределах аффинного расстояния, $2c^{-1}$ должна быть точка, сопряженная $\mathcal{T}$, причем c равно значению ${}_n\hat{\chi}_{ab}g^{ab}$ в точке, где данная изотропная геодезическая пересекает $\mathcal{T}$. Согласно предложению 4.5.14, те точки такой изотропной геодезической, которые лежат за точкой, сопряженной $\mathcal{T}$, принадлежат области $J^+(\mathcal{T})$. Таким образом, каждый образующий сегмент поверхности $\dot{J}^+(\mathcal{T})$ должен иметь конечную точку в будущем в сопряженной к $\mathcal{T}$ точке или до нее. Вблизи $\mathcal{T}$ на каждой ортогональной к $\mathcal{T}$ изотропной геодезической можно задать непрерывным образом аффинный параметр. Рассмотрим непрерывное

отображение $\beta: \mathscr{T} \times [0, b] \times Q \to \mathscr{M}$ (Q — дискретное множество 1, 2), определяемое переносом точки $p \in \mathscr{T}$ на аффинное расстояние $v \in [0, b]$ вдоль одной из двух направленных в будущее изотропных геодезических, проходящих через p ортогонально к $\mathscr{T}$. Ввиду компактности поверхности $\mathscr{T}$ будет существовать некоторое минимальное значение c_0 величин $(-{}_1\hat{\chi}_{ab}g^{ab})$ и $(-{}_2\hat{\chi}_{ab}g^{ab})$. Тогда, $\dot{J}^+(\mathscr{T}) \subset \beta(\mathscr{T} \times [0, b_0] \times Q)$ при $b_0 = 2c_0^{-1}$. Следовательно, $\dot{J}(\mathscr{T})$, будучи замкнутым подмножеством компактного множества, компактно. Это было бы возможно, если бы поверхность Коши $\mathscr{H}$ была компактной, потому что тогда граница могла бы замкнуться и образовать компактную поверхность Коши, гомеоморфную $\mathscr{H}$ (рис. 49). Однако ясно, что мы встретим затруднения, если потребуем, чтобы поверхность $\mathscr{H}$ была некомпактной. Чтобы доказать это строго, мы можем воспользоваться тем фактом (см. разд. 2.6), что $\mathscr{M}$ допускает направленное в прошлое, времениподобное векторное C^1-поле. Каждая интегральная кривая этого поля будет пересекать $\mathscr{H}$ (поскольку это поверхность Коши) и пересекает $\dot{J}^+(\mathscr{T})$ самое большее один раз. Таким образом, эти интегральные кривые задают непрерывное взаимнооднозначное отображение $\dot{J}^+(\mathscr{T}) \to \mathscr{H}$. Если множество $\dot{J}^+(\mathscr{T})$ компактно, его образ $\alpha(\dot{J}^+(\mathscr{T}))$ тоже компактен и поэтому гомеоморфен $\dot{J}(\mathscr{T})$. Однако ввиду некомпактности $\mathscr{H}$ множество $\alpha(\dot{J}^+(\mathscr{T}))$ может и не содержать всю поверхность $\mathscr{H}$ и в таком случае должно иметь границу в $\mathscr{H}$. Это, однако, невозможно, поскольку, согласно предложению 6.3.1, $\dot{J}^+(\mathscr{T})$, а следовательно, и $\alpha(\dot{J}^+(\mathscr{T}))$ должны быть трехмерными многообразиями (без края). Отсюда видно, что допущение изотропной геодезической полноты $\mathscr{M}$ (оно было сделано для доказательства компактности $\dot{J}^+(\mathscr{T})$) неверно. □

Условие (1) этой теоремы (именно, что $R_{ab}K^aK^b \geqslant 0$ для любого изотропного вектора $\mathbf{K}$) обсуждалось в разд. 4.3. Чтобы оно выполнялось, не имеет значения, какого порядка величины постоянная Λ, лишь бы плотность энергии была положительной для любого наблюдателя. В гл. 9 будет показано, что условие (3) (существования замкнутой ловушечной поверхности) должно удовлетворяться по крайней мере в некоторой области пространства-времени. Остается рассмотреть условие (2) (существования некомпактной пространственноподобной поверхности $\mathscr{H}$, являющейся поверхностью Коши). Согласно предложению 6.4.9, если допустить устойчивую причинность, то существование пространственноподобных поверхностей обеспечено. Условие компактности пространственноподобной поверхности — не слишком серьезное ограничение, поскольку единственный пункт, в котором

мы его использовали, это когда показывали, что $\alpha(J^+(\mathscr{T}))$ не может совпадать со всей поверхностью $\mathscr{H}$. Однако можно доказать то же самое, если вместо некомпактности $\mathscr{H}$ потребовать, чтобы с поверхности $\mathscr{H}$ можно было провести направленную в будущее непродолжимую кривую, которая не пересечет $\dot{J}^+(\mathscr{T})$. Другими словами, теорема была бы справедлива даже тогда, когда поверхность $\mathscr{H}$ компактна, если при этом существуют наблюдатели, которые могут избежать падения на коллапсирующую звезду. Это было бы невозможно, если бы коллапсировала сама Вселенная; но и при этих условиях, надо думать, сингулярности имеются, и это мы скоро покажем. Действительно тонкий пункт теоремы — это требование, чтобы $\mathscr{H}$ была поверхностью Коши. Оно используется в двух местах: во-первых, при доказательстве того, что $\mathscr{M}$ — причинно простое многообразие, из чего следовало наличие у образующих поверхности $\dot{J}^+(\mathscr{T})$ конечных точек $\mathscr{T}$, и, во-вторых, когда мы проверяли, что при отображении α каждая точка $\dot{J}^+(\mathscr{T})$ отображается в точку поверхности $\mathscr{H}$. Что это условие (чтобы $\mathscr{H}$ была поверхностью Коши) необходимо, видно из примера, данного Бардином. Глобальная структура в нем такая же, как у решения Райсснера — Нордстрема, за тем исключением, что реальные сингулярности при $r = 0$ устроены так, что они становятся в точности началом полярных координат. Это пространство-время удовлетворяет условию $R_{ab}K^aK^b \geqslant 0$ для любого изотропного, но не времениподобного вектора $\mathbf{K}$, и в нем имеются замкнутые ловушечные поверхности. Единственно, чем это пространство-время не удовлетворяет условиям теоремы 1, тем, что в нем нет поверхностей Коши.

Следовательно, теорема 1, по всей видимости, утверждает, что в коллапсирующей звезде будет или сингулярность, или горизонт Коши. Это очень важный результат, ибо в любом случае мы теряем возможность предсказывать будущее. Однако он не дает ответа на вопрос, появляются ли сингулярности в физически реалистических решениях? Для того чтобы разрешить его, нам нужна теорема, в которой не предполагается существование поверхностей Коши. Одно из условий такого рода теоремы должно состоять в том, что $R_{ab}K^aK^b \geqslant 0$ не только для изотропных, но и для всех *времениподобных* векторов, поскольку невыполнение этого условия — единственная возможность для того, чтобы пример Бардина перестал быть аргументом. Теорема, которую мы приведем ниже, включает это условие и, кроме того, хронологические условия (условие отсутствия замкнутых времениподобных кривых). С другой стороны, она приложима к более широкому классу ситуаций, поскольку существование замкнутых ловушечных поверхностей является теперь только одним из трех возможных условий. Другое из этих альтернативных условий

состоит в том, чтобы существовала компактная частичная поверхность Коши, а третья — в том, чтобы существовала точка, световой конус прошлого (или будущего) которой начинает сходиться снова (рис. 50). Первое из этих двух последних условий выполняется в пространственно замкнутых решениях, тогда как второе тесно связано с существованием замкнутых ловушечных поверхностей, но его формулировка лучше подходит для наших целей: например, в том случае, когда сходящийся вдали све-

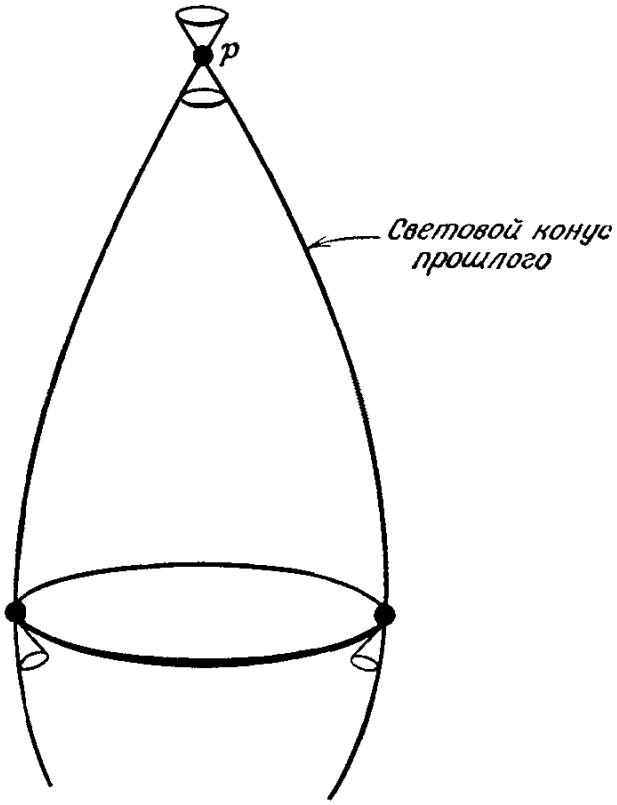


Рис. 50. Точка p, изотропный конус прошлого которой начинает сходиться вторично.

товой конус — это наш собственный световой конус прошлого; тогда мы можем непосредственно установить, выполняется ли это условие. В последней главе будет показано, что наблюдения микроволнового фона свидетельствуют о его выполнении.

Точная формулировка такова:

Теорема 2 (*Хокинг и Пенроуз* [75])

Пространство-время $(\mathscr{M}, \mathbf{g})$ не является времениподобно или изотропно геодезически полным, если

1) $R_{ab}K^aK^b \geqslant 0$ для любого непространственноподобного вектора $\mathbf{K}$ (ср. с разд. 4.3);

2) выполняется типовое условие разд. 4.4, т. е. если каждая непространственноподобная геодезическая содержит точку, в которой $K_{[a}R_{b]cd[e}K_{f]}K^cK^d \neq 0$, где $\mathbf{K}$ — касательный вектор к этой геодезической;

3) на $\mathscr{M}$ выполняется хронологическое условие (т. е. нет замкнутых времениподобных кривых);

4) справедливо по меньшей мере одно из трех следующих утверждений:

а) существует компактное ахрональное множество без границы;

б) существует замкнутая ловушечная поверхность;

в) существует такая точка p, что на любой направленной в прошлое (или на любой направленной в будущее) изотропной геодезической из точки p расхождение $\hat{\theta}$ изотропных геодезических из p становится отрицательным (т. е. изотропные геодезические, выходящие из p, фокусируются материей или кривизной и начинают сходиться).

Замечание. Приведем альтернативную формулировку теоремы.

Следующие три условия не могут выполняться одновременно:

I) каждая непродолжимая непространственноподобная геодезическая содержит пару сопряженных точек;

II) на $\mathscr{M}$ выполняется хронологическое условие;

III) имеется ахрональное множество $\mathscr{S}$, для которого $E^+(\mathscr{S})$ или $E^-(\mathscr{S})$ компактны. (Будем говорить, что такое множество — соответственно *ловушечное для будущего* или *ловушечное для прошлого.*)

Мы докажем теорему 2 именно в этой формулировке. Доказательство первого варианта отсюда следует непосредственно, поскольку, если бы многообразие $\mathscr{M}$ было времениподобно и изотропно полным, то из условий (1) и (2), в силу предложений 4.4.2 и 4.4.5, следовало бы условие (I); условие (3) — то же самое, что и (II), а из (1) и (4) вытекало бы условие (III): в случае (а) $\mathscr{S}$ было бы компактным ахрональным множеством без границы и

$$E^+(\mathscr{S}) = E^-(\mathscr{S}) = \mathscr{S};$$

а в случаях (б) или (в) множество $\mathscr{S}$ будет соответственно замкнутой ловушечной поверхностью или точкой p; тогда, согласно предложениям 4.4.4, 4.4.6, 4.5.12 и 4.5.14, соответственно $E^+(\mathscr{S})$ или $E^-(\mathscr{S})$ будут компактны как пересечения замкнутых множеств $J^+(\mathscr{S})$ и $J^-(\mathscr{S})$ с компактными множествами, состоящими из всех изотропных геодезических некоторой конечной длины от $\mathscr{S}$.

Доказательство. Поскольку доказательство довольно длинное, мы разделим его на части и докажем сначала лемму и следствие из нее. Отметим, что из рассуждений, аналогичных доказательству предложения 6.4.6, вытекает, что в силу условий (I) и (II) на $\mathscr{M}$ имеет место сильная причинность.

Лемма 8.2.1

Если $\mathscr{S}$ — замкнутое множество и на $\bar{J}^+(\mathscr{S})$ выполняется условие сильной причинности, то множество $H^+(\bar{E}^+(\mathscr{S}))$ или некомпактно, или пусто (рис. 51).

По лемме 6.3.2 через каждую точку $q \in \dot{J}^+(\mathscr{S}) - \mathscr{S}$ проходит направленный в прошлое и лежащий в $\dot{J}^+(\mathscr{S})$ изотропный

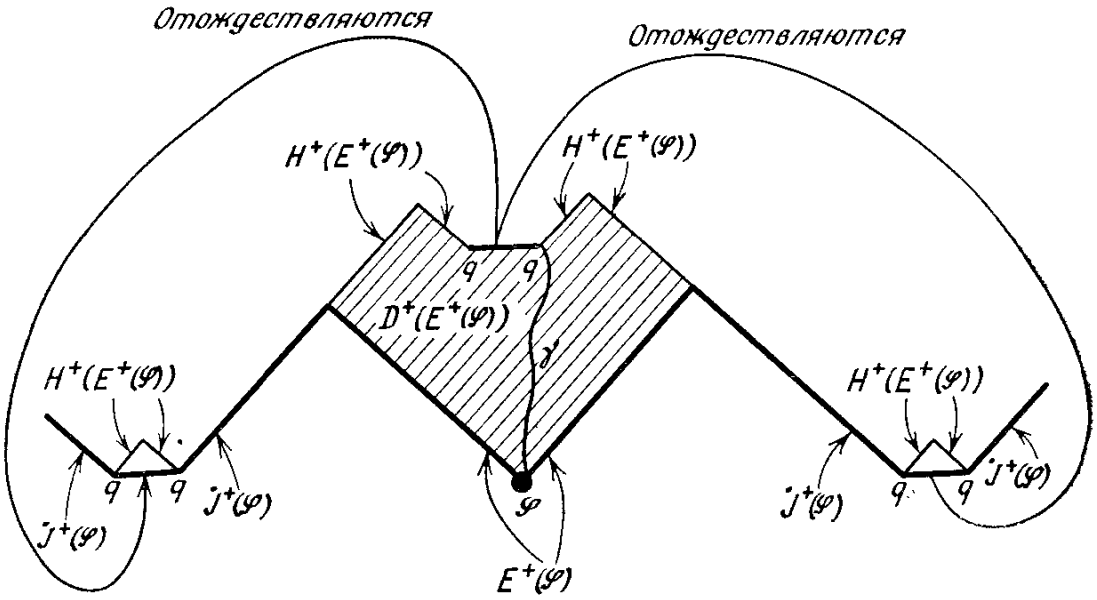


Рис. 51. Ловушечное множество для будущего $\mathscr{S}$; изотропные линии проходят под $\pm 45°$; три линии отождествлены, а точки q находятся на бесконечности. Приведены ахрональные множества $E^+(\mathscr{S})$, $\dot{J}^+(\mathscr{S})$, $H^+(E^+(\mathscr{S}))$ и непродолжимая в будущее времениподобная кривая $\gamma \in D^+(E^+(\mathscr{S}))$.

геодезический сегмент, который обладает конечной точкой в прошлом тогда и только тогда, когда $q \in E^+(\mathscr{S})$. (Заметим, что мы более не предполагаем существования поверхности Коши, $\mathscr{M}$ может и не быть причинно-простым многообразием и соответственно $\dot{J}^+(\mathscr{S}) - E^+(\mathscr{S})$ может быть непустым множеством.) Отсюда, если $q \in \dot{J}^+(\mathscr{S}) - E^+(\mathscr{S})$, то через q проходит непродолжимая в прошлое изотропная геодезическая, которая лежит в $\dot{J}^+(\mathscr{S})$ и, следовательно, не пересекает $I^-(\dot{J}^+(\mathscr{S}))$. Тогда из леммы 6.6.4 следует, что $q \notin D^+(\dot{J}(\mathscr{S})) - H^+(\dot{J}^+(\mathscr{S}))$. Поэтому

$$D^+(\bar{E}^+(\mathscr{S})) - H^+(\bar{E}^+(\mathscr{S})) = D^+(\dot{J}^+(\mathscr{S})) - H^+(\dot{J}^+(\mathscr{S}))$$

и

$$H^+(\bar{E}^+(\mathscr{S})) \subset H^+(\dot{J}^+(\mathscr{S})).$$

Теперь допустим, что множество $H^+(\bar{E}^+(\mathscr{S}))$ не пусто и компактно. Тогда его можно покрыть конечным числом окрестностей локальной причинности $\mathscr{U}_i$. Пусть p_1 — точка множества $J^+(\mathscr{S}) \cap [\mathscr{U}_1 - D^+(\dot{J}^+(\mathscr{S}))]$; из нее можно провести непродол-

жимую в прошлое непространственноподобную кривую, которая не пересечет ни $I^+(\mathscr{S})$, ни $D^+(\bar{E}^+(\mathscr{S}))$. Поскольку окрестности $\mathscr{U}_i$ имеют компактные замыкания, λ_1 выходит из $\mathscr{U}_1$. Пусть q_1 — точка на λ_1 вне $\mathscr{U}_1$. Тогда, поскольку $q_1 \in I^+(\mathscr{S})$, найдется непространственноподобная кривая μ_1 от q_1 до $\mathscr{S}$. Эта кривая пересечет $D^+(\bar{E}^+(\mathscr{S}))$, а значит, и некоторые из окрестностей $\mathscr{U}_i$, отличные от $\mathscr{U}_1$ (скажем, $\mathscr{U}_2$). Теперь пусть p_2 — точка из $\mu_1 \cap [\mathscr{U}_2 - D^+(J^+(\mathscr{S}))]$, поступим, как и прежде, и т. д.

Эта процедура ведет к противоречию, ибо окрестностей локальной причинности предполагалось лишь конечное число, и в то же время мы не можем вернуться в предыдущие $\mathscr{U}_j$, поскольку ни одна непространственноподобная кривая не может пересечь $\mathscr{U}_i$ больше одного раза. Следовательно, множество $H^+(\bar{E}^+(\mathscr{S}))$ должно быть или некомпактным, или пустым. □

Следствие

Если $\mathscr{S}$ — ловушечное множество для будущего, существует непродолжимая в будущее времениподобная кривая, которая целиком содержится в $D^+(E^+(\mathscr{S}))$.

Введем на $\mathscr{M}$ времениподобное векторное поле. Если каждая интегральная кривая этого поля, пересекающая $E^+(\mathscr{S})$, пересекает также и $H^+(E^+(\mathscr{S}))$, то эти кривые задают непрерывное взаимно-однозначное отображение $E^+(\mathscr{S})$ на $H^+(E^+(\mathscr{S}))$, и потому область $H^+(E^+(\mathscr{S}))$ будет компактной. Пересечение $I^+(\mathscr{S})$ с кривой, которая не пересекает $H^+(E^+(\mathscr{S}))$, дает искомую кривую γ (рис. 51 изображает одну из возможных ситуаций). □

Рассмотрим теперь компактное множество $\mathscr{F}$, определяемое как $E^+(\mathscr{S}) \cap \overline{J^-(\gamma)}$. Поскольку γ содержится в $\operatorname{int} I^+(E^+(\mathscr{S}))$, $E^-(\mathscr{F})$ будет состоять из $\mathscr{F}$ и части $\dot{J}^-(\gamma)$. Поскольку кривая γ непродолжима в будущее, изотропные геодезические образующие сегменты поверхности $\dot{J}^-(\gamma)$ не могут иметь конечных точек в будущем. Но по условию (а) всякая непродолжимая непространственноподобная геодезическая содержит пару сопряженных точек. Следовательно, в силу предложения 4.5.12 непродолжимое в прошлое продолжение ν' каждого образующего сегмента ν поверхности $\dot{J}^-(\gamma)$ должно войти в область $I^-(\gamma)$. При этом ν должна иметь конечную точку в прошлом в первой точке p множества $\nu' \cap I^-(\gamma)$ или до нее. $I^-(\gamma)$ — открытое множество, и некоторая окрестность точки p должна содержать точки области $I^-(\gamma)$, лежащие на соседних геодезических. Следовательно, аффинное расстояние до точек p от $\mathscr{F}$ должно быть полунепрерывно сверху, а множество $E^-(\mathscr{F})$ должно быть компактным как пересечение замкнутого множества $\dot{J}^-(\gamma)$ с компактным множе-

ством, которое образовано проведенными из $\mathscr{F}$ геодезическими сегментами некоторой ограниченной аффинной длины. Тогда из леммы следует, что найдется непродолжимая в прошлое времениподобная кривая λ, содержащаяся в $\operatorname{int} D^-(E^-(\mathscr{F}))$ (рис. 52).

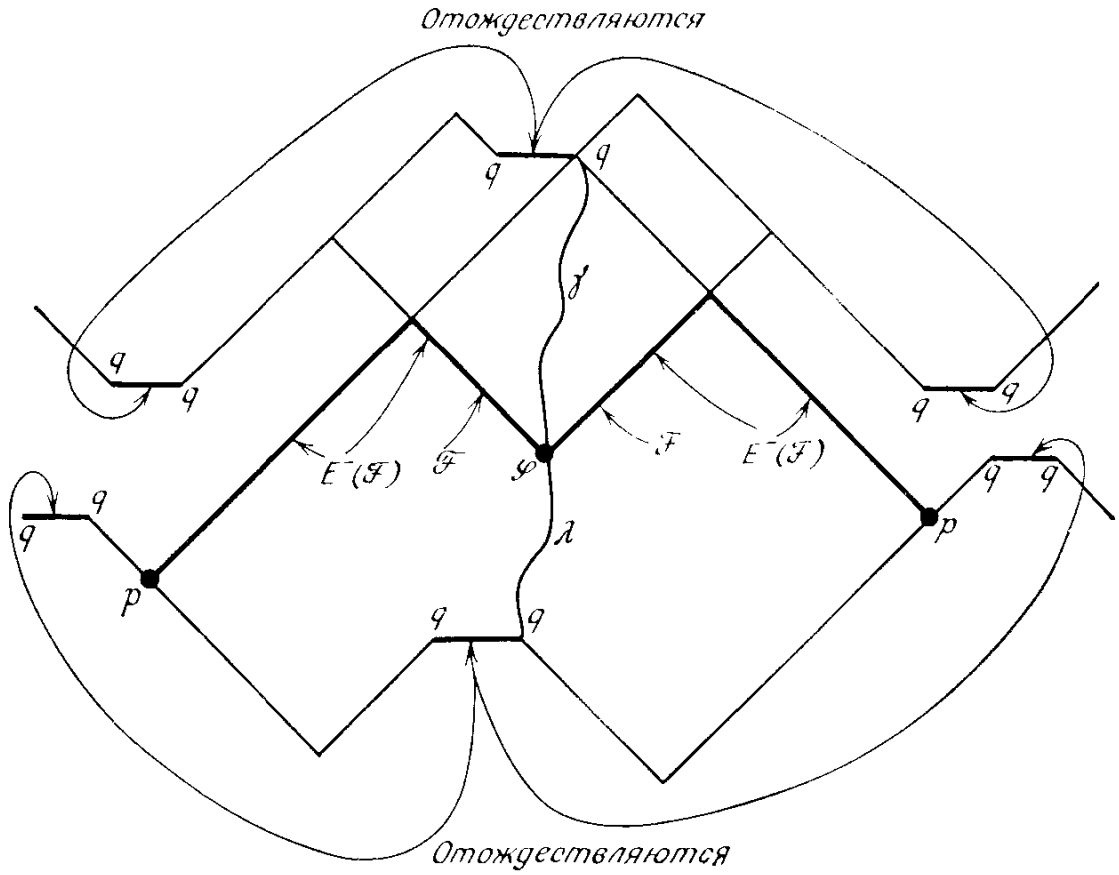


Рис. 52. Подобен рис. 51, но с отождествлением еще трех линий. $\mathscr{F}$ — множество $E^+(\mathscr{P}) \cap \overline{I}^-(\gamma)$. Точки p являются конечными точками прошлого изотропных образующих сегментов поверхности $E^-(\mathscr{F})$. Кривая λ — непродолжимая в прошлое времениподобная кривая, содержащаяся в $\operatorname{int} D^-(E^-(\mathscr{F}))$

Пусть a_n — бесконечная последовательность точек на λ, обладающая следующими свойствами:

I. $a_{n+1} \in I^-(a_n)$.

II. Любой компактный сегмент λ содержит лишь конечное число точек a_n.

Пусть b_n — аналогичная последовательность на γ, но с заменой I^- в свойстве I на I^+ и с $b_1 \in I^+(a_1)$.

Из того что γ и λ должны содержаться в глобально гиперболическом множестве $\operatorname{int} D(E^-(\mathscr{F}))$ (предложение 6.6.3), следует существование непространственноподобной геодезической μ_n максимальной длины между каждой точкой a_n и соответствующей точкой b_n (предложение 6.7.1). Любая из этих геодезических будет пересекать компактное множество $E^+(\mathscr{P})$. Следовательно, должна существовать $q \in E^+(\mathscr{P})$, которая будет предельной точ-

кой кривой на $\mu_n \cap E^+(\mathscr{S})$ и непространственноподобное направление в q, которое будет пределом направлений геодезических μ_n. (Точка q и направление в q задают точку в расслоении направлений над $\mathscr{M}$. Такая предельная точка существует в силу компактности той части расслоения, которая является расслоением направлений над $E^+(\mathscr{S})$.) Пусть μ'_n — подпоследовательность μ_n, такая, что $\mu'_n \cap E^+(\mathscr{S})$ сходится к q и направления μ'_n в $E^+(\mathscr{S})$ сходятся к предельному направлению. (Точнее, точки, определяемые кривой μ'_n в расслоении направлений над $E^+(\mathscr{S})$, сходятся к предельной точке.) Пусть μ будет непродолжимой геодезической через q в предельном направлении. По условию I теоремы, на μ существуют сопряженные точки x и y, причем $y \in I^+(x)$. Пусть точки x' и y' расположены на μ соответственно в прошлом и будущем точек x и y. Согласно предложению 4.5.8, найдутся $\varepsilon > 0$ и некоторая времениподобная кривая α от x' до y', длина которой равна ε плюс длина μ от x' до y'. Пусть $\mathscr{U}$ и $\mathscr{V}$ — нормальные выпуклые координатные окрестности точек x' и y' соответственно, причем в каждой из них нет ни одной кривой длины $^1/_1\varepsilon$. Пусть $x'' = \dot{\mathscr{U}} \cap \alpha$ и $y'' = \dot{\mathscr{V}} \cap \alpha$, а x'_n и y'_n — точки на μ'_n, сходящиеся соответственно к x' и y'. При достаточно больших n длина μ'_n от x'_n до y'_n будет меньше, чем $\varepsilon/4$ плюс длина μ от x' до y'. Кроме того, при достаточно больших n x'_n и y'_n будут лежать соответственно в $I^-(x'', \mathscr{U})$ и $I^+(y'', \mathscr{V})$. Тогда, двигаясь от x'_n к x'' вдоль α до y'' и затем от y'' до y'_n, мы получим непространственноподобную кривую, более длинную, чем μ'_n. Но в силу свойства II при достаточно больших n a'_n и b'_n будут лежат на μ'_n соответственно в прошлом точки x'_n и в будущем точки y'_n. Следовательно, μ'_n должна быть непространственноподобной кривой наибольшей длины от x'_n до y'_n. Мы приходим к противоречию, доказывающему теорему. □

Хотя эта теорема устанавливает существование сингулярностей при весьма общих предположениях, у нее есть один недостаток: она ничего не говорит о том, где эта сингулярность — в прошлом или в будущем. В случае (б) условия (4), когда имеется компактная пространственноподобная поверхность, нет никаких оснований считать, что будущее в этом отношении предпочтительнее прошлого, но в случае (а), когда имеется замкнутая ловушечная поверхность, можно думать, что сингулярность находится в будущем, а в случае (в), когда изотропный конус прошлого начинает стягиваться, можно полагать, что сингулярность — в прошлом. Если условие (в) несколько усилить, а именно потребовать, чтобы все направленные в прошлое времениподобные и изотропные геодезические начинали сходиться в пределах компактной области в $J^-(p)$, то можно показать, что сингулярность находится в прошлом.

Теорема 3 (*Хокинг* [73])
Если

1) $R_{ab}K^aK^b \geqslant 0$ для любого непространственноподобного вектора $\mathbf{K}$ (ср. разд. 4.3);

2) на $(\mathscr{M}, \mathbf{g})$ выполняется условие сильной причинности;

3) существуют направленный в прошлое единичный вектор $\mathbf{W}$ в точке p и положительная постоянная b, такие, что если $\mathbf{V}$ — единичный касательный вектор к направленной в прошлое, проходящей через p времениподобной геодезической, то на каждой такой геодезической расхождение $\theta \equiv V^a{}_{;a}$ этих геодезических становится меньше, чем $-3c/b$ на расстоянии b/c от p, где $c \equiv \equiv -W^aV_a$,

тогда через p проходит неполная в прошлое, непространственноподобная геодезическая.

Пусть K^a — параллельно перенесенный касательный вектор к направленной в прошлое, проходящей через p непространственноподобной геодезической со следующей нормировкой: $K^aW_a = -1$. Тогда для упомянутой в условии (б) времениподобной геодезической $K^a = c^{-1}V^a$, откуда $K^a{}_{;a} = c^{-1}V^a{}_{;a}$. Поскольку функция $K^a{}_{;a}$ непрерывна на непространственноподобной геодезической, она становится меньше, чем $-3/b$ на изотропной геодезической, проходящей через p, в пределах аффинного расстояния b. Если $\mathbf{Y}_1$, $\mathbf{Y}_2$, $\mathbf{Y}_3$ и $\mathbf{Y}_4$ — псевдоортонормированная тетрада на этих изотропных геодезических, причем $\mathbf{Y}_1$ и $\mathbf{Y}_2$ — единичные пространственноподобные векторы, а $\mathbf{Y}_3$ и $\mathbf{Y}_4$ — изотропные векторы, для которых $Y_3{}^aY_{4a} = -1$ и $\mathbf{Y}_4 = \mathbf{K}$, то расхождение $\hat{\theta}$ изотропных геодезических в p определяется формулой

$$\hat{\theta} = K_{a;b}\,(Y_1{}^aY_1{}^b + Y_2{}^aY_2{}^b) = K^a{}_{;a} + K_{a;b}\,(Y_3{}^aY_4{}^b + Y_4{}^aY_3{}^b).$$

Второй член равен нулю, так как K^a переносится параллельно. Третий член можно представить в виде $\frac{1}{2}(K_aK^a)_{;b}Y_3{}^b$; эта функция отрицательна, ибо $K^aK_a = 0$ на изотропных и $K^aK_a < 0$ — на времениподобных геодезических. Отсюда видно, что $\hat{\theta}$ становится меньше $-3/b$ в пределах аффинного расстояния b вдоль каждой изотропной геодезической из точки p. Таким образом, если бы все направленные из p в прошлое изотропные геодезические были бы полны, то контур (horismos) $E^-(p)$ был бы компактен. Любая точка $q \in J^-(E^-(p)) - E^-(p)$ при этом будет лежать в $I^-(p)$. Следовательно, она не может находиться в $J^+(E^-(p))$, поскольку $E^-(p)$ — ахрональное множество. Отсюда

$$J^+(E^-(p)) \cap J^-(E^-(p)) = E^-(p),$$

т. е. будет компактным множеством. Тогда по предложению 6.6.7 $D^-(E^-(p))$ будет глобально гиперболической областью. Согласно предложению 6.7.1, каждую точку $r \in D^-(E^-(p))$ можно со-

единить с p непространственноподобной геодезической, на которой между r и p не будет ни одной точки, сопряженной p. Отсюда в силу предложения 4.4.1 область $D^-(E^-(p))$ должна содержаться в $\exp_p(F)$, где F — компактная область в T_p, которая состоит из всех направленных в прошлое непространственноподобных векторов K^a, удовлетворяющих условию $K^aW_a \leqslant -2b$. Если бы все непространственноподобные геодезические из p в прошлое были полны, множество $\exp_p(K^a)$ было бы определено для любого $K^a \in F$ и было бы компактно как образ компактного множества при непрерывных отображениях. Однако, согласно следствию леммы 8.2.1, $D^-(E^-(p))$ содержит непродолжимую в прошлое времениподобную кривую. По предложению 6.4.7 она не может быть полностью заключена в компактном множестве $\exp_p(F)$, и, следовательно, предположение, что все направленные в прошлое из p непространственноподобные геодезические полны, не может быть верным. □

Теоремы 2 и 3 являются наиболее полезными теоремами о сингулярностях, ибо можно показать, что их условия удовлетворяются в ряде физических ситуаций (см. следующую главу). Однако не может ли оказаться, что эти теоремы свидетельствуют не о сингулярности, а о замкнутой времениподобной кривой, нарушающей условия причинности? Это было бы много хуже, чем просто потеря возможности делать предсказания, что является альтернативой теореме 1; по нашему мнению, такое нарушение причинности более неприемлемо, чем сингулярность. Тем не менее хотелось бы знать, может ли нарушение причинности воспрепятствовать появлению сингулярностей? Из нижеследующей теоремы следует, что в определенных ситуациях не может. Это означает, что сингулярности нужно принимать всерьез, и придает нам уверенность, что вообще нарушение причинности не является выходом из положения.

Теорема 4 (*Хокинг* [73])

Пространство-время не является времениподобно геодезически полным, если

1) $R_{ab}K^aK^b \geqslant 0$ для любого непространственноподобного вектора $\mathbf{K}$ (ср. с разд. 4.3);

2) существует компактная пространственноподобная гиперповерхность (без края);

3) единичные нормали к $\mathscr{S}$ всюду на $\mathscr{S}$ сходятся или расходятся.

Замечания. Условие (2) можно трактовать как утверждение о пространственной замкнутости Вселенной, а условие (3) — как утверждение о том, что она расширяется (сжимается). Как было объяснено в разд. 6.5, мы можем взять накрывающее многообра-

зие $\hat{\mathscr{M}}$, в котором каждая связная компонента образа поверхности $\mathscr{S}$ диффеоморфна $\mathscr{S}$ и является частичной поверхностью Коши в $\hat{\mathscr{M}}$. Мы будем работать с многообразием $\hat{\mathscr{M}}$ и обозначим через $\hat{\mathscr{S}}$ одну связную компоненту образа $\mathscr{S}$. Рассматривая задачу Коши в $\hat{\mathscr{M}}$, мы видим, что наличие сингулярностей (но не обязательно их характер) является устойчивым свойством данных Коши на $\hat{\mathscr{S}}$, поскольку достаточно малые вариации данных на $\hat{\mathscr{S}}$ не нарушают условия (3). Это служит контрпримером к предположению Лифшица и Халатникова, что сингулярности появляются лишь для множества данных Коши меры нуль, хотя следует помнить, что определение сингулярности, которому мы следуем здесь, не совпадает с тем, которым пользуются Лифшиц и Халатников.

Доказательство. По условиям (2) и (3) свертка второй фундаментальной формы поверхности $\hat{\mathscr{S}}$ имеет на $\mathscr{S}$ отрицательную верхнюю грань. Таким образом, если бы $\mathscr{M}$ (и значит, $\hat{\mathscr{M}}$) было времениподобно геодезически полно, то на каждой направленной в будущее геодезической, ортогональной к $\hat{\mathscr{S}}$ в пределах конечной верхней грани b расстояния от $\hat{\mathscr{S}}$ (предложение 4.4.7) существовала бы точка, сопряженная $\hat{\mathscr{S}}$. Но, согласно дополнению к предложению 6.7.1, до каждой точки $q \in D^+(\hat{\mathscr{S}})$ можно провести направленную в будущее геодезическую, ортогональную к $\hat{\mathscr{S}}$, на которой между q и $\hat{\mathscr{S}}$ нет ни одной точки, сопряженной поверхности $\hat{\mathscr{S}}$. Пусть $\beta: \hat{\mathscr{S}} \times [0, b] \to \hat{\mathscr{M}}$ — дифференцируемое отображение, которое смещает точку $p \in \hat{\mathscr{S}}$ и расстояние $s \in [0, b]$ вдоль направленной в будущее геодезической, ортогональной к $\hat{\mathscr{S}}$ в точке p. Тогда множество $\beta: \hat{\mathscr{S}} \times [0, b]$ компактно и содержит $D^+(\hat{\mathscr{S}})$. Отсюда компактны $\bar{D}^+(\hat{\mathscr{S}})$ и $H^+(\hat{\mathscr{S}})$. Если бы мы предполагали выполнение условия строгой причинности, то искомое противоречие следовало бы из леммы 8.2.1. Однако даже и без этого условия можно прийти к противоречию. Рассмотрим точку $q \in H^+(\hat{\mathscr{S}})$. Поскольку каждая направленная в прошлое непространственноподобная кривая от q до $\hat{\mathscr{S}}$ будет состоять из (возможно, нулевого) изотропного геодезического сегмента в $H^+(\hat{\mathscr{S}})$ и из непространственноподобной кривой в $D^+(\hat{\mathscr{S}})$, функция расстояния $d(\hat{\mathscr{S}}, q)$ будет меньше или равна b. Отсюда, поскольку d полунепрерывна снизу, можно было бы найти бесконечную последовательность точек $r_n \in D^+(\hat{\mathscr{S}})$, сходящихся к q, для которых $d(\hat{\mathscr{S}}, r_n)$ будет сходиться к $d(\hat{\mathscr{S}}, q)$. Каждой r_n будет соответствовать по крайней мере один элемент $\beta^{-1}(r_n)$ из $\hat{\mathscr{S}} \times [0, b]$. Ввиду компактности $\hat{\mathscr{S}} \times [0, b]$ найдется элемент (p, s),

который будет предельной точкой последовательности $\beta^{-1}(r_n)$. В силу непрерывности $s = d(\hat{\mathscr{S}}, q)$ и $\beta(p, s) = q$. Таким образом, к каждой точке $q \in H^+(\hat{\mathscr{S}})$ можно будет от $\hat{\mathscr{S}}$ провести геодезическую длиной $d(\hat{\mathscr{S}}, q)$. Теперь пусть $q_1 \in H^+(\hat{\mathscr{S}})$ лежит в прошлом точки q на той же изотропной геодезической образующей λ поверхности $H^+(\hat{\mathscr{S}})$. Соединив геодезическую длиной $d(\hat{\mathscr{S}}, q_1)$ от $\hat{\mathscr{S}}$ до q_1 с сегментом λ между q_1 и q, мы получим непространственноподобную кривую $d(\hat{\mathscr{S}}, q_1)$ от $\hat{\mathscr{S}}$ до q, которую можно варьировать так, чтобы получить более длинную кривую между теми же конечными точками (предложение 4.5.10). Следовательно, $d(\hat{\mathscr{S}}, q)$, $q \in H^+(\hat{\mathscr{S}})$ будет строго убывающей функцией вдоль каждой направленной в прошлое образующей $H^+(\hat{\mathscr{S}})$. Но это ведет к противоречию, так как в силу полунепрерывности снизу функции $d(\hat{\mathscr{S}}, q)$ она должна иметь минимум на компактном множестве $H^+(\hat{\mathscr{S}})$. □

Условие (2) компактности $\mathscr{S}$ необходимо, так как в пространстве Минковского $(\mathscr{M}, \eta)$ некомпактная поверхность $\mathscr{S}: (x^1)^2 + (x^2)^2 + (x^3) - (x^4)^2 = -1$, $x^4 < 0$ является частичной поверхностью Коши с $\chi^a{}_a = -3$ во всех точках. Если возьмем область пространства Минковского, задаваемую неравенствами

$$x^4 < 0, \quad (x^1)^2 + (x^2)^2 + (x^3)^2 + (x^3)^2 - (x^4)^2 < 0,$$

мы можем отождествить точки, полученные действием дискретной группы изометрий G, так, что $\mathscr{S}/G$ будет компактно [102]. Согласно теореме 4, пространство $(\mathscr{M}/G, \eta)$ будет времениподобно геодезически неполным, поскольку отождествление относительно действия группы G не может быть распространено на все многообразие $\mathscr{M}$ [ни условие (1), ни условие (2) разд. 5.8 не выполняются в начале координат]. В этом случае сингулярность в смысле неполноты возникает из плохих глобальных свойств и не сопровождается сингулярностью кривизны. Этот пример был предложен Пенроузом.

Условия (2) и (3) можно заменить на следующие:

2′) $\hat{\mathscr{S}}$ есть поверхность Коши для $\hat{\mathscr{M}}$;

3′) $\chi^a{}_a$ имеет на $\mathscr{S}$ грань, отличную от нуля.

Хотя в этом случае не может быть горизонта Коши, однако длины всех направленных в будущее времениподобных кривых из $\hat{\mathscr{S}}$ должны иметь некоторую конечную верхнюю грань.

Герок [50] показал, что при сохранении условия (2) и замене условий (1) и (3) на условия

1″) $R_{ab}K^aK^b \geqslant 0$ для любого непространственноподобного вектора, причем равенство имеет место лишь при $R_{ab} = 0$;

3″) существует такая точка $p \in \hat{\mathscr{P}}$, что любая непродолжимая непространственноподобная кривая, пересекающая $\hat{\mathscr{P}}$, пересечет также $J^+(p)$ и $J^-(p)$; имеет место одно из двух: либо область зависимости поверхности $\hat{\mathscr{P}}$ плоская, либо многообразие $\hat{\mathscr{M}}$ времениподобно геодезически неполно.

Условие (3″) требует, чтобы наблюдатель в точке p мог увидеть любую частицу, которая пересечет $\mathscr{S}$ и чтобы он сам был видим с нее. Метод доказательства состоит в том, чтобы рассмотреть пространственноподобные поверхности без края, которые содержат точку p. Из этих поверхностей можно построить топологическое пространство $S(p)$ подобно тому, как мы строили топологическое пространство из всех непространственноподобных кривых между двумя точками. Условия (2) и (3″) означают тогда, что $S(p)$ компактно. Можно показать, что площади поверхностей являются полунепрерывными сверху функциями в $S(p)$ и должна быть некоторая поверхность $\mathscr{S}'$, проходящая через p, площадь которой больше или равна площади любой другой поверхности. Из вариационных соображений, аналогичных тем, что приводились в случае непространственноподобных кривых, следует, что $\chi^a{}_a$ обращается в нуль всюду на $\mathscr{S}'$, за исключением, быть может, точки p, где поверхность может и не быть дифференцируемой.

Рассмотрим однопараметрическое семейство пространственноподобных поверхностей $\mathscr{S}(u)$, где $\mathscr{S}(0) = \mathscr{S}'$. Вектор вариации $\mathbf{W} \equiv \partial/\partial u$ можно представить как $f\mathbf{n}$, где $\mathbf{n}$ — единичная нормаль к поверхностям, f — некоторая функция. Используя уравнение Райчаудхури для конгруэнции интегральных кривых векторного поля $\mathbf{W}$, можно показать, что

$$\frac{\partial\theta}{\partial u} = f\left\{-\frac{1}{3}\theta^2 - 2\sigma^2 - R_{ab}n^a n^b + f^{-1}f_{;ab}h^{ab}\right\},$$

где

$$\theta \equiv \chi^a{}_a, \quad \sigma_{ab} \equiv \chi_{ab} - \frac{1}{3}\theta h_{ab}, \quad h_{ab} \equiv g_{ab} + n_a n_b$$

и

$$\sigma^2 = \frac{1}{2}\sigma_{ab}\sigma^{ab}.$$

Если существует точка $q \in \mathscr{S}'$, в которой $R_{ab}n^a n^b \neq 0$ или $\chi_{ab} \neq 0$, мы можем подобрать такую функцию f, при которой $(\partial\theta/\partial u) < 0$ всюду на $\mathscr{S}'$. Если $R_{ab}n^a n^b = 0$, $\chi_{ab} = 0$ всюду на $\mathscr{S}'$, но при этом имеется некоторая точка $q \in \mathscr{S}'$, в которой $C_{abcd}n^b n^d \neq 0$, то тогда $\partial\sigma/\partial u \neq 0$ и можно найти такую функцию f, что $\partial\theta/\partial u = 0$ и $\partial^2\theta/\partial u^2 < 0$ всюду на $\mathscr{S}'$. В любом случае можно построить поверхность $\mathscr{S}''$, на которой всюду $\chi^a{}_a < 0$ и, следовательно, многообразие $\hat{\mathscr{M}}$ будет изотропно геодезически не-

полным в силу теоремы 4. Если R_{ab}, χ_{ab} и $C_{abcd}n^b n^d$ равны нулю всюду на $\mathscr{S}'$, то из тождеств Риччи для n^a видно, что $C_{abcd} = 0$ на $\mathscr{S}'$, т. е. в области $D(\mathscr{S})$ пространство-время является плоским. Примером, в котором условия (1″), (2) и (3″) выполнены, а область $D(\mathscr{S})$ плоская, служит пространство Минковского с отождествлением $\{x^1, x^2, x^3, x^4\}$ с $\{x^1+1, x^2, x^3, x^4\}$, $\{x^1, x^2+1, x^3, x^4\}$ и $\{x^1, x^2, x^3+1, x^4\}$. Оно геодезически полно. Однако приведенный ранее пример тоже удовлетворяет этим условиям и свидетельствует о том, что область $D(\mathscr{S})$ может быть одновременно геодезически неполной и плоской.

8.3. Описание сингулярностей

Приведенные выше теоремы свидетельствуют о наличии сингулярностей в широком классе решений, но содержат мало информации о природе этих сингулярностей. Чтобы исследовать этот вопрос более детально, нужно определить, что́ подразумевать под размерами, формой, местонахождением и прочими характеристиками сингулярности. Это было бы совсем легко, если бы сингулярные точки включались в пространственно-временное многообразие. Однако установление структуры многообразия в таких точках физическими измерениями было бы невозможно. В самом деле, может существовать много структур многообразия, совпадающих в регулярных областях, но различающихся в сингулярных точках. К примеру, в решении Робертсона — Уокера многообразие в окрестности сингулярности $t = 0$ можно описать координатами

$$\{t,\ r\cos\theta,\ r\sin\theta\cos\phi,\ r\sin\theta\sin\phi\}$$

или координатами

$$\{t,\ Sr\cos\theta,\quad Sr\sin\theta\cos\phi,\quad Sr\sin\theta\sin\phi\}.$$

В первом случае сингулярностью будет трехмерная поверхность, а во втором — единственная точка.

Выход заключается в том, чтобы разработать рецепт построения некоторой границы ∂ многообразия $\mathscr{M}$, которая однозначно устанавливается измерениями в несингулярных точках, т. е. структурой $(\mathscr{M}, \mathbf{g})$. Затем было бы желательно задать по меньшей мере топологию, а может быть, и дифференцируемую структуру в пространстве $\mathscr{M}^+ \equiv \mathscr{M} \cup \partial$. Одним из подходов мог бы служить метод неразложимых множеств для бесконечности, описанный в разд. 6.9. Но этот метод зависит лишь от конформной метрики и потому не позволяет отличить бесконечность от сингулярных точек на конечном расстоянии. Чтобы такое различие существовало, по-видимому, нужно при построении $\mathscr{M}^+$

исходить из того критерия, которым мы пользовались при доказательстве существования сингулярностей, а именно из b-неполноты. Изящный метод такого построения был разработан Шмидтом. Этим методом перекрываются более ранние конструкции Хокинга [72] и Герока [53], в которых сингулярные точки определяются как классы эквивалентности неполных геодезических. В этих конструкциях не требуется, чтобы все b-неполные кривые (скажем, неполные времениподобные кривые с ограниченным ускорением) обязательно имели конечные точки. В определении классов эквивалентности имеется некоторая неоднозначность. Конструкция Шмидта не страдает этими недостатками.

Метод Шмидта состоит в задании положительно определенной метрики $\mathbf{e}$ на расслоении ортонормированных реперов $\pi : O(\mathscr{M}) \to \mathscr{M}$. Здесь $O(\mathscr{M})$ — множество всех ортонормированных четверок векторов $\{\mathbf{E}_a\}$, $\mathbf{E}_a \in T_p$ для каждого $p \in \mathscr{M}$ (a пробегает значения от 1 до 4), а π — проекция, которая отображает базис в точке p в саму точку p. Оказывается, что множество $O(\mathscr{M})$ неполно по метрике $\mathbf{e}$, если и только если многообразие $\mathscr{M}$ b-неполно. Если $O(\mathscr{M})$ m-неполно, можно с помощью фундаментальных последовательностей образовать пополнение $\overline{O(\mathscr{M})}$ до метрического пространства. Проекция π может быть продолжена на $\overline{O(\mathscr{M})}$, и тогда фактор-пространство $\overline{O(\mathscr{M})}$ по π есть по определению многообразие $\mathscr{M}^+$, которое является объединением $\mathscr{M}$ с множеством дополнительных точек ∂. Множество ∂ состоит из сингулярных точек многообразия $\mathscr{M}$ в том смысле, что оно является множеством конечных точек всех b-неполных кривых.

Чтобы выполнить построение, вспомним (разд. 2.9), что связность на $\mathscr{M}$, задаваемая метрикой $\mathbf{g}$, определяет четырехмерное *горизонтальное подпространство* H_u десятимерного касательного пространства T_u в точке $u \in O(\mathscr{M})$. Таким образом, T_u представляет собой прямую сумму подпространства H_u и вертикального подпространства V_u, состоящего из всех векторов в T_u, которые касательны к слою $\pi^{-1}(\pi(u))$. Теперь мы построим базис $\{\mathbf{G}_A\} = \{\bar{\mathbf{E}}_a, \mathbf{F}_i\}$ для T_u, где A пробегает значения от 1 до 10, а i от 1 до 6; $\{\bar{\mathbf{E}}_a\}$ есть базис для H_u, и $\{\mathbf{F}_i\}$ — базис для V_u.

Для данного вектора $\mathbf{X} \in T_{\pi(u)}(\mathscr{M})$ существует единственный вектор $\bar{\mathbf{X}} \in H_u(O(\mathscr{M}))$, удовлетворяющий условию $\pi_* \bar{\mathbf{X}} = \mathbf{X}$. Следовательно, на $O(\mathscr{M})$ однозначно определены четыре горизонтальных векторных поля $\bar{\mathbf{E}}_a$, которые являются горизонтальными лифтами ортонормированных базисных векторов $\mathbf{E}_a$ для каждой точки $u \in O(\mathscr{M})$. Интегральные кривые поля $\bar{\mathbf{E}}_a$ в $O(\mathscr{M})$ соответствуют параллельному перенесению базиса $\{\mathbf{E}_a\}$ в $\mathscr{M}$ вдоль геодезической в направлении вектора $\mathbf{E}_a$. Группа $O(3, 1)$, мультипликативная группа всех несингулярных действительных

лоренцевых 4×4-матриц A_{ab}, действует в слоях θ множества $O(\mathscr{M})$, переводя точку $u=\{p,\ \mathbf{E}_a\}\in O(\mathscr{M})$ в точку $A(u)=\{p, A_{ab}\mathbf{E}_b\}\in O(\mathscr{M})$. Можно рассматривать O (3, 1) как шестимерное многообразие и при этом описывать касательное пространство $T_I(O(3,\ 1))$ к $O(3,\ 1)$ в единице группы I векторным пространством всех 4×4-матриц a, обладающих свойством $a_{ab}G_{bc}=-a_{cb}G_{ba}$. Тогда при $a\in T_I(O(3,\ 1))$ мы можем определить кривую в $O(3,\ 1)$ через $A_t=\exp(ta)$, где

$$\exp(b)=\sum_{n=0}^{\infty}\frac{b^n}{n!}.$$

Отсюда, если $u\in O(\mathscr{M})$, то в $\pi^{-1}(\pi(u))$ можно задать кривую, проходящую через u, полагая $\lambda_{au}(t)=A_t(u)$. Поскольку кривая $\lambda_{au}(t)$ лежит в слое, ее касательный вектор $(\partial/\partial t)_{\lambda_{au}}$ вертикален. Поэтому для каждого $a\in T_I$ можно определить вертикальное векторное поле $\mathbf{F}(a)$:

$$\mathbf{F}(a)\,|_u=(\partial/\partial t)_{\lambda_{au}|\,u}$$

для всякого $u\in O(\mathscr{M})$. Пусть $\{a_i\}$ $(i=1,\ 2,\ \ldots,\ 6)$ есть базис в T_I, тогда $\mathbf{F}_i\equiv\mathbf{F}(a_i)$ представляют собой шесть векторных полей на $O(\mathscr{M})$ и могут служить базисом для V_u в каждой точке $u\in O(\mathscr{M})$.

Матрица $B\in O(3,\ 1)$ задает отображение $O(\mathscr{M})\to O(\mathscr{M})$ вида $u\to B(u)$. При индуцированном отображении $B_*: T_u\to T_{B(u)}$ вертикальные и горизонтальные векторные поля преобразуются следующим образом:

$$B_*(\bar{\mathbf{E}}_a)=B^{-1}_{ab}\bar{\mathbf{E}}_b,$$
$$B_*(\mathbf{F}_i)=C_i{}^j\mathbf{F}_j,$$

где $C_i{}^j=B_{ab}a_{ibc}B^{-1}{}_{cd}a^j{}_{da}$ и $\{a^j\}$ — базис в T^*_I, дуальный базису $\{a_i\}$ в T_I (следовательно, $a^i{}_{ab}a_{jab}=\delta^i{}_j$, $a^i{}_{ab}a_{jcd}=\delta_{ac}\delta_{bd}$). Для дальнейшего будет важна не конкретная форма этих индуцированных отображений, а их свойство быть постоянными на $O(\mathscr{M})$.

Теперь у нас есть базис $\{\mathbf{G}_A\}=\{\bar{\mathbf{E}}_a,\ \mathbf{F}_i\}$ $(A=1,\ \ldots,\ 10)$ в T_u в каждой точке $u\in O(\mathscr{M})$. Поэтому мы можем задать положительно-определенную метрику $\mathbf{e}$ на $O(\mathscr{M})$ соотношением $e(\mathbf{X},\ \mathbf{Y})=\sum\limits_A X^AY^A$, где $\mathbf{X},\ \mathbf{Y}\in T(u)$ и $X^A,\ Y^A$ — компоненты $\mathbf{X}$ и $\mathbf{Y}$ в базисе $\{\mathbf{G}_A\}$. Используя метрику $\mathbf{e}$, можно ввести функцию расстояния $p(u,\ v)$, где $u,\ v\in O(\mathscr{M})$, как нижнюю грань длин кривых от u до v (измеренных по метрике $\mathbf{e}$). Тогда можно поставить вопрос: является ли $O(\mathscr{M})$ полным множеством относительно функции расстояния ρ?

Предложение 8.3.1

$(O(\mathscr{M}), \mathbf{e})$ является m-полным пространством тогда и только тогда, когда пространство-время $(\mathscr{M}, \mathbf{g})$ b-полно.

Пусть $\gamma(t)$ — некоторая кривая в $\mathscr{M}$. Тогда, если задана точка $u \in \pi^{-1}(p)$, где $p \in \gamma$, можно построить горизонтальную кривую $\bar{\gamma}(t)$, которая проходит через u и для которой $\pi(\bar{\gamma}(t)) = \gamma(t)$. Из определения положительно-определенной метрики $\mathbf{e}$ следует, что длина дуги $\bar{\gamma}(t)$, измеренная по этой метрике, равна обобщенному аффинному параметру $\gamma(t)$, определяемому тем базисом в точке p, изображением которого является точка u. Поэтому, если у $\gamma(t)$ нет конечной точки, но длина ее, измеренная по обобщенному аффинному параметру, конечна, то у $\bar{\gamma}(t)$ тоже нет конечной точки, но конечна длина по метрике $\mathbf{e}$. Таким образом, m-полнота множества $O(\mathscr{M})$ означает b-полноту в многообразии $\mathscr{M}$.

Чтобы доказать обратное, нужно показать, что для C^1-кривой $\lambda(t)$ конечной длины в $O(\mathscr{M})$, не имеющей конечной точки, $\pi(\lambda(t))$ является C^1-кривой в $\mathscr{M}$ со следующими свойствами:

1) у нее конечная аффинная длина;

2) у нее нет конечных точек в $\mathscr{M}$.

Свойство (1) доказывается следующим образом. Пусть $u \in \lambda(t)$. Тогда через u можно провести горизонтальную кривую $\bar{\lambda}(t)$, такую, что $\pi(\bar{\lambda}(t)) = \pi(\lambda(t))$. Для каждого значения t $\lambda(t)$ и $\bar{\lambda}(t)$ будут лежать в одном и том же слое, так что в $O(3, 1)$ будет единственная кривая $B(t)$, для которой $\lambda(t) = B(t)\bar{\lambda}(t)$. Это означает, что

$$\left(\frac{\partial}{\partial t}\right)_{\lambda} = B_* \left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}} + F(\dot{B}B^{-1}),$$

где $\dot{B} \equiv \mathrm{d}B/\mathrm{d}t$. Отсюда

$$e\left(\left(\frac{\partial}{\partial t}\right)_{\lambda}, \left(\frac{\partial}{\partial t}\right)_{\lambda}\right) = \sum_b \left(\left\langle \bar{\mathbf{E}}^a, \left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}} \right\rangle B^{-1}{}_{ab}\right)^2 + \\ + \sum_i \left(\dot{B}_{ab} B^{-1}{}_{bc} a^t{}_{ca}\right)^2,$$

где $\{\bar{\mathbf{E}}^a\}$ — базис $H^*{}_u$, дуальный базису $\{\bar{\mathbf{E}}_a\}$ (т. е. $\langle \bar{\mathbf{E}}^a, \bar{\mathbf{E}}_b \rangle = \delta^a{}_b$), и $a^i{}_{ab}$ — базис T_I^*, дуальный базису $a_{i\,ab}$ (т. е. $a_{i\,ab} a^j{}_{ab} = \delta_i{}^j$).

Матрица B_{ab} удовлетворяет равенству $B_{ab} G_{bc} B_{dc} = G_{ad}$. Отсюда

$$B_{ab} G_{ac} B_{cd} = G_{bd},$$

так как $G_{ab} = G^{-1}{}_{ab}$. Дифференцируя по t, имеем

$$\dot{B}_{ab} B^{-1}{}_{bc} G_{cd} = - G_{ac} \dot{B}_{db} B^{-1}{}_{bc}.$$

Таким образом, $\dot{B}_{ab}B^{-1}{}_{bc} \in T_I(O(3,1))$. Поскольку матрицы $a^i{}_{at}$ образуют базис $T^*{}_I$, существует такая постоянная C, что

$$\sum_i (\dot{B}_{ab}B^{-1}{}_{bc}a^i{}_{ca})^2 \geqslant C(\dot{B}_{ab}B^{-1}{}_{bc}\dot{B}_{ad}B^{-1}{}_{dc}).$$

Любая матрица $B \in O(3,1)$ может быть представлена в виде $B = \bar{\Omega}\Delta\Omega$, где

$$\Delta = \begin{pmatrix} \operatorname{ch}\xi & 0 & 0 & \operatorname{sh}\xi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \operatorname{sh}\xi & 0 & 0 & \operatorname{ch}\xi \end{pmatrix}$$

и $\bar{\Omega}$ и Ω — ортогональные матрицы вида

$$\left(\begin{array}{c|c} \bar{O} & \\ \hline & 1 \end{array}\right) \quad \text{и} \quad \left(\begin{array}{c|c} O & \\ \hline & 1 \end{array}\right),$$

причем $\bar{O}$ и O — ортогональные 3×3-матрицы, а векторы базиса $\{\mathbf{E}_a\}$ пронумерованы так, что $\mathbf{E}_4$ — времениподобный вектор. Матрица Δ соответствует изменению скорости в направлении x^1, а матрицы $\bar{\Omega}$ и Ω соответствуют вращениям. При таком разложении

$$\dot{B}_{ab}B^{-1}{}_{bc}\dot{B}_{ad}B^{-1}{}_{dc} \geqslant 2(\dot{\xi})^2.$$

Для любого векторного поля $\mathbf{X} \in T_u$

$$\sum_b (\langle \bar{\mathbf{E}}^a, \mathbf{X}\rangle \Omega_{ab})^2 = \sum_a (\langle \bar{\mathbf{E}}^a, \mathbf{X}\rangle)^2.$$

Следовательно,

$$\sum_b \left(\left\langle \bar{\mathbf{E}}^a, \left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}}\right\rangle B^{-1}{}_{ab}\right)^2 \geqslant \sum_a \left(\left\langle \bar{\mathbf{E}}^a, \left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}}\right\rangle\right)^2 e^{-2|\xi|} =$$

$$= e\left(\left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}}, \left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}}\right) e^{-2|\xi|}.$$

Отсюда

$$e\left(\left(\frac{\partial}{\partial t}\right)_{\lambda}, \left(\frac{\partial}{\partial t}\right)_{\lambda}\right) \geqslant e\left(\left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}}, \left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}}\right) e^{-2|\xi|} + 2C(\dot{\xi})^2,$$

и поэтому

$$\left[e\left(\left(\frac{\partial}{\partial t}\right)_{\lambda}, \left(\frac{\partial}{\partial t}\right)_{\lambda}\right)\right]^{1/2} \geqslant \frac{1}{2}\left[e\left(\left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}}, \left(\frac{\partial}{\partial t}\right)_{\bar{\lambda}}\right)\right]^{1/2} e^{-|\xi|} + C^{1/2}|\dot{\xi}|.$$

Пусть $\xi_0 \leqslant \infty$ — наименьшая верхняя грань для $|\xi|$ на $\lambda(t)$. Тогда

$$L(\lambda) \geqslant \frac{1}{2} L(\bar{\lambda}) e^{-\xi_0} + C^{1/2}\xi_0,$$

где $L(\lambda)$ — длина кривой λ по метрике $\mathbf{e}$. Так как она конечна, должны быть конечны ξ_0 и $L(\bar{\lambda})$. Следовательно, аффинная длина кривой $\pi(\lambda(t))$ в $\mathscr{M}$, равная $L(\bar{\lambda})$, будет конечна.

В завершение доказательства предложения 8.3.1 мы должны показать, что нет такой точки $p \in \mathscr{M}$, в любую окрестность $\mathscr{U}$ которой входит и остается там кривая $\pi(\lambda(t))$. Поскольку точка p обладает нормальной окрестностью, это вытекает из следующего утверждения:

Предложение 8.3.2 (*Шмидт* [*147*])
Пусть $\mathscr{N}$ — компактное подмножество многообразия $\mathscr{M}$. Допустим, что в $O(\mathscr{M})$ имеется кривая $\lambda(t)$ конечной длины и без конечных точек, которая входит в $\pi^{-1}(\mathscr{N})$ и остается там. Тогда существует непродолжимая изотропная геодезическая γ, целиком содержащаяся в $\mathscr{N}$.

Пусть $\bar{\lambda}(t)$ — горизонтальная кривая, проходящая через некоторую точку $u \in \lambda(t)$, причем $\pi(\bar{\lambda}(t)) = \pi(\lambda(t))$. Кривая $\lambda(t)$ не имеет конечной точки. Допустим, что есть точка $v \in O(\mathscr{M})$, которая является конечной точкой горизонтальной кривой $\bar{\lambda}(t)$. Тогда должна существовать такая открытая окрестность $\mathscr{W}$ точки v с компактным замыканием, что $\bar{\lambda}(t)$ входит в нее и остается там. Пусть $\mathscr{W}'$ — множество $\{x \in O(\mathscr{M}) : Bx \in \mathscr{W}$ для всех матриц B с $|\xi| \leqslant \xi_0\}$. В силу компактности $\overline{\mathscr{W}}$ и ограниченности ξ_0 множество $\overline{\mathscr{W}'}$ должно быть компактным. Кривая $\lambda(t)$ войдет в $\overline{\mathscr{W}'}$ и останется в нем. Но любое компактное множество m-полно относительно положительно определенной метрики $\mathbf{e}$. Следовательно, кривая $\lambda(t)$, обладая конечной длиной, должна будет иметь конечную точку в $\overline{\mathscr{W}'}$. Из этого видно, что у кривой $\bar{\lambda}(t)$ нет конечной точки.

Пусть $\{x_n\}$ — последовательность точек на $\bar{\lambda}(t)$, не имеющая какой-либо предельной точки. Поскольку $\mathscr{N}$ — компактное множество, найдется точка $x \in \mathscr{N}$, которая является предельной точкой $\pi(x_n)$. Пусть $\mathscr{U}$ — нормальная окрестность точки x с компактным замыканием, и пусть $\sigma : \mathscr{U} \to O(\mathscr{M})$ — сечение $O(\mathscr{M})$ над $\mathscr{U}$, т. е. $\sigma(p)$, $p \in \mathscr{U}$, — ортонормированный базис в p. Пусть $\tilde{\lambda}(t) \equiv \sigma(\pi(\lambda(t)))$ при $\lambda(t) \in \pi^{-1}(\mathscr{U})$. Тогда, как и в предыдущем предложении, будет существовать единственное семейство матриц $A(t) \in O(3, 1)$, такое, что $\bar{\lambda}(t) = A(t)\tilde{\lambda}(t)$ и матрицы A можно представить в виде $A = \bar{\Omega}\Delta\Omega$. Допустим, что $|\xi(t_{n'})|$ имеет конечную верхнюю грань ξ_1, где $x_{n'} = \bar{\lambda}(t_{n'})$ — подпоследовательность последовательности x_n, сходящаяся к x. Тогда точки $x_{n'}$ должны принадлежать множеству $\mathscr{U}' = \{v \in O(\mathscr{M}) : A^{-1}v \subset \sigma(\mathscr{U})$ при некоторой матрице $A \in O(3, 1)$ с $|\xi| < \xi_1\}$.

Однако множество $\overline{\mathcal{U}'}$ должно быть компактным и потому должно содержать предельную точку последовательности $\{x_{n'}\}$, что противоречит нашему выбору $\{x_n\}$. Таким образом, $|\xi(t_{n'})|$ не имеет конечной верхней грани. Поскольку ортогональная группа компактна, мы можем выбрать такую подпоследовательность $\{x_{n''}\}$, что $\bar{\Omega}_{n''}$ будет сходиться к некоторой матрице $\bar{\Omega}'$, а $\Omega_{n''}$ — к некоторой Ω' (здесь $\bar{\Omega}_{n''} = \bar{\Omega}(t_{n''})$ и т. д.); $\xi_{n''} \to \infty$ и

$$\xi_{n''+1} - \xi_{n''} > a > 0 \tag{8.1}$$

для некоторой постоянной a.

Пусть

$$\lambda'(t) = (\bar{\Omega}')^{-1}\bar{\lambda}(t) \quad \text{и} \quad \hat{\lambda}_{n''}(t) \equiv \Delta_{n''}{}^{-1}(\bar{\Omega}')^{-1}\bar{\lambda}(t).$$

Тогда $\hat{\lambda}_{n''}(t_{n''})$ стремится к $\hat{x} \equiv \Omega'\sigma(x)$. Поскольку длина кривой $\bar{\lambda}(t)$ конечна, кривая $\lambda'(t)$ тоже имеет конечную длину. Это означает, что стремится к нулю интеграл

$$\int_{t_{n''}}^{t_{n''+1}} ((X^u)^2 + (X^v)^2 + (X^2)^2 + (X^3)^2)^{1/2}\, dt,$$

где

$$X^A \equiv \left\langle \bar{\mathbf{E}}^A, \left(\frac{\partial}{\partial t}\right)_{\lambda'} \right\rangle, \quad A = u, v, 2, 3$$

и

$$\bar{\mathbf{E}}^u = \frac{1}{\sqrt{2}}(\bar{\mathbf{E}}^4 + \bar{\mathbf{E}}^1), \quad \bar{\mathbf{E}}^v = \frac{1}{\sqrt{2}}(\bar{\mathbf{E}}^4 - \bar{\mathbf{E}}^1).$$

Следовательно,

$$\int_{t_{n''}}^{t_{n''+1}} |X^A|\, dt$$

стремится к нулю при каждом A. Компоненты $Y_{n''}{}^A$ касательного вектора к горизонтальной кривой $\hat{\lambda}_{n''}(t)$ равны

$$Y_{n''}{}^u = e^{-\xi_{n''}}X^u, \quad Y_{n''}{}^v = e^{\xi_{n''}}X^v, \quad Y_{n''}{}^2 = X^2, \quad Y_{n''}{}^3 = X^3.$$

Поэтому

$$\int_{t_{n''}}^{t_{n''+1}} |\mathbf{Y}_{n''}{}^A|\, dt \quad (A = u, 2, 3) \tag{8.2}$$

стремится к нулю.

Пусть μ — интегральная кривая горизонтального векторного поля $\bar{\mathbf{E}}^v$, проходящая через $\hat{x}$. Тогда $\pi(\mu)$ будет изотропной геодезической в $\mathcal{M}$. Допустим, что $\pi(\mu)$ выходит из $\mathcal{N}$ как в

направлении будущего, так и в направлении прошлого. Тогда найдется некоторая окрестность $\mathscr{V}$ точки $\hat{x}$ с компактным замыканием и с тем свойством, что в каждом из этих направлений μ покидает множество $\overline{\mathscr{V}'}$, где $\mathscr{V}' \equiv \{v \in O(\mathscr{M}) :$ существует Δ, для которой $\Delta v \in \mathscr{V}\}$. Окрестность $\mathscr{V}$ можно выбрать настолько малой, чтобы она обладала упомянутым свойством для любой интегральной кривой векторного поля $\bar{\mathbf{E}}^v$, которая пересекает $\overline{\mathscr{V}}$, т. е. так, чтобы любая такая кривая выходила из $\pi^{-1}(\mathscr{N})$ в обоих направлениях. Пусть $\mathscr{U}$ — трубка, состоящая из всех точек на интегральных кривых поля $\bar{\mathbf{E}}^v$, пересекающих $\overline{\mathscr{V}}$. Тогда множество $\mathscr{U} \cap \pi^{-1}(\mathscr{N})$ компактно. При достаточно больших n'' кривая $\hat{\lambda}_{n''}(t_{n''})$ будет содержаться в $\mathscr{V}$. Согласно (8.2), поперечные к направлению $\bar{\mathbf{E}}^v$ компоненты касательного к $\hat{\lambda}_{n''}$ вектора настолько малы, что при достаточно больших n'' и $t > t_{n''}$ кривая $\hat{\lambda}_{n''}(t)$ нигде не сможет выйти за пределы трубки $\mathscr{U} \cap \pi^{-1}(\mathscr{N})$, за исключением ее концов, т. е. там, где $\mathscr{U}$ выходит из $\pi^{-1}(\mathscr{N})$. Однако $\hat{\lambda}_{n''}(t)$ не может покинуть $\pi^{-1}(\mathscr{N})$, поскольку $\lambda(t)$ не выходит из $\pi^{-1}(\mathscr{N})$. Следовательно, $\hat{\lambda}_{n''}$ должна лежать в $\mathscr{U} \cap \pi^{-1}(\mathscr{N})$ при $t \geqslant t_{n''}$. Это ведет к следующему противоречию. Кривая $\hat{\lambda}_{n''+1}(t_{n''+1})$ содержится в $\mathscr{V}$; однако, согласно (8.1), окрестность можно выбрать настолько малой, что

$$\hat{\lambda}_{n''}(t_{n''+1}) = \Delta_{n''+1}\Delta_{n''}{}^{-1}\hat{\lambda}_{n''+1}(t_{n''+1})$$

будет лежать в $\mathscr{V}'$, но не в $\mathscr{V}$. Отсюда следует, что наше предположение о том, что изотропная геодезическая $\pi(\mu)$ выходит из $\mathscr{N}$ в обоих направлениях, ложно. Следовательно, должна быть какая-то точка $p \in \mathscr{N}$, которая является предельной точкой $\pi(\mu)$. Тогда по предложению 6.3.1 существует непродолжимая изотропная геодезическая γ, проходящая через p, которая содержится в $\mathscr{N}$ и является предельной кривой для $\pi(\mu)$. □

Если расслоение $O(\mathscr{M})$ m-неполно, можно построить его пополнение по метрике $\overline{O(\mathscr{M})}$. Оно по определению есть множество классов эквивалентности фундаментальных последовательностей из точек расслоения $O(\mathscr{M})$. Если $x \equiv \{x_m\}$ и $y \equiv \{y_m\}$ — две фундаментальные последовательности в $O(\mathscr{M})$, расстоянием $\bar{\rho}(x, y)$ между x и y является по определению $\lim_{n\to\infty} \rho(x_n, y_n)$, где ρ — функция расстояния на $O(\mathscr{M})$, определяемая положительно определенной метрикой $\mathbf{e}$; будем говорить, что x и y эквивалентны, если $\bar{\rho}(x, y) = 0$. $\overline{O(\mathscr{M})}$ можно разложить на часть, гомеоморфную $O(\mathscr{M})$, и множество граничных точек $\bar{\partial}$ (т. е. $\overline{O(\mathscr{M})} = O(\mathscr{M}) \cup \bar{\partial}$). Функция расстояния ρ задает топологию

на $\overline{O(\mathscr{M})}$. Из (8.1) следует, что эта топология на O не зависит от выбора базиса $\{a_i\}$ в T_1.

Действие группы $O(3, 1)$ можно распространить на $\overline{O(\mathscr{M})}$. В самом деле, преобразование базиса $\{\mathbf{G}_A\}$ под действием $A \in O(3, 1)$ не зависит от положения в $O(\mathscr{M})$. Значит, существуют такие положительные постоянные C_1 и C_2 (зависящие только от A), что

$$C_1\rho(u, v) \leqslant \rho(A(u), A(v)) \leqslant C_2\rho(u, v).$$

Это означает, что под действием A последовательности Коши будут отображаться на фундаментальные последовательности и классы эквивалентности фундаментальных последовательностей будут отображаться на такие же классы эквивалентности. Поэтому действие $O(3, 1)$ распространяется на $\overline{O(\mathscr{M})}$ единственным образом. Тогда $\mathscr{M}^+$ можно определить как фактор-пространство $O(\mathscr{M})$ по группе $O(3, 1)$. Поскольку $O(\mathscr{M})/O(3, 1) = \mathscr{M}$ и поскольку под действием $O(3, 1)$ неполная фундаментальная последовательность остается неполной, мы можем представить $\mathscr{M}^+$ как объединение многообразия $\mathscr{M}$ и множества ∂, называемого b-границей $\mathscr{M}$. Точки ∂ можно считать изображениями конечных точек классов эквивалентности b-неполных кривых в $\mathscr{M}$.

Проекция $\pi : \overline{O(\mathscr{M})} \to \mathscr{M}^+$, которая относит точку в $\overline{O(\mathscr{M})}$ к ее классу эквивалентности относительно действия группы $O(3, 1)$, индуцирует топологию на $\mathscr{M}^+$ по топологии на $\overline{O(\mathscr{M})}$. Однако $\bar{\pi}$ не индуцирует на $\mathscr{M}^+$ функцию расстояния, ибо функция $\bar{\rho}$ не инвариантна относительно $O(3, 1)$. Таким образом, хотя топология пространства $\overline{O(\mathscr{M})}$ — метрическая и соответственно хаусдорфова, топология на $\mathscr{M}^+$ не обязательно хаусдорфова. Это означает, что могут найтись такая точка $p \in \mathscr{M}$ и такая точка $q \in \partial$, что любая окрестность p в $\mathscr{M}^+$ будет пересекать любую окрестность q. Это случится, когда точка q будет соответствовать неполной кривой, которая полностью или частично захвачена в $\mathscr{M}$. Мы обсудим неполноту, связанную с захватом, позднее, в разд. 8.5.

Если $\mathbf{g}$ — положительно определенная метрика на $\mathscr{M}$, то $\mathscr{M}^+$ гомеоморфно пополнению $(\mathscr{M}, \mathbf{g})$ фундаментальными последовательностями. Конструкция Шмидта обладает, в частности, тем желательным свойством, что при вырезании из пространства замкнутого множества $\mathscr{A}$ для каждой точки из $\mathscr{A}$ возникает по меньшей мере одна точка в b-границе, которая окажется конечной точкой для некоторой кривой в $\mathscr{M} - \mathscr{A}$. Примером появления более одной точки b-границы для одной точки из $\mathscr{A}$ служит двумерное пространство Минковского, в котором в качестве $\mathscr{A}$ взят отрезок оси t между -1 и $+1$. Тогда при $-1 < t < 1$

для каждой точки $(0, t)$ будут две точки b-границы. Примером, когда точка на $\dot{\mathscr{A}}$ не может быть достигнута какой-либо кривой из $\mathscr{M} - \mathscr{A}$, служит множество

$$\mathscr{A} = \left\{t = \sin\frac{1}{x},\ t \neq 0\right\} \cup \{-1 \leqslant t \leqslant 1,\ x = 0\}.$$

В $\mathscr{M} - \mathscr{A}$ нет ни одной кривой, которая имела бы конечную точку в начале координат, и, следовательно, эта точка не может лежать в $(\mathscr{M} - \mathscr{A})^+$, хотя она и принадлежит $\dot{\mathscr{A}}$.

Несмотря на изящную формулировку, метод Шмидта, к сожалению, крайне трудно применить на практике. Единственным решением, для которого можно найти $\mathscr{M}^+$, помимо пространств постоянной кривизны является двумерное решение Робертсона — Уокера с обычной материей. В нем доказывается, что ∂ — это пространственноподобная 1-поверхность, как и можно было бы ожидать из конформной картины. В этом случае можно наделить ∂ естественной дифференцируемой структурой и превратить $\mathscr{M}^+$ в многообразие с краем. Однако, видимо, не существует какого-либо общего способа введения структуры многообразия на ∂. Конечно, естественно думать, что в типовых ситуациях ∂ будет в высшей степени иррегулярным и на нем нельзя будет ввести гладкую структуру.

8.4. Характер сингулярностей

В этом и следующем разделах мы исследуем характер сингулярностей, предсказываемых теоремой 4. Мы рассматриваем именно эту теорему, а не другие, потому что из нее можно получить больше информации о сингулярности. Однако сингулярности, предсказываемые остальными теоремами, обладают сходными свойствами.

Сразу встает вопрос: сколь сильное нарушение дифференцируемости метрики допустимо? Теоремы предыдущего раздела утверждают, что пространство-время должно быть геодезически неполным, если метрика принадлежит классу C^2. Это условие необходимо для того, чтобы были хорошо определены сопряженные точки и вариации длин дуг; иными словами, чтобы решения уравнения геодезических зависели от начальных положений и направлений дифференцируемым образом. Однако о геодезической неполноте можно говорить при условии, что решения уравнения геодезических существуют. Они существуют, когда метрика принадлежит классу C^1; единственность решений и их непрерывная зависимость от начальных положений и направлений имеет место, если метрика класса C^{2-} (т. е. если связность локально липшицева). Фактически можно обсуждать b-неполноту, опираясь лишь на существование положительно-опреде-

ленной метрики $\mathbf{e}$ в расслоении реперов $O(\mathscr{M})$, которая почти всюду определена и локально ограничена. Это имеет место, если компоненты связности $\Gamma^a{}_{bc}$ определены почти всюду и локально ограничены, т. е. если метрика класса C^{1-}.

Итак, может показаться, что содержание теорем о сингулярностях не связано с неограниченным ростом кривизны, а свидетельствует лишь о наличии разрыва кривизны (т. е. метрика становится класса C^{2-} вместо C^2). Мы покажем, что это не так: при выполнении условий теоремы 4 пространство-время должно быть времениподобно геодезически неполным (и, следовательно, b-неполным), даже если потребовать, чтобы метрика была класса C^{2-}. Метод доказательства состоит в том, чтобы аппроксимировать C^{2-}-метрику C^2-метрикой и произвести вариацию длины дуги в этой метрике.

Предположим, что данное пространство-время нерасширяемо при метрике класса C^{2-} и что выполнены условия теоремы 4. Теперь требуется, чтобы условие времениподобного схождения $R_{ab}K^aK^b \geqslant 0$ выполнялось «почти всюду», причем тензор Риччи R_{ab} определяется через обобщенные производные. Единственный пункт в доказательстве теоремы 4, который теряет силу в C^{2-}-метрике, — это тот, где используется вариация длины дуги для того, чтобы показать, что не может существовать точка $p \in \in D^+(\hat{\mathscr{S}})$, для которой $d(\hat{\mathscr{S}}, p) > -3/\theta_0$, причем θ_0 — максимальное значение $\chi^a{}_a$ на $\mathscr{S}$. Итак, если бы многообразие $\mathscr{M}$ было геодезически полным, то нашлись бы такая точка p и ортогональная к $\hat{\mathscr{S}}$ геодезическая от $\hat{\mathscr{S}}$ до p длиной $d(\hat{\mathscr{S}}, p)$. Пусть $\mathscr{U}$ — открытое множество с компактным замыканием, которое содержит $J^-(p) \cap J^+(\hat{\mathscr{S}})$, и пусть $\mathbf{e}$ и $\hat{\mathbf{g}}$ — соответственно положительно определенная и лоренцева C^∞-метрика. Для $\varepsilon > 0$ можно найти лоренцеву C^∞-метрику g_ε^{ab}, удовлетворяющую на $\overline{\mathscr{U}}$ условиям

1) $|g_\varepsilon{}^{ab} - g^{ab}| < \varepsilon$,

2) $|g_\varepsilon{}^{ab}{}_{|c} - g^{ab}{}_{|c}| < \varepsilon$,

3) $|g_\varepsilon{}^{ab}{}_{|cd}| < C$, где C — постоянная, зависящая от $\mathscr{U}$, $\mathbf{e}$, $\hat{\mathbf{g}}$, $\mathbf{g}$,

4) $R_{\varepsilon ab}K^aK^b > -\varepsilon|K^a|^2$ для любого вектора $\mathbf{K}$, для которого $g_{\varepsilon ab}K^aK^b \geqslant 0$. (Метрику $g_\varepsilon{}^{ab}$ можно построить, покрывая $\overline{\mathscr{U}}$ конечным числом локальных координатных окрестностей $(\mathscr{V}_\alpha, \phi_\alpha)$, интегрируя координатные компоненты g^{ab} с подходящей сглаживающей функцией $\rho_\varepsilon(x)$ и суммируя интегралы с разложением единицы $\{\psi_\alpha\}$, т. е.

$$g_\varepsilon{}^{ab}(q) = \sum_\alpha \psi_\alpha(q) \int_{\phi_\alpha(\mathscr{V}_\alpha)} g^{ab}(x)\,\rho_\varepsilon(x - \phi_\alpha(q))\,\mathrm{d}^4x,$$

где $\int \rho_\varepsilon(x)\,d^4x = 1$.)

Свойство (1) означает, что при достаточно малых значениях ε точка ρ будет лежать в $D^+(\hat{\mathscr{S}}, \mathbf{g}_\varepsilon)$ и $I^-(p, \mathbf{g}_\varepsilon) \cap I^+(\hat{\mathscr{S}}, \mathbf{g}_\varepsilon)$ будет содержаться в $\mathscr{U}$. Поэтому в метрике $\mathbf{g}_\varepsilon$ будет существовать геодезическая γ_ε от $\hat{\mathscr{S}}$ до p длиной $d_\varepsilon(\hat{\mathscr{S}}, p)$. Кроме того, $|d_\varepsilon(\hat{\mathscr{S}}, p) - d(\hat{\mathscr{S}}, p)|$ будет стремиться к нулю при $\varepsilon \to 0$.

В силу свойств (1), (2), (3) и стандартных теорем для обыкновенных дифференциальных уравнений при $\varepsilon \to 0$ касательные векторы к геодезическим в метрике $\mathbf{g}_\varepsilon$ будут стремиться к касательным к геодезическим в метрике $\mathbf{g}$ с теми же начальными точками и направлениями. При этом, если V^a-единичный касательный вектор к геодезической в метрике $\mathbf{g}$, ортогональной к $\hat{\mathscr{S}}$, то на

$$\bar{\mathscr{U}} \cap \beta(\hat{\mathscr{S}} \times [0,\ 2d(\hat{\mathscr{S}},\ p)])$$

величины $|V^a|$ будут ограничены сверху. Следовательно, для любого $\delta > 0$ найдется такая постоянная $\varepsilon_1 > 0$, что при $\varepsilon < \varepsilon_1$ будет удовлетворяться неравенство $R_{\varepsilon ab} V_\varepsilon{}^a V_\varepsilon{}^b > -\delta$. Теперь мы можем установить противоречие, показав, что достаточно малое изменение энергетического условия не будет препятствовать появлению сопряженных точек в метрике $\mathbf{g}_\varepsilon$ на расстоянии, меньшем $d_\varepsilon(\hat{\mathscr{S}}, p)$. В самом деле, расхождение θ_ε геодезических в метрике $\mathbf{g}_\varepsilon$ подчиняется уравнению Райчаудхури:

$$\frac{d\theta_\varepsilon}{ds} = -\frac{1}{3}\theta_\varepsilon{}^2 - 2\sigma_\varepsilon{}^2 - R_{\varepsilon ab} V_\varepsilon{}^a V_\varepsilon{}^b.$$

Отсюда

$$\frac{d(\theta_\varepsilon{}^{-1})}{ds} \geqslant \frac{1}{3} + R_{\varepsilon ab} V^a V^b \theta_\varepsilon{}^{-2}.$$

Поэтому, если бы начальное значение θ_{ε_0} было отрицательным, а $3\delta\theta_{\varepsilon_0}{}^{-2}$ было меньше единицы, то величина $\theta_\varepsilon{}^{-1}$ стала бы отрицательной на расстоянии от $\mathscr{S}$, не превышающем $3/\theta_0(1-3\delta\theta_0{}^{-2})$. Но $\theta_{\varepsilon_0} \to \theta_0$ при $\varepsilon \to 0$. Отсюда видно, что при достаточно малых ε в метрике $\mathbf{g}_\varepsilon$ на любой геодезической, ортогональной $\hat{\mathscr{S}}$, будет сопряженная точка на расстоянии от $\hat{\mathscr{S}}$, меньшем чем $d_\varepsilon(\hat{\mathscr{S}}, p)$. Поэтому многообразие $\mathscr{M}$ должно быть времениподобно геодезически полным, даже если мы требуем от метрики лишь принадлежности классу C^{2-}.

Из этого результата вытекает, что при попытке расширить пространство-время с целью сделать неполные геодезические полными или метрика перестанет быть лоренцевой, или кривизна должна быть локально неограниченной, т. е. должна быть сингулярность кривизны. Однако, даже если кривизна локально не ограничена, все же метрику можно интерпретировать как решение уравнений Эйнштейна в обобщенных функциях при условии, что объемные интегралы компонент тензора кривизны

по любой компактной области будут конечны. Последнее будет справедливо, если метрика лоренцева, непрерывна и обладает квадратично интегрируемыми первыми производными. В частности, это будет выполнено, если метрика лоренцева и класса C^{1-} (т. е. локально липшицева). Примерами таких C^{1-}-решений служат гравитационные ударные волны (кривизна имеет поведение типа δ-функции на изотропных 3-поверхностях; см., например, [29, 130]), массивные тонкие оболочки (где кривизна ведет себя как δ-функция на времениподобных 3-поверхностях; см., например, [83]) и решения, содержащие вещество без давления («пыль»), в которых геодезические линии тока обладают двумерными и трехмерными каустиками (см. [65, 120]).

Ввиду нелинейной зависимости кривизны от метрики не всегда возможно аппроксимировать C^{-1}-решение в обобщенных функциях C^2-метрикой, подчиняющейся условию схождения или хотя бы нарушающей это условие на не слишком большую величину, как в предыдущем случае (свойство 4). Однако во всех упомянутых выше примерах это возможно. Для этого в сущности есть физическое оправдание: их надо рассматривать как математические идеализации C^2- или C^∞-решений, которые подчиняются условию сближения и в которых кривизна велика в малых областях. К этим C^2-решениям можно было бы применить теоремы разд. 8.2 и доказать существование в них неполных геодезических. Из этого ясно, что предсказываемые сингулярности не могут быть лишь гравитационными ударными волнами или каустиками линий тока, а должны представлять собой более серьезные нарушения метрики. (С обычными гидродинамическими ударными волнами связаны лишь разрывы плотности и давления, и в этом случае метрика может быть класса C^{2-}.) Хотя мы и не можем доказать это строго, мы уверены, что сингулярности должны оказаться такими, чтобы метрику нельзя было расширить даже как решение уравнений Эйнштейна в обобщенных функциях, т. е. наряду с неограниченностью в сингулярной точке компонент тензора кривизны их объемные интегралы в любой окрестности такой точки должны быть не ограничены. Так обстоит дело во всех известных примерах сингулярностей, за исключением особого случая решения Тауба — НУТ, которым мы займемся в следующем разделе. Если эта догадка верна для «характерных» особенностей (т. е. для всех особенностей, кроме тех, начальные условия для которых образуют множество меры нуль), то мы можем рассматривать сингулярность как точку, в которой уравнения Эйнштейна (и, наверное, остальные известные законы физики) теряют силу.

Другой вопрос, на который хотелось бы иметь ответ: каково число неполных геодезических, если есть сингулярность? Если только одна, то может создаться впечатление, что эту сингуляр-

ность можно игнорировать. Из доказательства теоремы 4 можно видеть, что при отсутствии горизонта Коши, т. е. если $\hat{\mathscr{S}}$ — поверхность Коши, ни одну времениподобную кривую с поверхности $\hat{\mathscr{S}}$ (неважно, геодезическая она или нет) нельзя продолжить до длины, превышающей $-3/\theta_0$, где θ_0 — максимальное

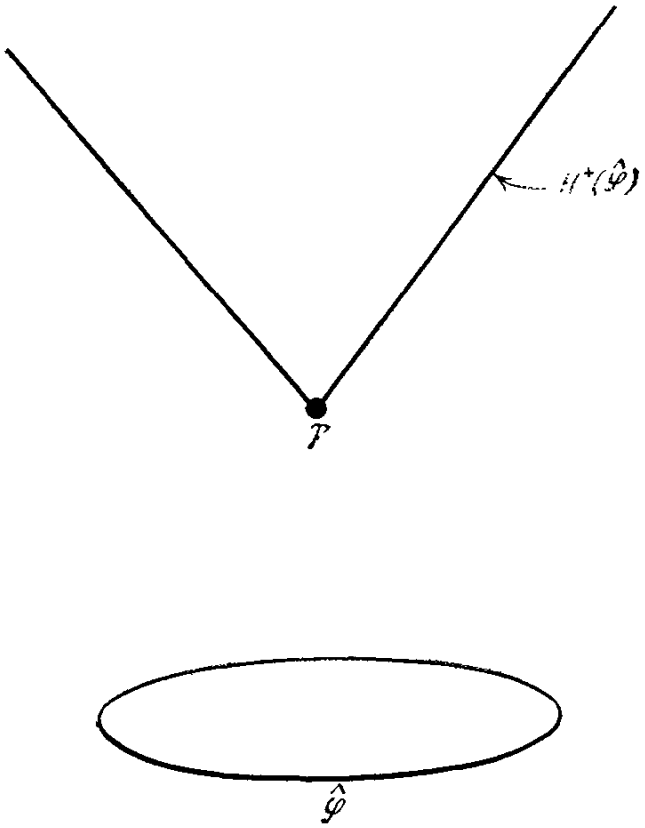


Рис. 53. Точка p удалена из пространства-времени, поскольку на ней находится сингулярность. Соответственно для поверхности $\hat{\mathscr{S}}$ существует горизонт Коши $H^+(\hat{\mathscr{S}})$.

значение метрики $\chi^a{}_a$ на $\mathscr{S}$. На самом деле этот результат справедлив даже тогда, когда поверхность $\mathscr{S}$ некомпактна, если $\chi^a{}_a$ при этом имеет отрицательную нижнюю грань. Однако это не обязательно свидетельствует о том, что любая времениподобная кривая должна встретиться с сингулярностью. Скорее, сингулярность будет сочетаться с горизонтом Коши, и потому мы лишимся возможности предсказывать будущее. Пример этого дан на рис. 53. Здесь метрика сингулярна в точке p и соответственно эта точка вырезана из пространственно-временного многообразия. Существует горизонт Коши, расходящийся от этой дыры. Этот пример показывает, что самое большее, на что можно надеяться, — это доказать существование трехмерного семейства геодезических, которые неполны и остаются в пределах области Коши поверхности $\hat{\mathscr{S}}$ (в нашем примере это — геодезические, которые проходят через p). Могут быть иные геодезиче-

ские, которые уходят из области Коши $\hat{\mathscr{S}}$ и которые неполны, но мы не можем предсказать их поведение по условиям, заданным на $\hat{\mathscr{S}}$.

Ясно, что в $D^+(\hat{\mathscr{S}})$ должно быть более одной неполной геодезической. Действительно, из теоремы 4 следует, что должна быть геодезическая γ, ортогональная к $\hat{\mathscr{S}}$, которая остается в $D^+(\hat{\mathscr{S}})$, но неполна. Пусть p — точка, в которой γ пересекает $\hat{\mathscr{S}}$. Тогда можно слегка изменить поверхность $\hat{\mathscr{S}}$ в окрестности точки p так, чтобы получилась новая поверхность $\hat{\mathscr{S}}$, для которой по-прежнему $\chi^a{}_a < 0$, но которая не ортогональна к γ. Тогда по теореме 4 должна найтись другая, ортогональная к $\hat{\mathscr{S}}'$ времениподобная геодезическая γ', которая неполна и не пересекает горизонт $H^+(\hat{\mathscr{S}}')$, совпадающий с $H^+(\hat{\mathscr{S}})$.

В действительности можно доказать, что имеется по меньшей мере трехмерное семейство времениподобных геодезических (по одной через каждую точку некоторой ахрональной поверхности), которые остаются в $D^+(\hat{\mathscr{S}})$ и неполны. Все эти геодезические соответствуют одной и той же точке границы в смысле неразложимых прошлых множеств (разд. 6.8), т. е. все они имеют одно и то же прошлое. Однако в силу построения предыдущего раздела они не могут соответствовать одним и тем же точкам. Схема доказательства такова. В теореме 4 было показано, что должна существовать направленная в будущее ортогональная к $\mathscr{S}$ времениподобная геодезическая, которую нельзя продолжить до длины $3/\theta_0$. Можно высказать более сильное утверждение: должна существовать такая геодезическая γ, которая остается в пределах $D^+(\hat{\mathscr{S}})$ и в каждой своей точке является кривой максимальной длины, т. е. для каждой $q \in \gamma$ длина γ от $\hat{\mathscr{S}}$ до q равна $d(\mathscr{S}, q)$. Идея теперь состоит в том, чтобы рассмотреть функцию $d(r, \gamma)$ при $r \in J^-(\gamma)$. Ясно, что она ограничена на $J^+(\hat{\mathscr{S}}) \cap J^-(\gamma)$. Из того факта, что γ — кривая максимальной длины от $\hat{\mathscr{S}}$, следует, что в окрестности γ функция $d(r, \gamma)$ непрерывна, и поверхности $d(r, \gamma) = \text{const}$ являются пространственноподобными поверхностями, пересекающими γ ортогонально. Тогда времениподобные геодезические, ортогональные к этим поверхностям, остаются внутри $J^-(\gamma)$ и, следовательно, неполны.

8.5. Захваченная неполнота

В разд. 8.1 мы предложили использовать b-неполноту в качестве определения сингулярности. Наш замысел состоял в том, чтобы поставить b-неполную кривую в соответствие сингулярной точке, которая удалена из пространства-времени. Допустим,

однако, что имеется b-неполная кривая λ, у которой существует предельная точка $p \in \mathscr{M}$, т. е. λ частично или полностью захвачена в компактной окрестности p. Тогда мы не можем вложить $\mathscr{M}$ в большее четырехмерное хаусдорфово паракомпактное многообразие $\mathscr{M}'$ так, чтобы λ продолжилась в $\mathscr{M}'$. Действительно, если q — точка, в которой λ пересекает границу $\mathscr{M}$ в $\mathscr{M}'$, то любая окрестность q должна пересекать любую окрестность p, что невозможно, поскольку $\mathscr{M}'$ хаусдорфово и $q \neq p$. По существу, захваченную неполноту в $\mathscr{M}^+$ можно охарактеризовать нехаусдорфовым поведением пополнения $\mathscr{M}^+$ из конструкции Шмидта.

Предложение 8.5.1

Точка $p \in \mathscr{M}$ не отделена по Хаусдорфу в $\mathscr{M}^+$ от точки $r \in \partial$, если в $\mathscr{M}$ существует неполная кривая λ, для которой p служит предельной точкой, а r — конечной точкой в $\mathscr{M}^+$.

Допустим, что $p \in \mathscr{M}$ — предельная точка b-неполной кривой λ. Можно построить горизонтальный лифт $\bar{\lambda}$ кривой λ в расслоении ортонормированных реперов $O(\mathscr{M})$. Некоторая точка

$$x \in \pi^{-1}(r) \subset \bar{\partial} \equiv \overline{O(\mathscr{M})} - O(\mathscr{M})$$

будет служить для $\bar{\lambda}$ конечной точкой. Если $\mathscr{V}$ — некоторая окрестность точки r в $\mathscr{M}^+$, то $\pi^{-1}(\mathscr{V})$ является открытой окрестностью x в $\overline{O(\mathscr{M})}$. Таким образом, $\pi^{-1}(\mathscr{V})$ содержит все точки на $\bar{\lambda}$ после некоторой точки y. Поэтому все точки на λ после $\pi(y)$ будут лежать в $\mathscr{V}$ и $\mathscr{V}$ будет пересекать любую окрестность точки p, так как p — предельная точка кривой λ. □

Пространство Тауба — НУТ (разд. 5.8) служит примером, в котором существуют неполные геодезические, причем все они полностью захвачены в компактных окрестностях горизонтов прошлого и будущего $U(t) = 0$. Поскольку в этих компактных окрестностях метрика вполне регулярна, неполные геодезические не соответствуют сингулярностям скалярных полиномов кривизны. Рассмотрим неполную в будущем замкнутую изотропную геодезическую $\lambda(v)$ на горизонте будущего $V(t) = 0$. Пусть $p = \lambda(0)$, и пусть v_1 — первое положительное значение v, для которого $\lambda(v) = p$. Тогда, как и в разд. 6.4, параллельно перенесенный касательный к λ вектор будет удовлетворять равенству

$$\left(\frac{\partial}{\partial v}\right)\Big|_{v=v_1} = a\left(\frac{\partial}{\partial v}\right)\Big|_{v=0},$$

где $a > 1$. $\lambda(v_n) = p$ для каждого n, где

$$v_n = v_1 \sum_{r=1}^{n} a^{1-r} = v_1 \frac{1 - a^{-n}}{1 - a^{-1}}$$

и

$$\left(\frac{\partial}{\partial v}\right)\Big|_{v=v_n} = a^n \left(\frac{\partial}{\partial v}\right)\Big|_{v=0}.$$

Таким образом, если на $\lambda(v)$ взять параллельно перенесенный псевдоортонормированный базис $\{\mathbf{E}_a\}$, где $\mathbf{E}_4 = \partial/\partial v$, то для другого изотропного базисного вектора $\mathbf{E}_3$ выполняется равенство $\mathbf{E}_3|_{v=v_n} = a^{-n}\mathbf{E}_3|_{v=0}$. При каждом обороте вдоль замкнутой изотропной геодезической λ вектор $\mathbf{E}_4$ становится больше, а вектор $\mathbf{E}_3$ — меньше. Векторы $\mathbf{E}_1$ и $\mathbf{E}_2$ не меняются. Поэтому если тензор Римана имеет ненулевую компоненту, которая включает $\mathbf{E}_4$ и, возможно, $\mathbf{E}_1$ и $\mathbf{E}_2$, то она будет казаться все больше и больше при каждом обороте по λ и, следовательно, будет иметь место сингулярность кривизны в параллельно перенесенном базисе. Однако в пространстве Тауба — НУТ вектор $\mathbf{E}_3$ можно выбрать так, что у тензора Римана будет только одна независимая компонента, а именно $R(\mathbf{E}_3, \mathbf{E}_4, \mathbf{E}_3, \mathbf{E}_4)$. Она включает $\mathbf{E}_3$ и $\mathbf{E}_4$ одинаковым образом и поэтому не меняет своего значения с оборотами по λ. Поскольку такой же вывод, по всей вероятности, справедлив для любой захваченной кривой, похоже, что в пространстве Тауба — НУТ нет никакой сингулярности кривизны, хотя оно, согласно нашему определению, сингулярно. Хотелось бы знать, может ли случиться нечто подобное в физически реалистических решениях, содержащих вещество, или же пространство Тауба — НУТ является особым патологическим случаем. Этот вопрос важен потому, что в следующей главе мы будем обосновывать такую интерпретацию предыдущих теорем, согласно которой они свидетельствуют не о том, что непременно должна быть сингулярность, а о том, что в сильных гравитационных полях общая теория относительности теряет силу. Таких полей нет в ситуации, подобной пространству Тауба — НУТ. Этот вывод основан на весьма специальном виде тензора Римана в пространстве Тауба — НУТ. Следует ожидать, что в общем случае на захваченной кривой будут отличны от нуля некоторые другие компоненты тензора Римана и соответственно появятся сингулярности кривизны в параллельно перенесенном базисе, хотя сингулярности скалярных полиномов могут и отсутствовать. Действительно, можно доказать

Предложение 8.5.2

Если $p \in \mathscr{M}$ — предельная точка b-неполной кривой λ и если в p для всех непространственноподобных векторов $\mathbf{K}$

$$R_{ab}K^aK^b \neq 0,$$

то кривой λ соответствуют сингулярности кривизны в параллельно перенесенном базисе. (Условие $R_{ab}K^aK^b \neq 0$ может быть

заменено условием отсутствия каких-либо изотропных направлений K^a, для которых $K^aK^cC_{abc[d}K_{e]} = 0$.)

Пусть $\mathcal{U}$ — выпуклая нормальная координатная окрестность точки p с компактным замыканием, и пусть $\{\mathbf{Y}_i\}$, $\{\mathbf{Y}^i\}$ — поле дуальных ортонормированных базисов в $\mathcal{U}$. Пусть $\{\mathbf{E}_a\}$, $\{\mathbf{E}^a\}$ — параллельно перенесенные дуальные ортонормированные базисы на кривой $\lambda(t)$, а $\tilde{t}$ — параметр на λ, такой, что в $\mathcal{U}$

$$\frac{\mathrm{d}\tilde{t}}{\mathrm{d}t} = \left(\sum_i X^i X^i\right)^{1/2},$$

где X^i — компоненты касательного вектора $\partial/\partial t$ в базисе $\{\mathbf{Y}_i\}$. Тогда параметром $\tilde{t}$ измеряется длина дуги в положительно определенной метрике в $\mathcal{U}$, в которой базисы $\{\mathbf{Y}_i\}$, $\{\mathbf{Y}^i\}$ ортонормированы.

Поскольку в p $R_{ab}K^aK^b \neq 0$ для любого непространственноподобного вектора K^a, существует окрестность $\mathcal{V} \subset \mathcal{U}$, в которой

$$R_{ab} = CZ_aZ_b + R'_{ab},$$

где $C \neq 0$ — постоянная, Z_a — единичный времениподобный вектор и R'_{ab} удовлетворяет неравенству $CR'_{ab}K^aK^b > 0$ при любом непространственноподобном векторе K^a. Допустим, что после некоторого значения $\tilde{t}_0$ параметра $\tilde{t}$ кривая λ пересечет $\mathcal{V}$. Поскольку у λ нет конечных точек, а p — предельная точка λ, та часть λ, которая лежит в $\mathcal{V}$, будет иметь бесконечную длину, если она измерена по параметру $\tilde{t}$. Однако обобщенный аффинный параметр дается формулой

$$\frac{\mathrm{d}u}{\mathrm{d}\tilde{t}} = \left\{\sum_a (\mathbf{E}^a{}_i\tilde{X}^i)^2\right\}^{1/2},$$

где $\tilde{X}^i$ — компоненты касательного вектора $(\partial/\partial\tilde{t})_\lambda$, так что $\sum_i \tilde{X}^i\tilde{X}^i = 1$ и $\mathbf{E}^a{}_i$ — компоненты базиса $\{\mathbf{E}^a\}$ в базисе $\{\mathbf{Y}^i\}$. Поскольку параметр u конечен на кривой λ, модуль вектора-столбца $E^a{}_i\tilde{X}^i$ должен стремиться к нулю и соответственно преобразования Лоренца, определяемые компонентами $\mathbf{E}^a{}_i$, должны стать неограниченно большими. $\mathbf{Z}$ — единичный времениподобный вектор, и поэтому компоненты $\mathbf{Z}$ в базисе $\{\mathbf{E}_a\}$ будут неограниченно расти, а следовательно, и некоторые компоненты тензора Риччи в базисе $\{\mathbf{E}_a\}$ станут сколь угодно велики. □

Из этого результата видно, что наблюдатель, чья история представляет собой b-неполную, захваченную непространственноподобную кривую в характерном пространстве, будет сброшен с этой кривой за конечный промежуток времени, тогда как другой наблюдатель может путешествовать в том же районе, не

испытывая ничего подобного. Интересным примером в этом отношении служит такое пространство Тауба — НУТ, у которого метрика умножена на конформный множитель Ω, отличающийся от единицы лишь в малой окрестности точки p на горизонте. Это конформное преобразование не меняет причинной структуры пространства и не влияет на неполноту замкнутой изотропной геодезической, проходящей через точку p. Однако, вообще говоря, $R_{ab}K^aK^b \neq 0$, где K^a — касательный вектор к этой замкнутой изотропной геодезической. После каждого цикла $R_{ab}K^aK^b$ умножается на a^2, и поэтому имеет место сингулярность кривизны в параллельно перенесенном базисе. Все же метрика вполне регулярна в компактной окрестности горизонта, и потому с этой неполнотой не связана никакая сингулярность скалярных полиномов кривизны.

Было бы желательно исключить ситуацию такого рода, когда неполные кривые полностью захвачены в компактной области. Однако такое могло бы произойти в бесконечном счетном множестве различных областей пространства-времени. Поэтому при рассмотрении ее нельзя ограничиться случаем, когда *все* неполные кривые полностью заключены в каком-то одном компактном множестве. Вместо этого правильно будет сказать, что мы хотим исключить ситуацию, когда множество неполных кривых, которое в каком-то смысле компактно, полностью захвачено в компактной области многообразия $\mathscr{M}$. Чтобы выразить это в более строгой форме, мы введем следующее определение b-ограниченности.

Определим пространство $B(\mathscr{M})$ как множество всех пар (λ, u), где u — точка в расслоении линейных реперов $L(\mathscr{M})$ и λ — C^1-кривая в $\mathscr{M}$, имеющая только одну конечную точку, лежащую в $\pi(u)$. Пусть $\mathscr{U}$ — открытое множество в $\mathscr{M}$ и $\mathscr{V}$ — открытое множество в $L(\mathscr{M})$. Мы определим открытое множество $O(\mathscr{U}, \mathscr{V})$ как множество всех элементов из $B(\mathscr{M})$, таких, что λ пересекают $\mathscr{U}$ и $u \in \mathscr{V}$. Множества вида $O(\mathscr{U}, \mathscr{V})$ при всех $\mathscr{U}$, $\mathscr{V}$ образуют подбазис для топологии $B(\mathscr{M})$. Напомним, что отображение $\exp : T(\mathscr{M}) \to \mathscr{M}$ определяется заданием вектора $\mathbf{X}$ в точке p и смещением из p вдоль геодезической в направлении $\mathbf{X}$ на единицу расстояния, измеренного по аффинному параметру, который определяется вектором $\mathbf{X}$. Аналогично мы можем задать отображение $\mathrm{Exp} : B(\mathscr{M}) \to \mathscr{M}$, следуя из $\pi(u)$ вдоль кривой λ на единицу расстояния, измеренного по обобщенному аффинному параметру, который определен точкой $u \in L(\mathscr{M})$. Отображение Exp определено непрерывным образом для любого $B(\mathscr{M})$, когда $\mathscr{M}$ — полное многообразие. Будем говорить, что пространство $(\mathscr{M}, \mathbf{g})$ b-ограничено, если для любого компактного множества $W \subset B(\mathscr{M})$ $\mathrm{Exp}(W)$ обладает компактным замыканием в $\mathscr{M}$. Ввиду непрерывности Exp пространство $(\mathscr{M}, \mathbf{g})$

b-ограничено, если оно b-полно. Однако пространство Тауба — НУТ, например, b-ограничено, но b-неполно. Мы покажем, что это оказалось возможным только из-за того, что пространство Тауба — НУТ везде пустое. Присутствие любой материи на поверхности $\mathscr{S}$ (теорема 4) будет означать, что пространство и b-неполно, и b-неограниченно.

Теорема 5

Пространство-время b-неполно, если выполняются условия (1) — (3) теоремы 4 и, кроме того,

4) тензор энергии-импульса не равен нулю где-либо на $\mathscr{S}$,

5) тензор энергии-импульса подчиняется условию энергодоминантности (разд. 4.3) в несколько более сильной форме (разд. 4.3): если K^a — непространственноподобный вектор, то $T^{ab}K_a$ равен нулю или непространственноподобен и $T_{ab}K^aK^b \geqslant 0$, причем равенство имеет место лишь при $T^{ab}K_b = 0$.

Замечание. Условие (4) можно заменить типовым условием (см. теорему 2).

Доказательство. Рассмотрим накрывающее пространство $\mathscr{M}_G$ (разд. 6.5), определенное как множество всех пар $(p, i[\lambda])$, где λ — кривая от q до p, p, $q \in \mathscr{M}$ и $i[\lambda]$ — разность чисел пересечений поверхности $\mathscr{S}$ кривой λ в направлении будущего и в направлении прошлого. Для каждого положительного a множество

$$\mathscr{S}_a \equiv \{(p, i[\lambda]) : p \in \mathscr{S},\ i[\lambda] = a\}$$

диффеоморфно поверхности $\mathscr{S}$ и является частичной поверхностью Коши в $\mathscr{M}_G$. Вообще, если многообразие $\mathscr{M}$ b-ограничено, $\mathscr{M}_G$ еще не обязательно является таковым, но в рассматриваемой ситуации справедливо следующее утверждение:

Лемма 8.5.3

Пусть выполнены условия (1) — (3), и пусть $D^+(\mathscr{S}_0)$ не обладает компактным замыканием в $\mathscr{M}_G$; тогда, если ψ — накрывающая проекция, $\psi : \mathscr{M}_G \to \mathscr{M}$, $\psi(D^+(\mathscr{S}_0))$ не будет иметь компактного замыкания в $\mathscr{M}$.

Многообразие $\mathscr{M}$ диффеоморфно или многообразию $\mathscr{M}_G$, или той части $\mathscr{M}_a$ многообразия $\mathscr{M}_G$, которая лежит между $\mathscr{S}_a$ и $\mathscr{S}_{a+1}$; в последнем случае $\mathscr{S}_a$ и $\mathscr{S}_{a+1}$ отождествляются. Если для любого $a \geqslant 0$ $\mathscr{M}_a \cap D^+(\mathscr{S}_0)$ не имеет компактного замыкания в $\mathscr{M}_G$, то $\psi(D^+\mathscr{S}_0))$ не будет иметь компактного замыкания в $\mathscr{M}$. Если, однако, множество $\mathscr{M}_a \cap D^+(\mathscr{S}_0)$ обладает компактным замыканием для всех $a \geqslant 0$, в силу некомпактности $\bar{D}^+(\mathscr{S}_0)$, оно должно быть также и непустым для всех $a \geqslant 0$. Но для $p \in \mathscr{S}_a$ собственный объем $I^-(p) \cap \mathscr{M}_{a-1}$ ограничен снизу некоторой постоянной c. Следовательно, для любого $a \geqslant 0$ собствен-

ный объем области $\mathscr{M}_a \cap D^+(\mathscr{S}_0)$ не должен быть меньше c. Но это невозможно, поскольку, в силу условий (1)—(3) и предложения 6.7.1, собственный объем $D^+(\mathscr{S}_0)$ меньше, чем $3/(-\theta_0) \times$ $\times$(площадь поверхности $\mathscr{S}$), где θ_0 — отрицательная верхняя грань $\chi^a{}_a$ на $\mathscr{S}$. □

Используя этот результат, можно доказать такое утверждение:

Лемма 8.5.4

Если $D^+(\mathscr{S}_0)$ не имеет компактного замыкания, многообразие $\mathscr{M}$ не является b-ограниченным.

Пусть $\mathscr{W}$ — подмножество пространства $B(\mathscr{M}_G)$, состоящее из всех пар (λ, u), где λ — любая непродолжимая в будущее, ортогональная к $\mathscr{S}_0$ времениподобная геодезическая кривая в $\mathscr{M}_G$ с конечной точкой $r \in \mathscr{S}_0$ и $u \in \pi^{-1}(r)$ — любой базис в r, один из векторов которого касателен к λ и имеет длину $-3/\theta_0$, а остальные векторы образуют ортонормированный базис в $\mathscr{S}_0$.

Пусть $\{\mathscr{P}_\alpha\}$ — набор открытых множеств, покрывающих $\mathscr{W}$. Каждое $\mathscr{P}_\alpha$ будет объединением конечного числа пересечений множеств вида $O(\mathscr{U}, \mathscr{V})$. Достаточно рассмотреть случай, когда $\mathscr{P}$ можно представить в виде

$$\mathscr{P}_\alpha = \bigcap_\beta O(\mathscr{U}_{\alpha\beta}, \mathscr{V}_\alpha),$$

где при каждом α $\mathscr{U}_{\alpha\beta}$ — набор некоторого конечного числа открытых множеств в $\mathscr{M}_G$ и $\mathscr{V}_\alpha$ — открытое множество в $L(\mathscr{M}_G)$. Пусть $(\mu, v) \in \mathscr{W}$. Тогда существует значение α, при котором $(\mu, v) \in \mathscr{P}_\alpha$. Это означает, что геодезическая μ пересекает открытое множество $\mathscr{U}_{\alpha\beta}$ при каждом значении β и что $v \in \mathscr{V}_\alpha$. Поскольку геодезические непрерывно зависят от их начальных значений, найдется окрестность $\mathscr{Y}_\alpha$ точки $\pi(v)$, такая, что каждая непродолжимая в будущее геодезическая, проходящая через $\mathscr{Y}_\alpha$ и ортогональная к $\mathscr{S}_0$, будет пересекать $\mathscr{U}_{\alpha\beta}$ при каждом значении β. Пусть $\mathscr{V}'_\alpha \subset \mathscr{V}_\alpha$ — открытое множество, такое, что $\pi(\mathscr{V}'_\alpha) \subset \mathscr{Y}_\alpha$. Тогда

$$(\mu, v) \in O(\pi(\mathscr{V}'_\alpha), \mathscr{V}'_\alpha)$$

содержится в $\mathscr{P}_\alpha$. Таким образом, множества $\{O(\pi(\mathscr{V}'_\alpha)\mathscr{V}'_\alpha)\}$ образуют измельчение покрытия $\mathscr{P}_\alpha$.

Рассмотрим подмножество $\mathscr{L}$ расслоения $L(\mathscr{M}_G)$, состоящего из всех базисов над $\mathscr{S}_0$, в которых один из базисных векторов ортогонален к $\mathscr{S}_0$ и имеет длину $-3/\theta_0$, а остальные векторы образуют ортонормированный базис в $\mathscr{S}_0$. В силу компактности $\mathscr{L}$, оно может быть покрыто конечным числом множеств $\mathscr{V}'_\alpha$. Следовательно, компактно и $\mathscr{W}$, поскольку его можно покрыть конечным числом множеств $O(\pi(\mathscr{V}'_\alpha), \mathscr{V}'_\alpha)$.

По предложению 6.7.1 каждая точка области $D^+(\mathscr{S}_0)$ лежит в пределах собственного расстояния $-3/\theta_0$ вдоль направленной в будущее времениподобной геодезической, ортогональной к $\mathscr{S}_0$. Это означает, что $\mathrm{Exp}(\mathscr{W})$ содержит $D^+(\mathscr{S}_0)$. Пусть $\psi_*: B(\mathscr{M}_G) \to B(\mathscr{M})$ есть отображение, переводящее $(\lambda, u) \in B(\mathscr{M}_G)$ в $(\psi(\lambda), \psi_*(u)) \in B(\mathscr{M})$. Тогда $\psi_*(\mathscr{W})$ будет компактным подмножеством пространства $B(\mathscr{M})$, для которого

$$\mathrm{Exp}(\psi_*\mathscr{W}) \supset \psi(D^+(\mathscr{S}_0)).$$

Таким образом, если $\overline{D^+(\mathscr{S}_0)}$ не компактно, $\bar{\psi}(D^+(\mathscr{S}_0))$ тоже не компактно и, следовательно, пространство $(\mathscr{M}, \mathbf{g})$ не является b-ограниченным. □

Отсюда видно, что достаточно доказать некомпактность $\overline{D^+(\mathscr{S}_0)}$. Допустим, что оно компактно. Тогда горизонт $H^+(\mathscr{S}_0)$ тоже должен быть компактным. Ниже мы покажем, что отсюда вытекает равенство нулю расхождения изотропных геодезических всюду на $H^+(\mathscr{S}_0)$. Это невозможно, если плотность материи где-либо на $H^+(\mathscr{S}_0)$ отлична от нуля.

Лемма 8.5.5

Если $\mathscr{L}$ — частичная поверхность Коши и если горизонт $H^+(\mathscr{L})$ компактен, то изотропные геодезические сегменты-образующие $H^+(\mathscr{L})$ геодезически полны в направлении прошлого.

Из предложения 6.5.2 следует, что сегменты-образующие не имеют конечных точек в прошлом. Следовательно, они должны образовывать «почти замкнутые» кривые в компактном множестве $H^+(\mathscr{L})$. Если бы они образовывали настоящие замкнутые кривые, то можно было бы воспользоваться предложением 6.4.4 и показать, что при их неполноте в направлении прошлого вариацией в направлении прошлого из них можно получить замкнутые времениподобные геодезические. Однако это оказывается невозможным из-за того, что такие кривые должны были бы лежать в $D^+(\mathscr{L})$. В случае, когда изотропные геодезические образующие горизонта $H^+(\mathscr{L})$ только «почти замкнуты», доказательство аналогичное, хотя несколько более тонкое.

Введем направленное в будущее единичное векторное поле $\mathbf{V}$, геодезическое в некоторой окрестности $\mathscr{U}$ с компактным замыканием горизонта $H^+(\mathscr{L})$. Введем, как в предложении 6.4.4, положительно определенную метрику

$$g'(\mathbf{X}, \mathbf{Y}) = g(\mathbf{X}, \mathbf{Y}) + 2g(\mathbf{X}, \mathbf{V})\, g(\mathbf{Y}, \mathbf{V}).$$

Пусть t — параметр, измеряющий собственное расстояние в метрике $\mathbf{g}'$ вдоль изотропной геодезической, образующей сегмент γ горизонта $H^+(\mathscr{L})$, причем $t = 0$ в некоторой точке $q \in \gamma$. Тогда

$g(\mathbf{V}, \partial/\partial t) = -1/\sqrt{2}$. Поскольку у γ нет конечной точки в прошлом, параметр t не ограничен снизу. Пусть f и h даны равенствами

$$f\frac{\partial}{\partial t} = \frac{\mathrm{D}}{\partial t}\left(\frac{\partial}{\partial t}\right), \quad \frac{\partial}{\partial v} = h\frac{\partial}{\partial t},$$

где v — некоторый аффинный параметр. Допустим, что γ геодезически неполна в прошлом, тогда аффинный параметр

$$v = \int_0^t h^{-1}\,\mathrm{d}t'$$

при $t \to -\infty$ должен иметь нижнюю грань v_0. Рассмотрим теперь вариацию α кривой γ с вектором вариации $\partial/\partial u = -x\mathbf{V}$. Тогда

$$\frac{\partial}{\partial u} g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right)\Big|_{u=0} = \frac{1}{\sqrt{2}}\left(\frac{\mathrm{d}x}{\mathrm{d}t} + xh^{-1}\frac{\mathrm{d}h}{\mathrm{d}t}\right). \tag{8.3}$$

Поскольку $h \to \infty$ при $t \to -\infty$, можно найти такую ограниченную функцию $x(t)$, что выражение (8.3) будет отрицательным при всех $t \leqslant 0$. Однако этого не будет достаточно для уверенности в том, что вариация дала всюду времениподобную кривую, поскольку может случиться, что область значений u, для которой (8.3) остается отрицательным, стремится к нулю при $t \to -\infty$. Чтобы учесть это, рассмотрим вторую производную при вариации:

$$\frac{\partial^2}{\partial u^2} g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) = \frac{\partial}{\partial u}\left(g\left(\frac{\partial}{\partial t}, \frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial u}\right)\right) =$$

$$= g\left(\frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial u}, \frac{\mathrm{D}}{\partial t}\frac{\partial}{\partial u}\right) + g\left(\frac{\partial}{\partial t}, \frac{\mathrm{D}}{\partial t}\frac{\mathrm{D}}{\partial u}\frac{\partial}{\partial u}\right) +$$

$$+ g\left(\frac{\partial}{\partial t}, R\left(\frac{\partial}{\partial u}, \frac{\partial}{\partial t}\right)\frac{\partial}{\partial u}\right).$$

Выбирая $(\partial x/\partial u) = 0$ и пользуясь тем фактом, что $\mathbf{V}$ — геодезическое векторное поле в окрестности $\mathcal{U}$ горизонта $H^+(\mathscr{L})$, полученное выражение при $0 \leqslant u \leqslant \varepsilon$ можно привести к виду

$$-\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + x^2\left[g\left(\frac{\mathrm{D}\mathbf{V}}{\partial t}, \frac{\mathrm{D}\mathbf{V}}{\partial t}\right) + g\left(\frac{\partial}{\partial t}, R\left(\mathbf{V}, \frac{\partial}{\partial t}\right)\mathbf{V}\right)\right].$$

В любом ортонормированном относительно метрики $\mathbf{g}'$ базисе компоненты тензора Римана и ковариантных производных (относительно $\mathbf{g}$) $\mathbf{V}$ будут на $\mathcal{U}$ ограниченными. Таким образом, существует такая постоянная $C > 0$, что

$$\frac{\partial^2}{\partial u^2} g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) \leqslant C^2 x^2 g'\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right).$$

Далее,

$$\frac{\partial}{\partial u}\left(g\left(\mathbf{V}, \frac{\partial}{\partial t}\right)\right) = -\frac{\mathrm{d}x}{\mathrm{d}t},$$

так что

$$g\left(\mathbf{V}, \frac{\partial}{\partial t}\right) = -\frac{1}{\sqrt{2}} - u\frac{\mathrm{d}x}{\mathrm{d}t}.$$

Отсюда при $0 \leqslant u \leqslant \varepsilon$

$$g'\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) = g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) + 1 - 2\sqrt{2}\, u\frac{\mathrm{d}x}{\mathrm{d}t} + 2u^2\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 \leqslant$$
$$\leqslant g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) + d,$$

где $d = 2\sqrt{2}\,\varepsilon C_1 + 2\varepsilon^2 C_1^2 + 1$ и C_1 — верхняя грань для $|\mathrm{d}x/\mathrm{d}t|$. Таким сбразом, имеем

$$\frac{\partial^2 y}{\partial u^2} \leqslant C^2 x^2 (y + d)$$

и

$$\left.\frac{\partial y}{\partial u}\right|_{u=0} = \frac{1}{\sqrt{2}} h^{-1}\frac{\mathrm{d}}{\mathrm{d}t}(hx), \quad y|_{u=0=0},$$

где $y = g(\partial/\partial t, \partial/\partial t)$. Поэтому

$$y \leqslant d(\operatorname{ch} Cxu - 1) + a \operatorname{sh} Cxu \leqslant \operatorname{sh} Cxu\left(d \operatorname{th}\frac{1}{2} Cxu + a\right),$$

где $a = \left(1/\sqrt{2}\right) C^{-1}\mathrm{d}(\ln hx)/\mathrm{d}t$.

Возьмем теперь

$$x = h^{-1}\left[-\int_t^0 h^{-1}\,\mathrm{d}t' + K\right]^{-1},$$

где

$$K = 2\int_{-\infty}^0 h^{-1}\,\mathrm{d}t';$$

тогда $a = -\left(1/\sqrt{2}\right) C^{-1} hx$. Поскольку функция $f = -h^{-1}(\mathrm{d}h/\mathrm{d}t)$ ограничена на компактном множестве $H^+(\mathscr{L})$ и поскольку предполагается, что интеграл

$$\int_t^0 h^{-1}\,\mathrm{d}t' = -v$$

сходится при $t \to -\infty$, x и $|\mathrm{d}x/\mathrm{d}t|$ должны иметь верхнюю грань, а h при $-\infty < t \leqslant 0$ — положительную нижнюю грань. Тогда для $0 < u < \min(\varepsilon, 2C^{-2}d^{-1}C_2)$ $y < 0$, если $-\infty < t \leqslant 0$.

Другими словами, вариация α должна дать непродолжимую в прошлое времениподобную кривую, лежащую в $\operatorname{int} D^+(\mathscr{L})$

и полностью заключенную в компактном множестве $\overline{\mathcal{U}}$. Но это невозможно, поскольку по лемме 6.6.5 в $\mathrm{int}\, D^+(\mathscr{L})$ выполняется условие сильной причинности. Следовательно, кривая γ должна быть геодезически полной в направлении прошлого. □

Рассмотрим расхождение $\hat{\theta}$ касательных векторов $\partial/\partial t$ к изотропным геодезическим образующим поверхности $H^+(\mathscr{S}_0)$. Допустим, что $\hat{\theta} > 0$ в некоторой точке q образующей γ и пусть $\mathscr{T}$ — пространственноподобная 2-поверхность в некоторой окрестности q на $H^+(\mathscr{S}_0)$, проходящая через q. Образующие поверхности $H^+(\mathscr{S}_0)$ будут ортогональны к $\mathscr{T}$ и будут сходиться в прошлое. Тогда по условию (1) и доказанной выше лемме должна быть точка $r \in \gamma$, сопряженная $\mathscr{T}$ вдоль γ (предложение 4.4.6). Точки γ, лежащие за точкой r, могли бы быть соединены с $\mathscr{T}$ времениподобными кривыми (предложение 4.5.14). Но это невозможно в силу того, что $H^+(\mathscr{S}_0)$ — ахрональное множество. Следовательно, $\hat{\theta} \leqslant 0$ на $H^+(\mathscr{S}_0)$.

Рассмотрим теперь семейство дифференцируемых отображений

$$\beta_z : H^+(\mathscr{S}_0) \to H^+(\mathscr{S}_0),$$

определяемых как перенос точки $q \in H^+(\mathscr{S}_0)$ на расстояние z (измеренное в метрике $\mathbf{g}'$) в прошлое вдоль изотропной геодезической, проходящей через q. Пусть $\mathrm{d}A$ — площадь малого элемента поверхности $H^+(\mathscr{S}_0)$, измеренная в метрике $\mathbf{g}'$. При отображении β_z

$$\frac{\mathrm{d}}{\mathrm{d}z}\,\mathrm{d}A = -\hat{\theta}\,\mathrm{d}A.$$

Следовательно,

$$\frac{\mathrm{d}}{\mathrm{d}z}\int\limits_{\beta_z(H^+(\mathscr{S}_0))} \mathrm{d}A = -\int\limits_{\beta_z(H^+(\mathscr{S}_0))} \hat{\theta}\,\mathrm{d}A. \tag{8.4}$$

Но β_z отображает $H^+(\mathscr{S}_0)$ в $H^+(\mathscr{S}_0)$ [и на $H^+(\mathscr{S}_0)$, если образующие сегменты не имеют будущих конечных точек]. Таким образом, интеграл (8.4) должен быть меньше или равен нулю. Вместе с предыдущим результатом это означает, что на $H^+(\mathscr{S}_0)$ $\hat{\theta} = 0$. Согласно уравнению распространения (4.35), это возможно лишь тогда, когда $R_{ab}K^aK^b = 0$ всюду на $H^+(\mathscr{S}_0)$, где $\mathbf{K}$ — касательный вектор изотропной геодезической образующей. Однако, согласно теореме сохранения (разд. 4.3), из условия (5) следует, что величина $T_{ab}K^aK^b$ где-нибудь на $H^+(\mathscr{S})$ отлична от нуля и, согласно уравнениям Эйнштейна (с Λ-членом или без него), $T_{ab}K^aK^b = R_{ab}K^aK^b$.

[Строго говоря, требуемая форма теоремы сохранения несколько отличается от приведенной в разд. 4.3. Поскольку нет

каких-либо подходящих пространственноподобных поверхностей, пересекающих $H^+(\mathscr{S}_0)$, мы используем семейство поверхностей, одной из которых является $H^+(\mathscr{S}_0)$, а остальные — пространственноподобные. Эти поверхности можно задать, взяв значение функции t в точке $p \in \bar{D}^+(\mathscr{S}_0)$ равным собственному объему $J^+(p) \cap D^+(\mathscr{S}_0)$, взятому со знаком минус. Поскольку вектор $t_{;a}$ становится изотропным на $H^+(\mathscr{S}_0)$, утверждение о существовании такой постоянной $C > 0$, при которой на $\bar{D}^+(\mathscr{S}_0)$ выполняется неравенство

$$T^{ab}t_{;ab} \leqslant CT^{ab}t_{;a}t_{;b},$$

более не является обязательно справедливым. Однако, если V^a — времениподобное векторное поле на $\bar{D}^+(\mathscr{S}_0)$, существует постоянная C, для которой

$$T^{ab}t_{;ab} \leqslant CT^{ab}(t_{;a}t_{;b} + t_{;a}V_b)$$

и

$$T^{ab}V_{a;b} \leqslant CT^{ab}(t_{;a}t_{;b} + t_{;a}V_b).$$

Затем можно поступить, как в разд. 4.3, используя $T^{ab}(t_{;ab} + V_{a;b})$ вместо $T^{ab}t_{;ab}$ и доказывая, что величина $T^{ab}(t_{;a}t_{;b} + t_{;a}V_b)$ не может быть равной нулю на $H^+(\mathscr{S}_0)$, если она отлична от нуля на $\mathscr{S}_0$. Тогда наше утверждение следует из условия (5).] □

Глава 9

Гравитационный коллапс и черные дыры

В этой главе мы покажем, что звезды массой более $\approx 1{,}5\, M_\odot$ должны претерпеть коллапс после того, как их ядерное топливо иссякнет. Если начальные условия не слишком асимметричны, то будут выполнены условия теоремы 2 и, следовательно, будет иметь место сингулярность. Однако эта сингулярность, по всей вероятности, скрыта от глаз внешнего наблюдателя, который видит только «черную дыру» там, где когда-то была звезда. Мы выведем ряд свойств таких черных дыр и покажем, что в результате их эволюции, по-видимому, устанавливается ситуация, описываемая решением Керра.

Коллапс звезд рассмотрен в разд. 9.1, в котором показано, как может возникнуть замкнутая ловушечная поверхность вокруг любой достаточно большой сферической звезды на поздних стадиях ее эволюции. В разд. 9.2 мы исследуем горизонт событий, который, по-видимому, образуется вокруг такого коллапсирующего тела. В разд. 9.3 рассматривается конечное стационарное состояние, к которому приводит решение вне горизонта. Вероятно, это должно быть одно из решений семейства Керра. В таком случае можно наложить определенные ограничения на количество энергии, способное выделиться в рассматриваемом процессе.

Для более глубокого изучения черных дыр см.: Black Holes. Les Houches 1972, B. Dewitt, ed. Gordon a. Breach, 1973.

9.1. Коллапс звезды

Вне статического сферически симметричного тела, такого, как звезда, решением уравнений Эйнштейна с необходимостью является та часть одной из асимптотически плоских областей решения Шварцшильда, для которой r больше некоторого значения r_0, соответствующего поверхности звезды. Это решение можно сшить при $r < r_0$ с решением, которое зависит от конкретного радиального распределения плотности и давления в звезде. Даже когда звезда не статическая, но сохраняет свою сферическую симметрию, внешнее решение будет фактически частью решения Шварцшильда, обрезанного на поверхности звезды. (Это и есть теорема Биркгофа, доказательство которой дано в приложении Б). Если звезда статическая, то r_0 должно

быть больше $2m$ (радиус Шварцшильда). Это следует из того, что поверхность статической звезды должна соответствовать орбите времениподобного вектора Киллинга, а в решении Шварцшильда времениподобный вектор Киллинга существует только при $r > 2m$. Если r_0 меньше $2m$, то поверхность звезды должна расширяться или сжиматься. Чтобы дать представление о величине радиуса Шварцшильда, отметим, что для Земли он равен ~ 1 см, а для Солнца 3,0 км. Отношения шварцшильдова радиуса к радиусу для Земли и Солнца равны соответственно $\sim 10^{-9}$ и $2 \cdot 10^{-6}$. Таким образом, радиусы нормальных звезд далеки от соответствующих шварцшильдовых радиусов.

Эволюция типичной звезды включает длительную ($\sim 10^9$ лет) квазистатическую фазу, в которой расходуется ядерное топливо, а тепловое давление и давление излучения уравновешивают силу тяжести. Но когда ядерное топливо выгорает, звезда начинает остывать, давление падает, и вследствие этого звезда сжимается. Предположим теперь, что давление не может остановить это сжатие, прежде чем радиус звезды не станет меньше ее радиуса Шварцшильда (ниже мы увидим, что это, по всей видимости, возможно для звезды с массой больше некоторого определенного значения). Тогда, поскольку решение вне звезды шварцшильдово, вокруг звезды будет существовать замкнутая ловушечная поверхность $\mathscr{T}$ (рис. 54) и, согласно теореме 2, должна существовать сингулярность, если только не нарушается причинность и выполняются соответствующие энергетические условия. Существенно, однако, что если звезда не является в точности сферически симметричной, при некотором не слишком большом отклонении звезды от сферической симметрии замкнутая ловушечная поверхность все равно возникает. Это следует из доказанной в разд. 7.5 устойчивости решения задачи Коши. В самом деле, мы можем рассматривать данное решение как развитие из некоторой частичной поверхности Коши $\mathscr{H}$ (рис. 55). Если теперь мы изменим начальные данные на малую величину в компактной области $J^-(\mathscr{T}) \cap \mathscr{H}$, то новое развитие из поверхности $\mathscr{H}$ в области $J^+(\mathscr{H}) \cap J^-(\mathscr{T})$ будет достаточно близким к прежнему, так что в возмущенном решении вокруг звезды по-прежнему будет существовать замкнутая ловушечная поверхность. Таким образом, мы показали, что имеется множество начальных данных ненулевой меры, которые приводят к замкнутой ловушечной поверхности и, следовательно, согласно теореме 2, к сингулярности.

Звезда может вращаться или обладать магнитным полем, и это две главные причины, по которым ее сферическая симметрия может быть нарушена. Изучая решение Керра, мы можем составить некоторое представление о том, каким быстрым должно быть вращение без того, чтобы оно воспрепятствовало

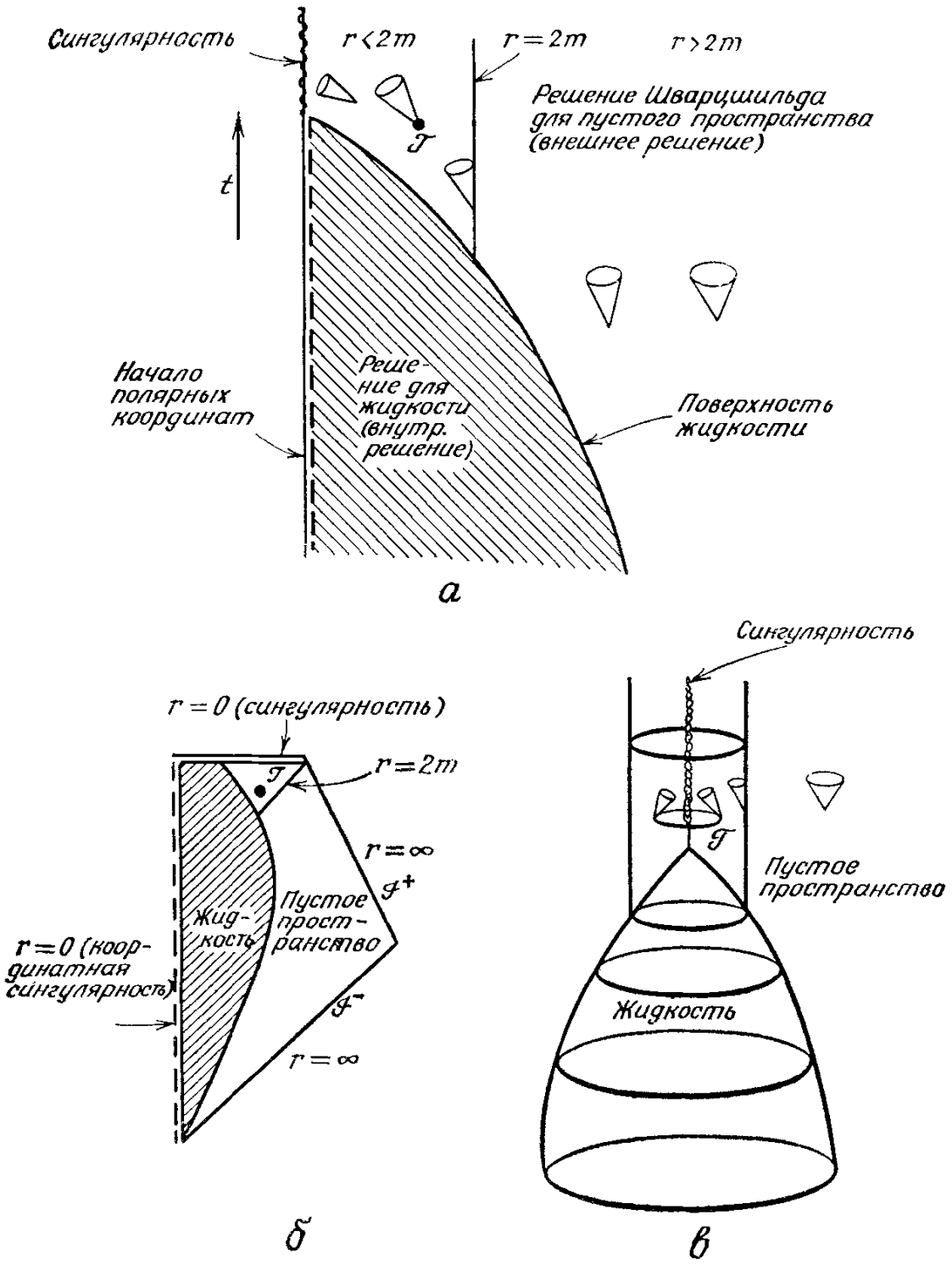


Рис. 54. Коллапс сферической звезды. *а* — диаграмма Финкельстейна [плоскость (r, t)] коллапсирующего сферически симметричного жидкого шара. Каждая точка изображает 2-сферу; *б* — диаграмма Крускала коллапсирующего жидкого шара; *в* — диаграмма коллапса, на которой опущена только одна пространственная координата.

появлению ловушечной поверхности. Предполагается, что решение Керра можно рассматривать как внешнее решение для тела массы m с моментом количества движения $L = am$. При $a < m$ имеются замкнутые ловушечные поверхности, а при $a > m$ они

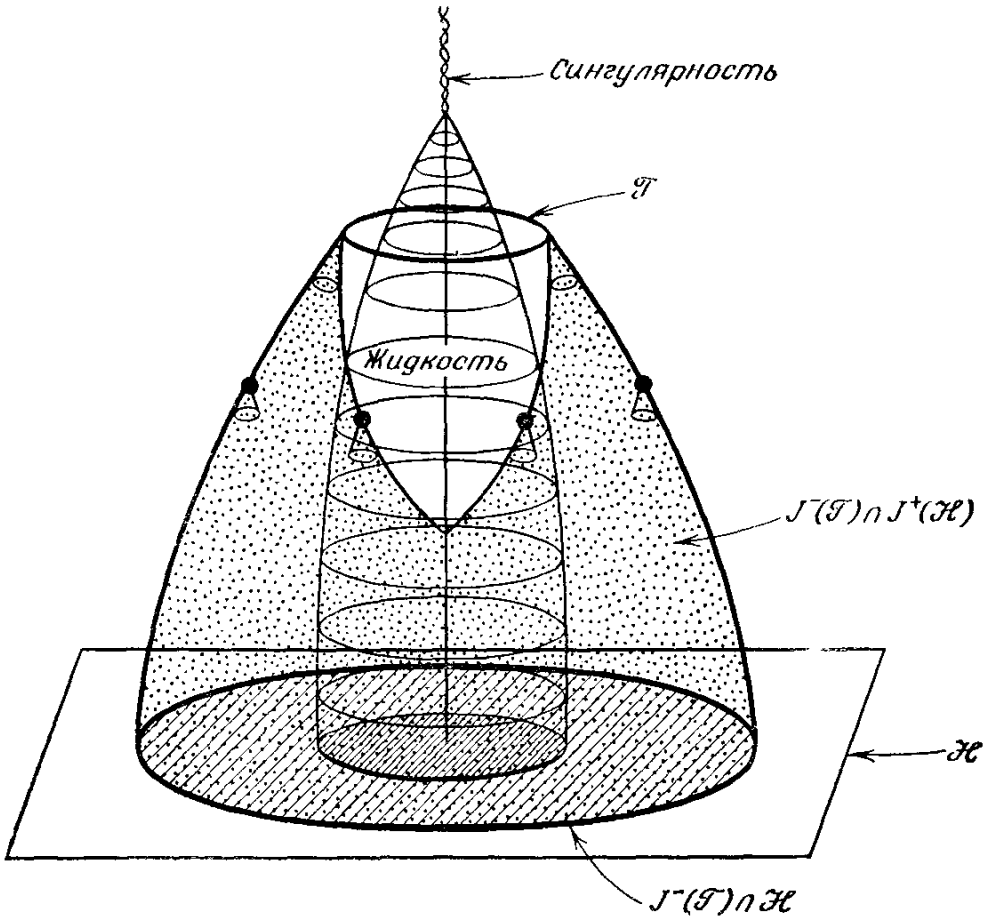


Рис. 55. Коллапс сферической звезды (как на рис. 54, *в*). Изображена частичная поверхность Коши $\mathcal{H}$. Начальные данные в компактной области $J^-(\mathcal{T}) \cap \mathcal{H}$ поверхности $\mathcal{H}$ ведут к появлению замкнутой ловушечной поверхности $\mathcal{T}$ в компактной области $J^-(\mathcal{T}) \cap J^+(\mathcal{H})$.

отсутствуют. Поэтому можно ожидать, что при моменте количества движения звезды больше квадрата ее массы вращение может прекратить сжатие звезды еще до появления замкнутой ловушечной поверхности. Подобный вывод следует из того, что при $L = m^2$ и сохранении момента количества движения в течение коллапса скорость вращения поверхности звезды приближается к скорости света, когда радиус звезды близок к ее радиусу Шварцшильда. В настоящее время у многих звезд моменты количества движения больше квадратов их масс (для Солнца $L \sim m^2$). Однако, по-видимому, есть основания ожидать, что в результате торможения магнитным полем и грави-

тационного излучения некоторая часть момента количества движения теряется. Поэтому положение таково, что в некоторых звездах, а вероятно в большинстве, момент количества движения не препятствует появлению замкнутых ловушечных поверхностей, а значит и появлению сингулярности.

При почти сферическом коллапсе магнитное поле **B**, вмороженное в звезду, будет расти как $\rho^{2/3}$, где ρ — плотность вещества звезды; следовательно, магнитное давление будет пропорционально $\rho^{4/3}$. Эта скорость роста настолько мала, что, если магнитное поле оказалось несущественным для поддержания равновесия звезды в исходном состоянии, оно никогда не станет достаточно сильным, чтобы заметно повлиять на коллапс.

Чтобы понять, почему звезда с массой больше некоторого значения, в которой выгорело ядерное топливо, не может противостоять собственному тяготению, проведем качественное исследование (основанное на неопубликованной работе Картера) уравнения состояния вещества при нулевой температуре.

В горячем веществе существуют давление, вызванное тепловым движением атомов, и давление излучения. Однако в холодном веществе при плотностях, меньших ядерной плотности ($\sim 10^{14}$ г/см^3), существенную роль играет только давление, связанное с принципом Паули. Чтобы оценить величину этого давления, рассмотрим концентрацию фермионов n массы m. По принципу Паули каждый фермион будет эффективно занимать объем n^{-1}. Таким образом, согласно принципу неопределенности, фермион будет иметь пространственную компоненту импульса порядка $\hbar n^{1/3}$. Если фермионы нерелятивистские, т. е. если $\hbar n^{1/3} < m$, то их скорость будет порядка $\hbar n^{1/3}/m$; если же они релятивистские (т. е. $\hbar n^{1/3} > m$), то их скорость практически равна единице (т. е. скорости света). Давление будет порядка произведения импульса на скорость и плотность числа частиц, т. е. $\sim \hbar^2 n^{5/3} m^{-1}$ при $\hbar n^{1/3} < m$ и $\sim \hbar n^{4/3}$ при $\hbar n^{1/3} > m$. Когда вещество нерелятивистское, главный вклад в давление вырождения вносят электроны, поскольку величина m^{-1} для них больше, чем для барионов. Однако при высоких плотностях, когда частицы становятся релятивистскими, давление не зависит от массы частиц, вызывающих его, а зависит только от их числа в единице объема.

Для малых холодных тел собственным тяготением можно пренебречь, и давление вырождения вещества будет уравновешено электростатическими силами притяжения между соседними частицами, образующими ту или иную решетку. (Мы предполагаем, что имеется равное число положительно и отрицательно заряженных частиц и приблизительно равное число электронов и барионов.) Эти силы вызывают отрицательное давление порядка $e^2 n^{4/3}$. Таким образом, плотность холодного тела

малых размеров будет равна по порядку величины

$$e^6 m_e^3 m_n \hbar^{-6} (\sim 1 \text{ г/см}^3), \tag{9.1}$$

где m_e — масса покоя электрона и m_n — масса покоя нуклона.

Для больших тел самогравитация становится существенной. Чтобы получить точное решение, необходимо проинтегрировать соответствующие уравнения Эйнштейна. Однако важные особенности можно установить более простым путем, проводя качественные расчеты на основе ньютоновской теории. В звезде массы M и радиуса r_0 сила тяготения, действующая на типичный единичный объем, имеет порядок $(M/r_0^2) n m_n$, где $n m_n \approx M/r_0^3$ — плотность. Эта сила уравновешена градиентом давления P/r_0, где P — среднее давление в звезде. Таким образом,

$$P = M^2/r_0^4 \approx M^{2/3} n^{4/3} m_n^{4/3}.$$

Если плотность достаточно мала и основной вклад в давление дают вырожденные нерелятивистские электроны, то

$$P = \hbar^2 n^{5/3} m_e^{-1} = M^{2/3} n^{4/3} m_n^{4/3},$$

так что

$$n = M^2 m_n^4 m_e^3 \hbar^{-6}.$$

Эта формула верна для таких тел, для которых она дает значение n больше, чем величина (9.1), и меньше, чем $m_e^3 \hbar^{-3}$, т. е. при $e^3 m_n^{-2} < M < \hbar^{3/2} m_n^{-2}$. Такие звезды называются белыми карликами.

Если плотность настолько велика, что электроны являются релятивистскими, т. е. $n > m_e^3 \hbar^{-3}$, то давление будет определяться релятивистским соотношением; следовательно,

$$P = \hbar n^{4/3} = M^{2/3} n^{4/3} m_n^{4/3}.$$

Теперь формула для давления не содержит n. Таким образом, можно было заключить, что мы получаем звезду массы

$$M_L = \hbar^{3/2} m_n^{-2} \approx 1{,}5 M_\odot,$$

которая может иметь любую плотность выше $m_e^3 m_n \hbar^{-3}$, т. е. любой радиус меньше $\hbar^{3/2} m_n^{-1} m_e^{-1}$. Давление вырожденных электронов не может предотвратить сжатие звезды массой больше M_L.

В действительности, когда электроны становятся релятивистскими, они стимулируют обратный бета-распад, образуя нейтроны

$$e^- + p \to \nu_e + n.$$

При этом содержание электронов в звезде уменьшается и давление вырожденных электронов падает, что вызывает сжатие

звезды, и электроны становятся еще более релятивистскими. Такая ситуация неустойчива, и процесс будет продолжаться до тех пор, пока почти все электроны и протоны не превратятся в нейтроны. На этой стадии может вновь наступить равновесие: звезда будет поддерживаться давлением вырожденных нейтронов. Если нейтроны нерелятивистские, получаем

$$n = M^2 m_n^{\ 7} \hbar^{-6},$$

если же они релятивистские, то звезда по-прежнему должна иметь массу M_L, а радиус, меньший или равный $\hbar^{3/2} m^{-2}{}_n$. Однако $M_L/\hbar^{3/2} m^{-2}{}_n = 1$, и, следовательно, такая звезда находится вблизи общерелятивистского предела $M_L/R \approx 2$.

Итак, мы приходим к выводу, что сжатие звезды массы, большей M_L, не может быть предотвращено давлением вырождения электронов или нейтронов. Чтобы показать это строго, рассмотрим ньютоновское уравнение равновесия

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\rho M(r) r^{-2}, \tag{9.2}$$

где

$$M(r) = 4\pi \int_0^r \rho r'^2 \, \mathrm{d}r$$

— масса, заключенная внутри сферы радиуса r. Умножим обе части равенства (9.2) на r^4 и проинтегрируем по частям от 0 до r_0. С учетом того, что $p = 0$ при $r = r_0$, получим

$$\int_0^{r_0} pr^3 \, dr = (M^2(r_0))^2/8\pi.$$

С другой стороны, всегда $\mathrm{d}p/\mathrm{d}r \leqslant 0$, и потому

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\int_0^r pr'^3 \, \mathrm{d}r'\right)^{3/4} = \frac{3}{4}\left(\int_0^r pr'^3 \, \mathrm{d}r'\right)^{-1/4} pr^3 =$$

$$= \frac{3}{4}\left(\frac{1}{4} pr^4 - \frac{1}{4}\int_0^r \frac{\mathrm{d}p}{\mathrm{d}r'} r'^4 \, \mathrm{d}r'\right)^{-1/4} pr^3 < \frac{3\sqrt{2}}{4} p^{3/4} r^2.$$

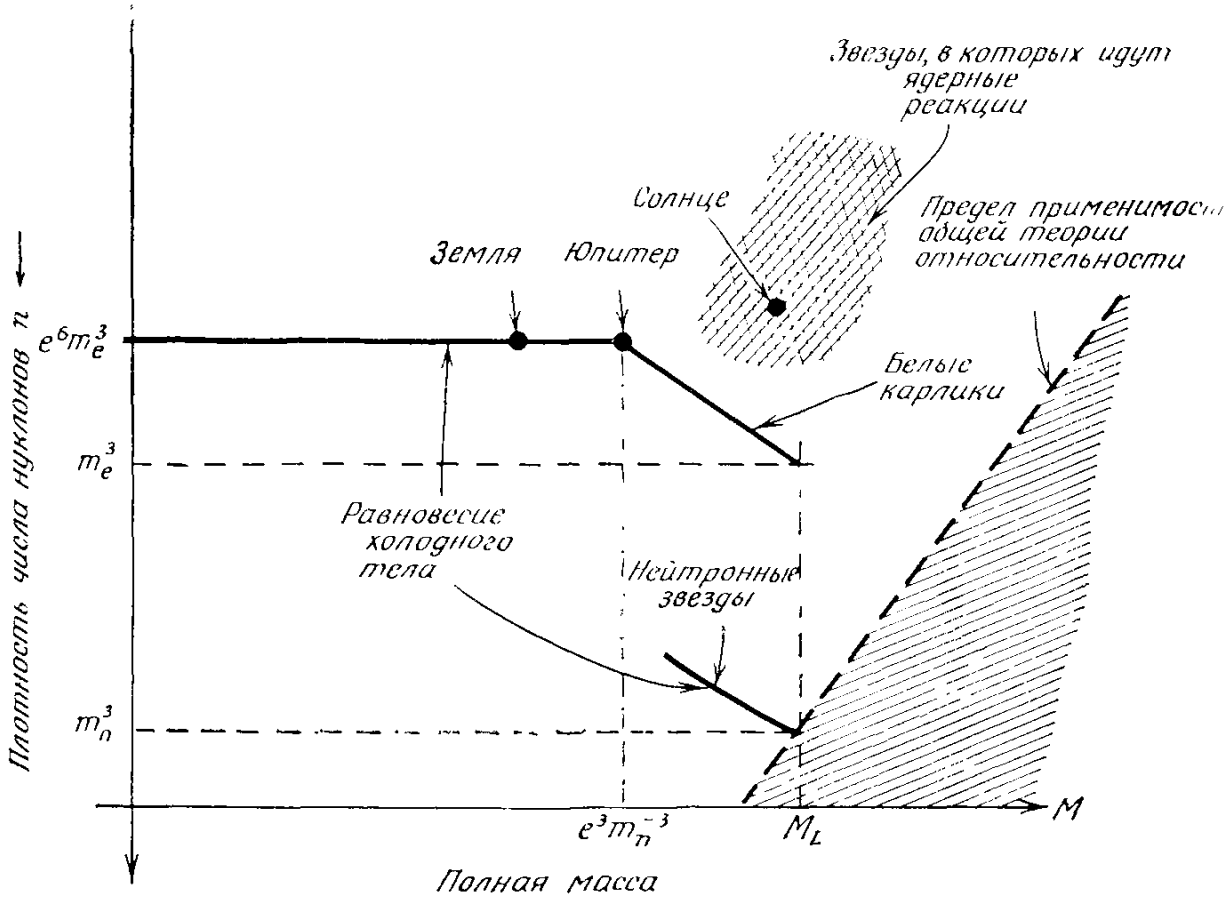
Поскольку p никогда не превышает $\hbar n^{4/3}$,

$$\int_0^{r_0} pr^3 \, \mathrm{d}r < \hbar\left(\int_0^{r_0} nr^2 \, \mathrm{d}r\right)^{4/3} = \hbar (M(r_0))^{4/3} (4\pi m_n)^{-4/3}.$$

Отсюда масса $M(r_0)$ должна быть меньше $(8\hbar)^{3/2} (4\pi)^{-1/2} m_n^{-2}$, т. е.

$$M(r_0) < 8\hbar^{3/2} m_n^{-2}.$$

Мы подытожили эти результаты на рис. 56. На диаграмме нанесена средняя концентрация нейтронов в зависимости от массы тела M. Жирная линия изображает приближенно равновесную конфигурацию холодных тел. В горячих телах помимо давления вырожденного вещества будут существовать тепло-



Р и с. 56. Плотность числа нуклонов в зависимости от полной массы статического тела M. Жирная линия — равновесное состояние холодных тел; горячие тела при соответствующих температурах могут быть в равновесии выше этой линии. Общая теория относительности запрещает статическое состояние любых тел в заштрихованной области.

вое давление и давление излучения, поэтому такие тела могут быть в равновесии в области выше жирной линии. Жирная штриховая линия справа соответствует равенству

$$M/r_0 \left(= M^{2/3} n^{1/3} m_p^{1/3}\right) = 2.$$

В области справа от этой линии вообще нет равновесных состояний, и эта область соответствует звезде, ушедшей внутрь своей сферы Шварцшильда. Далеко влево от этой линии различием между ньютоновской теорией и общей теорией относительности можно пренебречь. Вблизи нее нужно учитывать общерелятивистские эффекты. Для статического сферически симметричного тела, состоящего из идеальной жидкости, эйн-

штейновские уравнения поля можно свести (см. приложение Б) к такому уравнению:

$$\frac{dp}{dr} = -\frac{(\mu + p)\,[\hat{M}(r) + 4\pi r^3 p]}{r\,[r - 2\hat{M}(r)]}, \tag{9.3}$$

где радиальная координата такова, что площадь 2-поверхности $\{r = \text{const},\ t = \text{const}\}$ равна $4\pi r^2$. Величина $\hat{M}(r)$ теперь определяется как

$$\int_0^r 4\pi r^2 \mu\, dr,$$

где $\mu = \rho(1 + \varepsilon)$ — плотность полной энергии, $\rho = nm_n$ и ε — релятивистская добавка к массе, связанная со спином фермионов. Величина $\hat{M}(r_0)$ равна шварцшильдовой массе $\hat{M}$ для внешнего решения Шварцшильда при $r > r_0$. Для звезды конечных размеров она меньше, чем сохраняющаяся масса

$$\tilde{M} = \int_0^{r_0} \frac{4\pi\rho r^2\, dr}{\left(1 - 2\frac{M}{r}\right)^{1/2}} = Nm_n,$$

где N — полное число нуклонов в звезде, так как разность $(\tilde{M} - \hat{M})$ соответствует количеству энергии, излученной на бесконечность с момента образования звезды из диффузного первоначально покоящегося вещества. Практически эта разность никогда не превышает нескольких процентов и уж никак не больше $2\hat{M}$: Бонди [11] показал, что при $\mu > 0$, $p > 0$ и μ, убывающем к поверхности звезды, $(1 - 2\hat{M}/r)^{1/2}$ не может быть меньше $1/3$, а при $p \leqslant \mu$ — меньше $1/2$. Отсюда $\hat{M} < M < 3\,\hat{M}$.

Сравнивая соотношение (9.3) с (9.2), в которое вместо ρ и $\hat{M}$ подставлены соответственно μ и M, мы видим, что все дополнительные члены в правой части уравнения (9.3) при $\varepsilon \geqslant 0$ и $p \geqslant 0$ отрицательны. Следовательно, так как в ньютоновской теории сжатие холодной звезды массы $M > M_L$ неизбежно, то и в общей теории относительности сжатие холодной звезды шварцшильдовой массы $\hat{M} > M_L$ также неизбежно. Это означает, что холодная звезда, содержащая более $3\,M_L/m_n$ нуклонов, не может противостоять сжатию. Практически дополнительные члены в (9.3) означают, что предельное число нуклонов меньше M_L/m_n.

Говоря о нейтронных звездах, мы пренебрегали ядерными силами. Учет их несколько изменит положение линии равновесия для нейтронных звезд на рис. 56 [20, 68, 166, 167], однако

даже при учете ядерных сил звезда, содержащая больше M_L/m_n нуклонов, не придет к какому-либо равновесию при нулевой температуре. Это связано с тем, что точка, в которой нейтроны в звезде массы M_L становятся релятивистскими, почти совпадает с релятивистским пределом $M/R \approx 2$. Таким образом, звезда, содержащая несколько больше чем M_L/m_n нуклонов, достигнет ядерных плотностей, уже оказавшись внутри своей сферы Шварцшильда.

На рис. 56 эволюционный путь звезды будет лежать на вертикальной линии, если только звезда не потеряет в том или ином процессе значительное количество своего вещества. Звезда конденсируется из газового облака и разогревается вследствие сжатия газа. Если ее масса меньше некоторого значения порядка $10^{-2}\ M_L$, то температура никогда не возрастает настолько, чтобы начались ядерные реакции; звезда в конечном счете излучит свой запас тепла, и установится состояние, в котором ее тяготение уравновешивается давлением вырожденного газа нерелятивистских электронов. Если масса больше $\sim 10^{-2}\ M_L$, то температура достигнет значения, достаточного, чтобы началась ядерная реакция превращения водорода в гелий. Энергия, выделяемая в этой реакции, уравновесит потерю энергии вследствие излучения, и звезда длительный период [$\sim 10^{10}(M_L/M)^2$ лет] будет пребывать в квазистатическом равновесии. Когда водород в сердцевине звезды иссякнет, сердцевина сожмется и температура возрастет. При этом могут начаться новые ядерные реакции, в которых гелий в недрах звезды будет превращаться в более тяжелые элементы. Однако энергия, выделяющаяся при этих превращениях, не очень велика, и сердцевина не может оставаться слишком долго в этой фазе. Если масса меньше M_L, то звезда в итоге станет белым карликом и будет удерживаться давлением вырожденных нерелятивистских электронов или, возможно, превратится в нейтронную звезду и будет удерживаться давлением вырожденных нейтронов. Но когда масса заметно больше M_L, уже не существует равновесного низкотемпературного состояния. Поэтому звезда должна или уйти внутрь своей сферы Шварцшильда, или каким-то образом сбросить столько вещества, чтобы ее масса стала меньше M_L.

Выброс вещества наблюдается в сверхновых и планетарных туманностях, но в теории этого процесса еще много неясного. Звезды с массой 20 M_L, вероятно, способны сбросить бо́льшую часть своей массы и превратиться в белые карлики или нейтронные звезды с массой меньше M_L — вот все, что можно сказать на основании расчетов [2, 35, 98, 172, 176]. Однако вряд ли звезда с массой, превышающей 20 M_L, способна потерять более 95% своего вещества, и следует полагать, что внутренняя об-

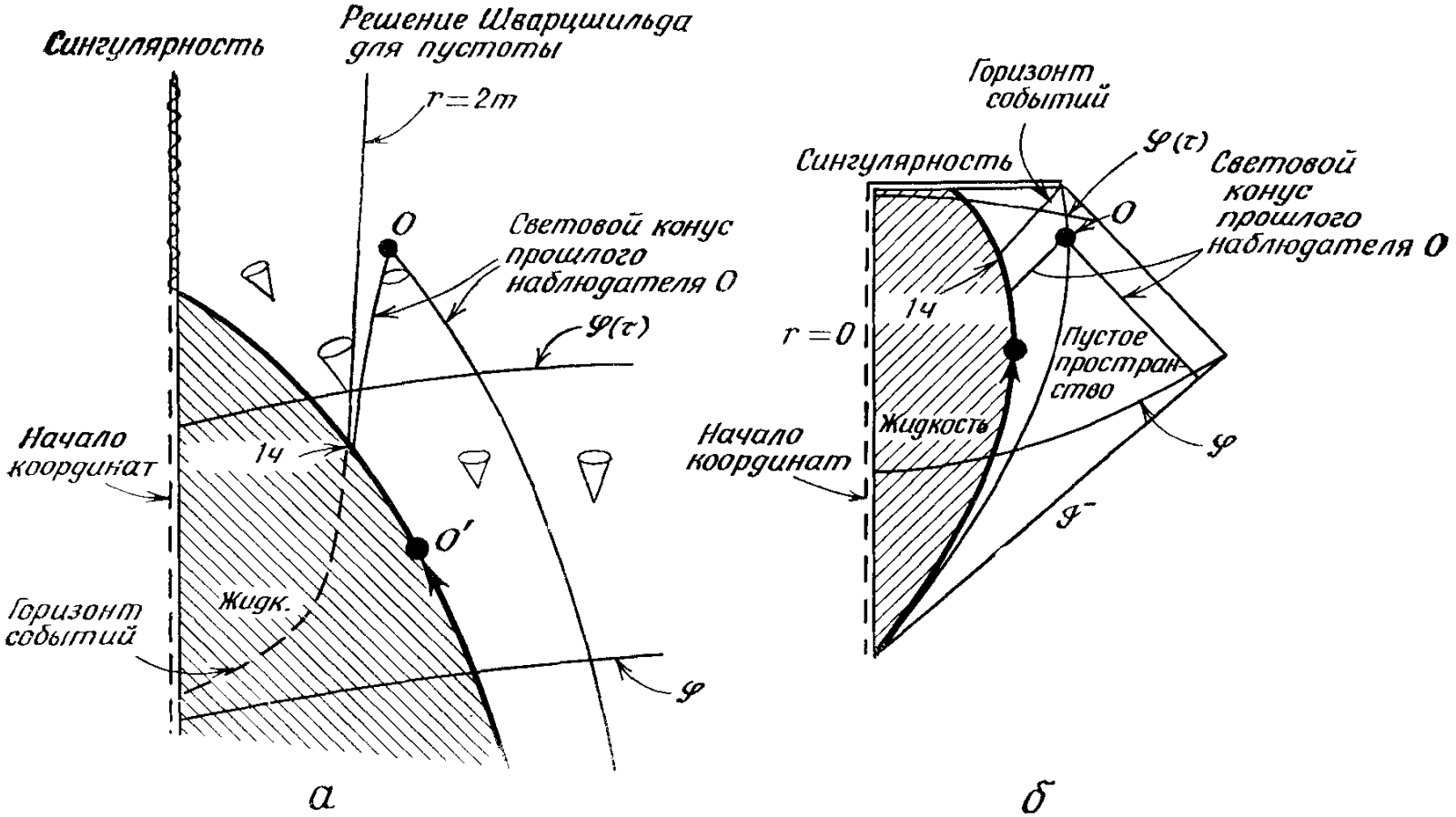
ласть этой звезды в любом случае должна сколлапсировать внутрь своей сферы Шварцшильда. (Последние вычисления действительно указывают на то, что звезды с массой $M > 5\ M_L$ не в состоянии сбросить достаточное количество вещества, чтобы избежать релятивистского коллапса.)

Обратимся к бо́льшим массам и рассмотрим тело массой $\sim 10^8\ M_L$. Если оно сколлапсирует до своего шварцшильдова радиуса, то плотность будет всего лишь порядка 10^{-4} г/см^3 (меньше плотности воздуха). Если вещество первоначально будет холодным, рост температуры будет недостаточен ни для уравновешивания сил тяготения, ни для «зажигания» ядерного горючего. Поэтому невозможна потеря массы и нет никакой неопределенности в выборе уравнения состояния. Кроме того, из этого примера видно, что условия, при которых тело пересекает свою сферу Шварцшильда, совсем не обязательно должны быть в том или ином смысле экстремальными.

Итак, по-видимому, некоторые тела с массой $> M_L$ (а, вероятно, большинство) коллапсируют в конце концов внутрь своих сфер Шварцшильда и порождают замкнутые ловушечные поверхности. В нашей Галактике имеется по меньшей мере 10^9 звезд с массой, превышающей M_L. Таким образом, существует огромное число ситуаций, для которых теорема 2 предсказывает существование сингулярностей. Наблюдаемые следствия коллапса звезды мы обсудим в последующих разделах.

9.2. Черные дыры

Как будет выглядеть коллапсирующее тело для удаленного наблюдателя O? На этот вопрос можно дать ответ, если коллапс строго сферически симметричен, так как тогда решение вне тела будет шварцшильдово. В этом случае наблюдатель O' на поверхности звезды уйдет в определенный момент времени (скажем, в 1 ч по своим часам) внутрь сферы $r = 2\ m$, причем он не заметит в этот момент ничего особенного. Однако после пересечения сферы $r = 2\ m$ он станет невидим наблюдателю O (рис. 57), остающемуся вне этой сферы. Как бы долго наблюдатель O ни ждал, он никогда не увидит O', после того как часы последнего покажут 1 ч. Вместо этого O будет наблюдать, как часы O' явно отстают и асимптотически приближаются к моменту 1 ч. Это значит, что частота света, принимаемого наблюдателем O от O', становится все меньше и меньше и соответственно интенсивность света убывает все сильнее. Таким образом, хотя поверхность звезды никогда не исчезает из поля зрения наблюдателя O, она скоро станови́тся столь слабой, что практически перестает быть видимой. Точнее, наблюдатель O увидит сначала, как уменьшится яркость центра диска звезды,



Р и с. 57. Наблюдатель O, который никогда не упадет внутрь коллапсирующей жидкой сферы, никогда не увидит в истории наблюдателя O', находящегося на поверхности этой сферы, событий происшедших после определенного момента (скажем, 1 часа). *а* — диаграмма Финкельстейна; *б* — диаграмма Пенроуза (вторая жирная точка на рисунке соответствует наблюдателю O').

а затем этот процесс падения яркости будет распространяться к его краю [1]. Характерное время этого спада интенсивности равно по порядку величины времени, за которое свет проходит расстояние $2m$.

Итак, остается объект, который во всех отношениях практически невидим. Однако он будет иметь ту же шварцшильдовскую массу и создавать такое же гравитационное поле, как и до коллапса. Можно было бы обнаружить присутствие этого объекта по его гравитационному действию, например по влиянию на орбиты близлежащих объектов или по отклонению луча света, проходящего вблизи него. Кроме того, возможно, что газ, падающий на такой объект, будет порождать ударные волны, которые будут служить источником рентгеновских лучей или радиоволн.

Наиболее удивительное свойство сферически симметричного коллапса состоит в том, что сингулярность оказывается в области $r < 2m$, откуда свет вообще не может уйти на бесконечность. Поэтому, если наблюдатель останется вне $r = 2m$, то он никогда не увидит сингулярность, предсказываемую теоремой 2. Кроме того, нарушение физических законов, происходящее в сингулярности, не может сказаться на нашей способности предсказывать будущее в асимптотически плоской области пространства-времени.

Спрашивается, справедлива ли эта картина, если коллапс не строго сферически симметричен? В предыдущем разделе мы воспользовались теоремой об устойчивости решения задачи Коши, чтобы показать, что малые отклонения от сферической симметрии не помешают появлению замкнутых ловушечных поверхностей. Однако эта теорема в приведенной форме утверждает лишь, что достаточно малое возмущение в начальных данных приведет к возмущению в решении, которое мало в компактной области. Из этого не следует, что возмущение решения останется малым сколь угодно долго.

Можно ожидать, что в общем случае наличие сингулярностей приведет к горизонтам Коши (как в решениях Райсснера — Нордстрема и Керра), а следовательно, к нарушению возможности предсказывать будущее. Однако если сингулярности не видны извне, возможность предсказания сохраняется во внешней асимптотически плоской области.

Чтобы показать это строго, мы предположим, что в $(\mathscr{M}, \mathbf{g})$ содержится область, асимптотически плоская в смысле слабой асимптотической простоты и пустоты (разд. 6.9). Тогда существует пространство $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$, в которое $(\mathscr{M}, \mathbf{g})$ может быть конформно вложено как многообразие с краем $\bar{\mathscr{M}} = \mathscr{M} \cup \partial\mathscr{M}$, где край $\partial\mathscr{M}$ многообразия $\mathscr{M}$ в $\tilde{\mathscr{M}}$ состоит из двух изотропных

поверхностей $\mathscr{I}^+$ и $\mathscr{I}^-$, которые соответствуют бесконечности в будущем и прошлом. Пусть $\mathscr{S}$ — частичная поверхность Коши в $\mathscr{M}$. Мы будем говорить, что в пространстве $(\mathscr{M}, \mathbf{g})$ *будущее асимптотически предсказуемо* с поверхности $\mathscr{S}$, если $\mathscr{I}^+$ содержится в замыкании множества $D^+(\mathscr{S})$ в конформном многообразии $\tilde{\mathscr{M}}$. Примерами пространств, в которых будущее асимптотически предсказуемо с некоторой поверхности $\mathscr{S}$, являются пространство Минковского, решение Шварцшильда при $m \geqslant 0$, решение Керра при $m \geqslant 0$, $|a| \leqslant m$ и решение Райсснера — Нордстрема при $m \geqslant 0$, $|e| \leqslant m$. Решение Керра с $|a| > m$ и решение Райсснера — Нордстрема с $|e| > m$ не являются пространствами с асимптотически предсказуемым будущим, поскольку в них для любой частичной поверхности Коши $\mathscr{S}$ существуют непродолжимые в прошлое непространственноподобные кривые от поверхности $\mathscr{I}^+$, которые не пересекают $\mathscr{S}$, а стремятся к сингулярности. Асимптотическую предсказуемость будущего можно рассматривать как условие отсутствия в будущем от $\mathscr{S}$ «голых» сингулярностей, т. е. сингулярностей, которые видимы с $\mathscr{I}^+$.

При сферическом коллапсе мы получаем пространство с асимптотически предсказуемым будущим. Вопрос в том, справедливо ли это и для несферического коллапса? Мы не можем дать строгий ответ на этот вопрос. Расчеты по теории возмущений Дорошкевича, Зельдовича, Новикова [43] и Прайса [137] свидетельствуют, по-видимому, о том, что малые отклонения от сферической симметрии не приводят к появлению голой сингулярности. Кроме того, Гиббонс и Пенроуз пытались, хотя и безуспешно, обнаружить противоречия, которые показали бы, что в некоторых ситуациях развитие пространства с асимптотически предсказуемым будущим не имеет смысла. Их неудача, конечно, не доказывает, что асимптотическая предсказуемость имеет место, но говорит в пользу этого. Если она невозможна, то мы не можем сказать ничего определенного об эволюции какой-либо области пространства, содержащего сингулярность, поскольку из сингулярности может поступать новая информация. Поэтому мы будем придерживаться предположения о том, что условие асимптотической предсказуемости будущего выполняется по крайней мере при достаточно малых отклонениях от сферической симметрии.

Можно ожидать, что частица на замкнутой ловушечной поверхности не в состоянии вырваться к поверхности $\mathscr{I}^+$. Однако если допускаются произвольные сингулярности, то можно всегда произвести нужные разрезы и отождествления, чтобы проложить путь для «побега» частицы. Из следующего утверждения видно, что это невозможно в пространстве с асимптотически предсказуемым будущим.

Предложение 9.2.1
Если

а) в пространстве $(\mathscr{M}, g)$ асимптотически предсказуемо будущее с частичной поверхности Коши $\mathscr{S}$,

б) $R_{ab}K^aK^b \geqslant 0$ для всех изотропных векторов K^a, то замкнутая ловушечная поверхность $\mathscr{T}$ в $D^+(\mathscr{S})$ не может пересечь $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, т. е. ее нельзя увидеть с $\mathscr{I}^+$.

В самом деле, допустим, что $\mathscr{T} \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ не пусто. Тогда должна существовать точка $p \in \mathscr{I}^+$ в $J^+(\mathscr{T}, \bar{\mathscr{M}})$. Пусть $\mathscr{U}$ — окрестность $\mathscr{M}$, изометричная окрестности $\mathscr{U}'$ границы $\partial\mathscr{M}'$ асимптотически простого и пустого пространства $(\mathscr{M}', g')$ в конформном многообразии $\bar{\mathscr{M}}'$. Пусть $\mathscr{S}'$ — поверхность Коши в $\mathscr{M}'$, которая совпадает с $\mathscr{S}$ на $\mathscr{U}' \cap \mathscr{M}'$. Тогда множество $\mathscr{S}' - \mathscr{U}'$ компактно и, следовательно, по лемме 6.9.3 каждая образующая поверхности $\mathscr{I}^+$ покидает область $J^+(\mathscr{S}' - \mathscr{U}', \bar{\mathscr{M}}')$. Отсюда видно, что для любого компактного множества $\mathscr{W} \subset \mathscr{S}$ каждая образующая $\mathscr{I}^+$ выходит из $J^+(\mathscr{W}, \bar{\mathscr{M}})$. Из этого следует, что любая образующая $\mathscr{I}^+$ выходит из области $J^+(\mathscr{T}, \bar{\mathscr{M}})$, так как она содержится в $J^+(J^-(\mathscr{T} \cap \mathscr{S}, \bar{\mathscr{M}})$. Поэтому изотропная геодезическая образующая μ поверхности $\dot{J}^+(\mathscr{T}, \bar{\mathscr{M}})$ должна пересечь $\mathscr{I}^+$. Образующая μ должна иметь конечную точку в прошлом на $\mathscr{T}$, так как иначе она пересекла бы $\mathscr{I}^-(\mathscr{S})$. Поскольку μ доходит до $\mathscr{I}^+$, ее аффинная длина бесконечна. Но по условию (б) любая изотропная геодезическая, ортогональная к $\mathscr{T}$, должна содержать на конечном аффинном расстоянии точку, сопряженную $\mathscr{T}$. Следовательно, она не может оставаться в $\dot{J}^+(\mathscr{T}, \bar{\mathscr{M}})$ на всем пути к $\mathscr{I}^+$. Этим доказывается, что $\mathscr{T}$ не может пересечь $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. □

Из всего этого следует, что замкнутая ловушечная поверхность в $D^+(\mathscr{S})$ в пространстве с асимптотически предсказуемым будущим должна лежать в $\mathscr{M} - J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Следовательно, должен существовать нетривиальный горизонт событий (будущего) $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, являющийся границей области, из которой частицы или фотоны могут уйти на бесконечность в направлении будущего. Согласно определению (разд. 6.3), горизонт событий есть ахрональная граница, образованная изотропными геодезическими сегментами, у которых могут быть конечные точки в прошлом, но нет конечных точек в будущем.

Лемма 9.2.2
Если условия (а) и (б) предложения 9.2.1 удовлетворены и если имеется непустой горизонт событий $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, то расхождение $\hat{\theta}$ изотропных геодезических образующих этого

горизонта неотрицательно в области

$$\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap D^+(\mathscr{S}).$$

Допустим, что есть открытое множество $\mathscr{U}$, такое, что $\hat{\theta} < 0$ на $\mathscr{U} \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Пусть $\mathscr{T}$ — пространственноподобная 2-поверхность в $\mathscr{U} \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Тогда $\hat{\theta} = \chi_2{}^a{}_a = 0$. Пусть $\mathscr{V}$ — открытое подмножество множества $\mathscr{U}$, которое пересекает $\mathscr{T}$ и обладает компактным замыканием, содержащимся в $\mathscr{U}$. В $\mathscr{V}$ мы можем немного изменить $\mathscr{T}$ так, чтобы по-прежнему было $\chi_2{}^a{}_a < 0$, но чтобы при этом в $\mathscr{U}$ поверхность $\mathscr{T}$ пересекала $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Как и прежде, это ведет к противоречию, поскольку любая образующая поверхности $\dot{J}^+(\mathscr{T}, \bar{\mathscr{M}})$ в $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ должна обладать конечной точкой в прошлом на $\mathscr{T} \cap \mathscr{V}$, где она должна быть ортогональна к $\mathscr{T}$. Однако ввиду того, что $\chi_2{}^a{}_a < 0$ в $\mathscr{V}$, любая выходящая из $\mathscr{V}$ изотропная ортогональная к $\mathscr{T}$ в $\mathscr{V}$ геодезическая будет содержать на конечном аффинном расстоянии точку, сопряженную $\mathscr{T}$, и потому не может оставаться в $\dot{J}^+(\mathscr{T}, \bar{\mathscr{M}})$ на всем пути к $\mathscr{I}^+$. □

В пространстве с асимптотически предсказуемым будущим $J^+(\mathscr{S}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ содержится в $D^+(\mathscr{S})$. Если на горизонте событий найдется точка $p \in J^+(\mathscr{S})$, которая не принадлежит $D^+(\mathscr{S})$, малейшее возмущение может привести к тому, что p попадет в $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, т. е. будет видна из бесконечности; это означает, что данное пространство не является более асимптотически предсказуемым. Ввиду этого, по-видимому, имеет смысл несколько уточнить определение асимптотически предсказуемого будущего: будем говорить, что в пространстве-времени будущее *асимптотически сильно предсказуемо* с частичной поверхности Коши $\mathscr{S}$, если $\mathscr{I}^+$ содержится в замыкании области $D^+(\mathscr{S})$ в $\bar{\mathscr{M}}$ и $J^+(\mathscr{S}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ содержится в $D^+(\mathscr{S})$. Другими словами, с поверхности $\mathscr{S}$ можно делать предсказания еще и относительно некоторой окрестности горизонта событий.

Предложение 9.2.3

Если $(\mathscr{M}, \mathbf{g})$ — пространство с сильно асимптотически предсказуемым будущим с частичной поверхности Коши $\mathscr{S}$, то существует гомеоморфизм

$$\alpha: (0, \infty) \times \mathscr{S} \to D^+(\mathscr{S}) - \mathscr{S},$$

для которого при каждом $\tau \in (0, \infty)$ поверхность $\mathscr{S}(\tau) \equiv \equiv (\{\tau\} \times \mathscr{S})$ представляет собой частичную поверхность Коши со следующими свойствами:

а) при $\tau_2 > \tau_1 \mathscr{S}(\tau_2) \subset I^+(\mathscr{S}(\tau_1))$;

б) при каждом τ граница поверхности $\mathscr{S}(\tau)$ в конформном многообразии $\tilde{\mathscr{M}}$ является пространственноподобной 2-сферой $\mathscr{L}(\tau)$ в $\mathscr{I}^+$, такой, что при $\tau_2 > \tau_1$ $\mathscr{L}(\tau_2)$ находится строго в будущем сферы $\mathscr{L}(\tau_1)$;

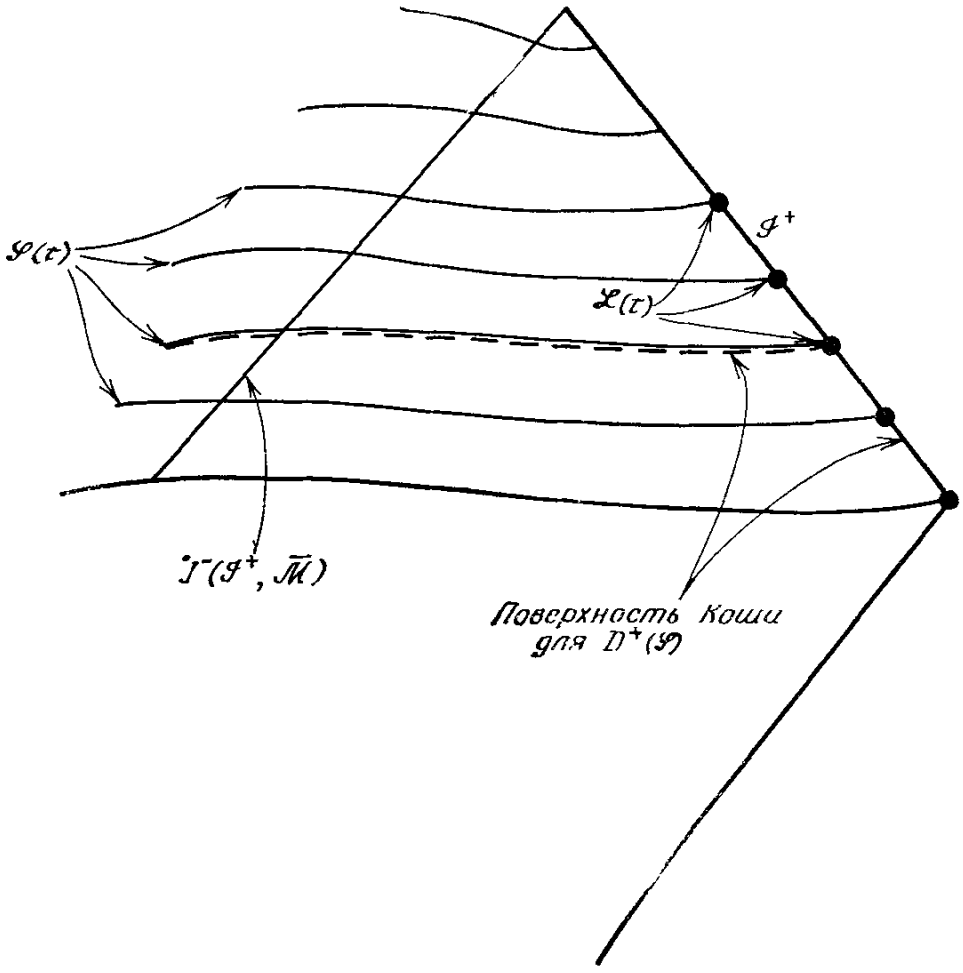


Рис. 58. Пространство $(\mathscr{M}, \mathbf{g})$ с асимптотически сильно предсказуемым с частичной поверхности Коши $\mathscr{S}$ будущим. Изображено семейство $\mathscr{S}(\tau)$ пространственноподобных поверхностей, которые покрывают $D^+(\mathscr{S}) - \mathscr{S}$ и пересекают $\mathscr{I}^+$ по семейству 2-сфер $\mathscr{L}(\tau)$.

в) при каждом τ множество $\mathscr{S}(\tau) \cup (\mathscr{I}^+ \cap J^-(\mathscr{L}(\tau), \tilde{\mathscr{M}}))$ есть поверхностьКоши в $\tilde{\mathscr{M}}$ для $D(\mathscr{S})$.

Другими словами, $\mathscr{S}(\tau)$ есть семейство пространственноподобных поверхностей, гомеоморфных $\mathscr{S}$, которое покрывает $D^+(\mathscr{S}) - \mathscr{S}$ и пересекает $\mathscr{I}^+$ (рис. 58). Эти поверхности можно рассматривать как поверхности постоянного времени в асимптотически предсказуемой области. Выберем их так, чтобы они пересекали $\mathscr{I}^+$; тогда при наличии гравитационного или иного излучения какая-либо масса, расположенная на этих поверхностях и измеренная из бесконечности, будет убывать.

Построение $\mathscr{S}(\tau)$ подобно доказательству предложения 6.4.9. Выберем непрерывное семейство $\mathscr{L}(\tau)$ $(\infty > \tau > 0)$ пространственноподобных 2-сфер, которые покрывают $\mathscr{I}^+$ так, что при $\tau_2 > \tau_1$ $\mathscr{L}(\tau_2)$ находится строго в будущем сферы $\mathscr{L}(\tau_1)$. Введем на $\mathscr{M}$ меру объема так, чтобы полный объем $\mathscr{M}$ по этой мере был конечен. В первую очередь докажем следующую лемму:

Лемма 9.2.4

Величина $k(\tau)$, объем множества $I^-(\mathscr{L}(\tau), \bar{\mathscr{M}}) \cap D^+(\mathscr{S})$, есть непрерывная функция τ.

Пусть $\mathscr{V}$ — какое-либо открытое множество с компактным замыканием в

$$I^-(\mathscr{L}(\tau), \bar{\mathscr{M}}) \cap D^+(\mathscr{S}).$$

Тогда из каждой точки множества $\mathscr{V}$ можно провести времениподобную кривую до $\mathscr{L}(\tau)$, которую можно деформировать так, чтобы получить времениподобные кривые до $\mathscr{L}(\tau - \delta)$ при некотором $\delta > 0$. При данном $\varepsilon > 0$ можно найти такое $\mathscr{V}$, объем которого больше $k(\tau) - \varepsilon$. Следовательно, найдется $\delta > 0$, при котором $k(\tau - \delta) > k(\tau) - \varepsilon$. С другой стороны, допустим, что существует открытое множество $\mathscr{W}$, которое не имеет пересечения с $I^-(\mathscr{L}(\tau), \bar{\mathscr{M}}) \cap D^+(\mathscr{S})$, но содержится в $\mathscr{I}^-(\mathscr{L}(\tau'), \bar{\mathscr{M}}) \cap$ $\cap D^+(\mathscr{S})$ при любом $\tau' > \tau$. Тогда, если $p \in \mathscr{W}$, должны существовать направленные в прошлое времениподобные кривые λ_τ от каждой $\mathscr{L}(\tau')$ до p. Поскольку область $\mathscr{I}^+$ между $\mathscr{L}(\tau)$ и $\mathscr{L}(\tau_1)$ компактна для любых $\tau_1 > \tau$, из $\mathscr{L}(\tau)$ можно будет провести направленную в прошлое непространственноподобную кривую λ, которая будет предельной кривой для $\{\lambda_{\tau'}\}$. Поскольку $\{\lambda_{\tau'}\}$ не пересекают область $I^-(\mathscr{L}(\tau), \bar{\mathscr{M}})$, λ тоже не будет ее пересекать и, следовательно, будет изотропной геодезической, лежащей в $I^-(\mathscr{L}(\tau), \bar{\mathscr{M}})$. Она войдет в $\mathscr{M}$ и потому должна или иметь конечную точку в прошлом в p, или пересечь $\mathscr{S}$. Первое невозможно, так как в этом случае $\mathscr{W}$ имеет пересечение с $I^-(\mathscr{L}(\tau), \bar{\mathscr{M}})$, а второе невозможно, поскольку $p \in \mathscr{I}^+(\mathscr{S})$. Следовательно, не существует такого открытого множества, которое лежало бы в $I^-(\mathscr{L}(\tau'), \bar{\mathscr{M}})$ при каждом $\tau' > \tau$, но не принадлежало бы $I^-(\mathscr{L}(\tau), \bar{\mathscr{M}}) \cap D^+(\mathscr{S})$. Таким образом, для данной постоянной ε найдется такая постоянная δ, что

$$k(\tau + \delta) < k(\tau) + \varepsilon,$$

т. е. $k(\tau)$ — непрерывная функция. □

Доказательство предложения 9.2.3. Введем функции $f(p)$ и $h(p, \tau)$, $p \in D^+(\mathscr{S})$, представляющие собой объем $I^+(p)$ и

$\overline{I^-}(\mathscr{L}(\tau), \bar{\mathscr{M}})$. Так же как в предложении 6.4.9, функция $f(p)$ непрерывна на глобально гиперболической области $D^+(\mathscr{S}) - \mathscr{S}$ и стремится к нулю на каждой непродолжимой в будущее непространственноподобной кривой. Поскольку $\overline{I^-}(\mathscr{L}(\tau), \bar{\mathscr{M}}) \cap \mathscr{M}$ есть множество в прошлом, область

$$D^+(\mathscr{S}) - \overline{I^-}(\mathscr{L}(\tau), \bar{\mathscr{M}}) - \mathscr{S}$$

глобально гиперболична. Следовательно, при каждом значении τ функция $h(p, \tau)$ непрерывна на $D^+(\mathscr{S}) - \mathscr{S}$. Это означает, что при данном $\varepsilon > 0$ можно найти такую окрестность $\mathscr{U}$ точки p, что $|h(q, \tau) - h(p, \tau)| < \varepsilon/2$ для любого $q \in \mathscr{U}$. По лемме 9.2.4 можно найти $\delta > 0$, такое, что $|k(\tau') - k(\tau)| < \varepsilon/2$, когда $|\tau' - \tau| < \delta$. Отсюда $|h(q, \tau') - h(p, \tau)| < \varepsilon$, т. е. функция $h(p, \tau)$ непрерывна на $(D^+(\mathscr{S}) - \mathscr{S}) \times (0, \infty)$. Поверхности $\mathscr{S}(\tau)$ можно определить, тогда как множества точек $p \in D^+(\mathscr{S}) - \mathscr{S}$, для которых $h(p, \tau) = \tau f(p)$. Ясно поэтому, что они представляют собой пространственноподобные поверхности, которые покрывают $D^+(\mathscr{S}) - \mathscr{S}$ и удовлетворяют условиям (а)—(в).

Чтобы задать гомеоморфизм α, нам понадобится времениподобное векторное поле на $D^+(\mathscr{S}) - \mathscr{S}$, пересекающее каждую поверхность $\mathscr{S}(\tau)$. Построим такое поле следующим образом. Пусть $\mathscr{V}$ — некоторая окрестность поверхности $\mathscr{I}^+$ в конформном многообразии $\tilde{\mathscr{M}}$ и $\mathbf{X}_1$ — некоторое непространственноподобное векторное поле на $\mathscr{V}$, которое на $\mathscr{I}^+$ касательно к образующим $\mathscr{I}^+$. Далее, пусть $x_1 \geqslant 0$ есть C^2-функция, равная нулю вне $\mathscr{V}$ и отличная от нуля на $\mathscr{I}^+$, $\mathbf{X}_2$ — времениподобное векторное поле на $\mathscr{M}$ и, наконец, $x_2 \geqslant 0$ — C^2-функция на $\tilde{\mathscr{M}}$, отличная от нуля на $\mathscr{M}$, но равная нулю на $\mathscr{I}^+$. Тогда векторное поле $\mathbf{X} = x_1\mathbf{X}_1 + x_2\mathbf{X}_2$ обладает требуемым свойством. Гомеоморфизм $\alpha: D^+(\mathscr{S}) - \mathscr{S} \to (0, \infty) \times \mathscr{S}$ отображает точку $p \in D^+(\mathscr{S}) - \mathscr{S}$ в пару (τ, q), где τ определяется тем, что $p \in \mathscr{L}(\tau)$, а интегральная кривая поля $\mathbf{X}$, проходящая через p, пересекает $\mathscr{S}$ в точке q. □

Если в области $D^+(\mathscr{S})$ пространства с асимптотически предсказуемым будущим существует горизонт событий $\dot{J}(\mathscr{I}^+, \bar{\mathscr{M}})$, то из свойства (в) предложения 9.2.3 следует, что при достаточно больших τ поверхности $\mathscr{S}(\tau)$ пересекут этот горизонт. Определим *черную дыру* на поверхности $\mathscr{S}(\tau)$ как связную компоненту множества $\mathscr{B}(\tau) \equiv \mathscr{S}(\tau) - J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Другими словами, черная дыра — это область поверхности $\mathscr{S}(\tau)$, из которой ни частицы, ни фотоны не могут уйти на поверхность $\mathscr{I}^+$.

По мере роста τ черные дыры могут сливаться, и в результате такого вторичного коллапса могут образоваться новые черные дыры. Вместе с тем из следующего утверждения видно, что черные дыры никогда не могут распасться.

Предложение 9.2.5

Пусть $\mathscr{B}_1(\tau_1)$ — черная дыра на $\mathscr{S}(\tau_1)$. Пусть $\mathscr{B}_2(\tau_2)$ и $\mathscr{B}_3(\tau_2)$ — черные дыры на более поздней поверхности $\mathscr{S}(\tau_2)$. Если $\mathscr{B}_2(\tau_2)$ и $\mathscr{B}_3(\tau_2)$ пересекают $J^+(\mathscr{B}_1(\tau_1))$, то $\mathscr{B}_2(\tau_2) = \mathscr{B}_3(\tau_2)$.

По свойству (в) предложения 9.2.3 каждая направленная в будущее непродолжимая времениподобная кривая из $\mathscr{B}_1(\tau_1)$ пересечет $\mathscr{S}(\tau_2)$. Следовательно, область

$$J^+(\mathscr{B}_1(\tau_1)) \cap \mathscr{S}(\tau_2)$$

связная и будет содержаться в связной компоненте $\mathscr{B}(\tau_2)$. □

Для физических приложений в первую очередь интересны черные дыры, образовавшиеся в результате гравитационного коллапса из первоначально несингулярного состояния. Уточним это понятие. Будем говорить, что частичная поверхность Коши $\mathscr{S}$ обладает *асимптотически простым прошлым*, если область $J^-(\mathscr{S})$ изометрична области $J^-(\mathscr{S}')$ некоторого асимптотически простого и пустого пространства-времени $(\mathscr{M}', \mathbf{g}')$, где $\mathscr{S}'$ — поверхность Коши для $(\mathscr{M}', \mathbf{g}')$. Согласно предложению 6.9.4, поверхность $\mathscr{S}'$ обладает топологией R^3, следовательно, $\mathscr{S}$ имеет такую же топологию. Тогда из предложения 9.2.3 следует, что каждая поверхность $\mathscr{S}(\tau)$ обладает топологией R^3 и объединение $\mathscr{S}(\tau)$ с граничной 2-сферой $\mathscr{L}(\tau)$ на $\mathscr{I}^+$ гомеоморфно единичному кубу I^3, если $(\mathscr{M}, \mathbf{g})$ — пространство со строго асимптотически предсказуемым с поверхности $\mathscr{S}$ будущим и с асимптотически простым прошлым.

Хотя мы интересуемся главным образом пространствами, имеющими асимптотически простые прошлые, будет уместно рассмотреть в следующем разделе пространства с асимптотически предсказуемым будущим, которые не имеют асимптотически простого прошлого, но при больших временах служат хорошим приближением к пространствам, имеющим его. Примером этого является сферически симметричный коллапс, рассмотренный нами в начале раздела. Если поверхность звезды ушла за горизонт событий, метрикой внешней области является метрика решения Шварцшильда, и на ней не сказывается дальнейшая судьба звезды. Поэтому при исследовании асимптотического поведения уместно попросту забыть о звезде и рассматривать решение Шварцшильда в пустом пространстве как пространство, в котором будущее сильно асимптотически предсказуемо с поверхности $\mathscr{S}$ типа изображенной на рис. 24. Эта поверхность не обладает асимптотически простым прошлым и топологически она представляет собой $S^2 \times R^1$, а не R^3. Однако часть $\mathscr{S}$ вне горизонта событий в области I имеет такую же топологию, как область поверхности $\mathscr{S}(\tau)$ вне горизонта событий на рис. 57. Мы хотим рассматривать пространства, в кото-

рых будущее сильно асимптотически предсказуемо с поверхности $\mathscr{S}$ и которые при этом таковы, что часть $\mathscr{S}$, находящаяся вне горизонта событий, обладает такой же топологией, как некоторая поверхность $\mathscr{S}(\tau)$ в пространстве с асимптотически простым прошлым. Конечно, в более сложных случаях может оказаться несколько компонент $\mathscr{B}(\tau)$, соответствующих коллапсу нескольких тел. Поэтому мы будем рассматривать пространства, которые сильно асимптотически предсказуемы с поверхности $\mathscr{S}$ и обладают тем свойством, что

(α) множество $\mathscr{S} \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ гомеоморфно пространству R^3 с вырезанным из него открытым множеством с компактным замыканием.

(Отметим, что это открытое множество может и не быть связным.) Кроме того, было бы удобно наложить такое требование:

(β) поверхность $\mathscr{S}$ односвязна.

Предложение 9.2.6

Пусть $(\mathscr{M}, \mathbf{g})$ — пространство, которое сильно асимптотически предсказуемо с частичной поверхности Коши $\mathscr{S}$, удовлетворяющей требованиям (α), (β). Тогда

1) поверхность $\mathscr{S}(\tau)$ также удовлетворяет требованиям (α), (β);

2) при любом τ $\partial\mathscr{B}_1(\tau)$ — граница черной дыры $\mathscr{B}_1(\tau)$ в $\mathscr{S}$ — компактна и односвязна.

Поскольку поверхности $\mathscr{S}(\tau)$ гомеоморфны $\mathscr{S}$, они удовлетворяют требованию (β). Можно определить инъективное отображение

$$\gamma: \mathscr{S}(\tau) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}) \to \mathscr{S} \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$$

как перенесение каждой точки $\mathscr{S}(\tau)$ по интегральной кривой векторного поля $\mathbf{X}$, введенного в предложении 9.2.3, до ее пересечения с $\mathscr{S}$. Пространство $(\mathscr{M}, \mathbf{g})$ — асимптотически простое в слабом смысле, поэтому в $\mathscr{S}(\tau) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ вблизи $\mathscr{I}^+$ можно построить 2-сферу $\mathscr{P}$. Часть $\mathscr{S}(\tau)$, которая находится вне $\mathscr{P}$, будет отображаться в область $\mathscr{S}$, расположенную вне 2-сферы $\gamma(\mathscr{P})$. Отсюда видно, что область множества $\mathscr{S} \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$, которая не содержится в $\gamma(\mathscr{S}(\tau) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}))$, должна иметь компактное замыкание. Следовательно, множество $\gamma(\mathscr{S}(\tau) \cap \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}))$ должно быть гомеоморфным пространству R^3, из которого вырезано открытое множество с компактным замыканием. Поскольку поверхность $\mathscr{S}(\tau)$ гомеоморфна $R^3 - \mathscr{V}$, где $\mathscr{V}$ — открытое подмножество пространства R^3 с компактным замыканием, граница $\partial\mathscr{B}(\tau)$ будет гомеоморфна $\partial\mathscr{V}$ и в силу этого

компактна. Граница $\partial\mathscr{B}_1(\tau)$, будучи замкнутым подмножеством $\partial\mathscr{B}(\tau)$, также компактна.

Допустим, что $\partial\mathscr{B}_1(\tau)$ состоит из двух несвязных компонент $\partial\mathscr{B}^1{}_1(\tau)$ и $\partial\mathscr{B}_1{}^2(\tau)$. Тогда в $\mathscr{S}(\tau)-\mathscr{B}(\tau)$ можно было бы провести кривые λ_1 и λ_2 от $\mathscr{L}(r)$ до $\partial\mathscr{B}_1{}^1(\tau)$ и $\partial\mathscr{B}_1{}^2(\tau)$ соответственно. Можно было бы также в $\operatorname{int}\mathscr{B}_1(\tau)$ провести кривую μ от $\partial\mathscr{B}_1{}^1(\tau)$ до $\partial\mathscr{B}_1{}^2(\tau)$. Соединив все эти кривые, мы получили бы замкнутую кривую в $\mathscr{S}(\tau)$, которая имела бы с $\partial\mathscr{B}_1{}^1(\tau)$ только одно пересечение. Эту кривую нельзя деформировать в $\mathscr{S}(\tau)$ к нулю, что противоречит тому факту, что $\mathscr{S}(\tau)$ односвязна. □

Нас интересуют лишь такие черные дыры, в которые реально можно «провалиться», т. е. такие, граница которых $\partial\mathscr{B}(\tau)$ содержится в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Поэтому к требованиям (α) и (β) мы добавим такое:

(γ) для достаточно больших τ $\mathscr{S}(\tau)\cap\overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ содержится в $\overline{J^+}(\mathscr{I}^-, \bar{\mathscr{M}})$.

Будем называть $(\mathscr{M}, \mathbf{g})$ *регулярно предсказуемым пространством*, если его будущее сильно предсказуемо с частичной поверхности Коши $\mathscr{S}$ и если оно удовлетворяет требованиям (α)—(γ). Все пространства, упомянутые в начале раздела, имея асимптотически предсказуемое будущее, являются в то же время и регулярно предсказуемыми. Из предложения 9.2.6 видно, что когда мы имеем дело с регулярно предсказуемыми пространствами, развивающимися из некоторой частичной поверхности Коши, существует взаимно-однозначное соответствие между черными дырами $\mathscr{B}_i(\tau)$ и их границами $\partial\mathscr{B}_i(\tau)$ в $\mathscr{S}(\tau)$. Поэтому в такой ситуации можно было бы дать эквивалентное определение черной дыры как связной компоненты множества $\mathscr{S}(\tau)\cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$.

Приведем утверждение, в котором устанавливается важное для следующего раздела свойство границы черной дыры.

Предложение 9.2.7

Пусть $(\mathscr{M}, \mathbf{g})$ — регулярно предсказуемое пространство, развивающееся из некоторой частичной поверхности Коши $\mathscr{S}$, причем $R_{ab}K^aK^b \geqslant 0$ для любого изотропного вектора K^a. Пусть $\mathscr{B}_1(\tau)$ — черная дыра на поверхности $\mathscr{S}(\tau)$, а $\{\mathscr{B}_i(\tau')\}$ $(i = 1, \ldots, N)$ — черные дыры на более ранней поверхности $\mathscr{S}(\tau')$, для которых $J^+(\mathscr{B}_i(\tau'))\cap\mathscr{B}_1(\tau)\neq\varnothing$. Тогда площадь $A_1(\tau)$ границы $\partial\mathscr{B}_1(\tau)$ больше или равна сумме площадей $A_i(\tau')$ границ $\partial\mathscr{B}_i(\tau')$, и равенство имеет место лишь при $N = 1$.

Другими словами, площадь границы черной дыры не может убывать со временем, и если две или более черные дыры

сливаются в одну, площадь границы последней будет больше суммы площадей границ исходных черных дыр.

Поскольку горизонт событий есть граница прошлого поверхности $\mathscr{I}^+$, его изотропные геодезические образующие будут обладать конечными точками в будущем, только если они пересекают $\mathscr{I}^+$. Это, однако, невозможно, так как изотропные геодезические образующие $\mathscr{I}^+$ не имеют конечных точек в будущем. Таким образом, изотропные образующие горизонта событий не имеют конечных точек в будущем. По лемме 9.2.2 их расхождение $\hat{\theta}$ неотрицательно; следовательно, площадь двумерных сечений этих образующих не может убывать по мере роста τ. В силу свойства (в) предложения 9.2.3 и согласно предложению 9.2.5, все изотропные геодезические образующие поверхности $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, пересекающие $\mathscr{S}(\tau')$ в одном из $\partial\mathscr{B}_i(\tau')$, должны пересечь $\mathscr{S}(\tau)$ в $\partial\mathscr{B}_1(\tau)$. Таким образом, площадь $\partial\mathscr{B}_1(\tau)$ больше или равна сумме площадей $\{\partial\mathscr{B}_i(\tau')\}$. Когда $N > 1$, $\partial\mathscr{B}_1(\tau)$ будет содержать N раздельных замкнутых подмножеств, которые соответствуют образующим поверхности $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, пересекающим каждую $\partial\mathscr{B}_i(\tau')$. Поскольку $\partial\mathscr{B}_1(\tau)$ связна, она должна содержать открытое множество образующих, которые не пересекают какую-либо из $\partial\mathscr{B}_i(\tau')$, а имеют прошлые конечные точки между $\mathscr{S}(\tau)$ и $\mathscr{S}(\tau')$. □

Определение черной дыры через горизонт событий $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ удобно тем, что он является изотропной гиперповерхностью и обладает потому рядом привлекательных свойств. Однако это определение зависит от поведения решения во все времена в будущем; при заданной частичной поверхности Коши $\mathscr{S}(\tau)$ мы не можем сказать, где расположен горизонт, не решив сначала задачи Коши во всей области Коши будущего этой поверхности. Поэтому будет полезно дать определение горизонта иного рода, который зависит лишь от свойств пространства-времени на поверхности $\mathscr{S}(\tau)$.

Из предложения 9.2.1 мы знаем, что в развивающемся из частичной поверхности Коши $\mathscr{S}$ регулярно предсказуемом пространстве любая замкнутая ловушечная поверхность на $\mathscr{S}(\tau)$ должна находиться в $\mathscr{B}(\tau)$. Этот результат связан только с тем фактом, что входящие ортогонально к этой 2-поверхности изотропные геодезические сходятся. При этом не важно, сходятся или нет входящие изотропные геодезические. Поэтому мы будем говорить, что ориентируемая компактная пространственная 2-поверхность в $D^+(\mathscr{S})$ есть *внешняя ловушечная поверхность*, если расхождение $\hat{\theta}$ выходящих ортогонально к ней изотропных геодезических неотрицательно. (Случай $\hat{\theta} = 0$ включается сюда для удобства.) Чтобы определить понятие семейства выходящих

изотропных геодезических, воспользуемся свойством (β) частичных поверхностей Коши $\mathscr{S}(\tau)$. Пусть $\mathbf{X}$ — времениподобное векторное поле из предложения 9.2.3. Тогда любую компактную ориентируемую 2-поверхность $\mathscr{P}$ в $D^+(\mathscr{S})$ можно отобразить посредством интегральных кривых поля $\mathbf{X}$ на компактную ориентируемую 2-поверхность $\mathscr{P}'$ в $\mathscr{S}(\tau)$ при любом данном значении τ. Пусть λ — кривая в $\mathscr{S}(\tau) \cup \mathscr{L}(\tau)$ от $\mathscr{L}(\tau)$ до $\mathscr{P}'$, пересекающая $\mathscr{P}'$ только в своей конечной точке. Тогда можно определить внешнее направление на $\mathscr{P}'$ в $\mathscr{S}(\tau)$ как направление, в котором λ приближается к $\mathscr{P}'$. Поскольку поверхность $\mathscr{S}(\tau)$ односвязна, это определение однозначно. Семейством выходящих изотропных геодезических, ортогональных $\mathscr{P}$, является тогда семейство, отображаемое полем $\mathbf{X}$ на кривые в $\mathscr{S}(\tau)$, которые для $\mathscr{P}'$ являются выходящими.

Зная решение на поверхности $\mathscr{S}(\tau)$, можно найти все внешние ловушечные поверхности $\mathscr{P}$, которые лежат в $\mathscr{S}(\tau)$. Мы определим *ловушечную область* $\mathscr{T}(\tau)$ на поверхности $\mathscr{S}(\tau)$ как множество всех точек $q \in \mathscr{S}(\tau)$, через которые проходит внешняя ловушечная поверхность $\mathscr{P}$, лежащая в $\mathscr{S}(\tau)$. Как показывает следующий результат, существование ловушечной области $\mathscr{T}(\tau)$ приводит к существованию черной дыры $\mathscr{B}(\tau)$, и при каждом значении τ $\mathscr{T}(\tau)$ лежит в $\mathscr{B}(\tau)$.

Предложение 9.2.8

Пусть $(\mathscr{M}, \mathbf{g})$ — регулярно предсказуемое пространство, развивающееся из частичной поверхности Коши $\mathscr{S}$, и пусть в этом пространстве $R_{ab}K^aK^b \geqslant 0$ для всякого изотропного вектора K^a. Тогда внешняя ловушечная поверхность $\mathscr{P}$ в $D^+(\mathscr{S})$ не пересекает $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$.

Это предложение доказывается подобно предложению 9.2.1. Допустим, что $\mathscr{P}$ пересекает $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Тогда $\dot{J}^+(\mathscr{P}, \bar{\mathscr{M}})$ должна пересекать $\mathscr{I}^+$. К каждой точке из $\mathscr{I}^+ \cap \dot{J}^+(\mathscr{P}, \bar{\mathscr{M}})$ можно провести направленную в прошлое изотропную геодезическую образующую поверхности $\dot{J}^+(\mathscr{P}, \bar{\mathscr{M}})$, имеющую конечную точку в прошлом на $\mathscr{P}$, но не имеющую точек, сопряженных $\mathscr{P}$. В силу (4.35), расхождение $\hat{\theta}$ этих образующих будет неположительно, поскольку оно неположительно на $\mathscr{P}$ и поскольку $R_{ab}K^aK^b \geqslant 0$. Таким образом, площадь двумерного сечения упомянутых образующих будет всегда меньше или равна площади $\mathscr{P}$. Это приводит к противоречию, так как $\mathscr{I}^+ \cap \dot{J}^+(\mathscr{P}, \bar{\mathscr{M}})$, находясь на бесконечности, имеет бесконечную площадь. □

Внешнюю границу $\partial\mathscr{T}_1(\tau)$ связной компоненты $\mathscr{T}_1(\tau)$ ловушечной области $\mathscr{T}(\tau)$ мы будем называть *кажущимся горизонтом*. Согласно предыдущему результату, из существования ка-

жущегося горизонта $\partial\mathcal{T}_1(\tau)$ следует, что существует вне его или совпадает с ним компонента $\partial\mathcal{B}_1(\tau)$ горизонта событий. Обратное, однако, не обязательно верно: внутри горизонта событий может и не быть внешних ловушечных поверхностей.

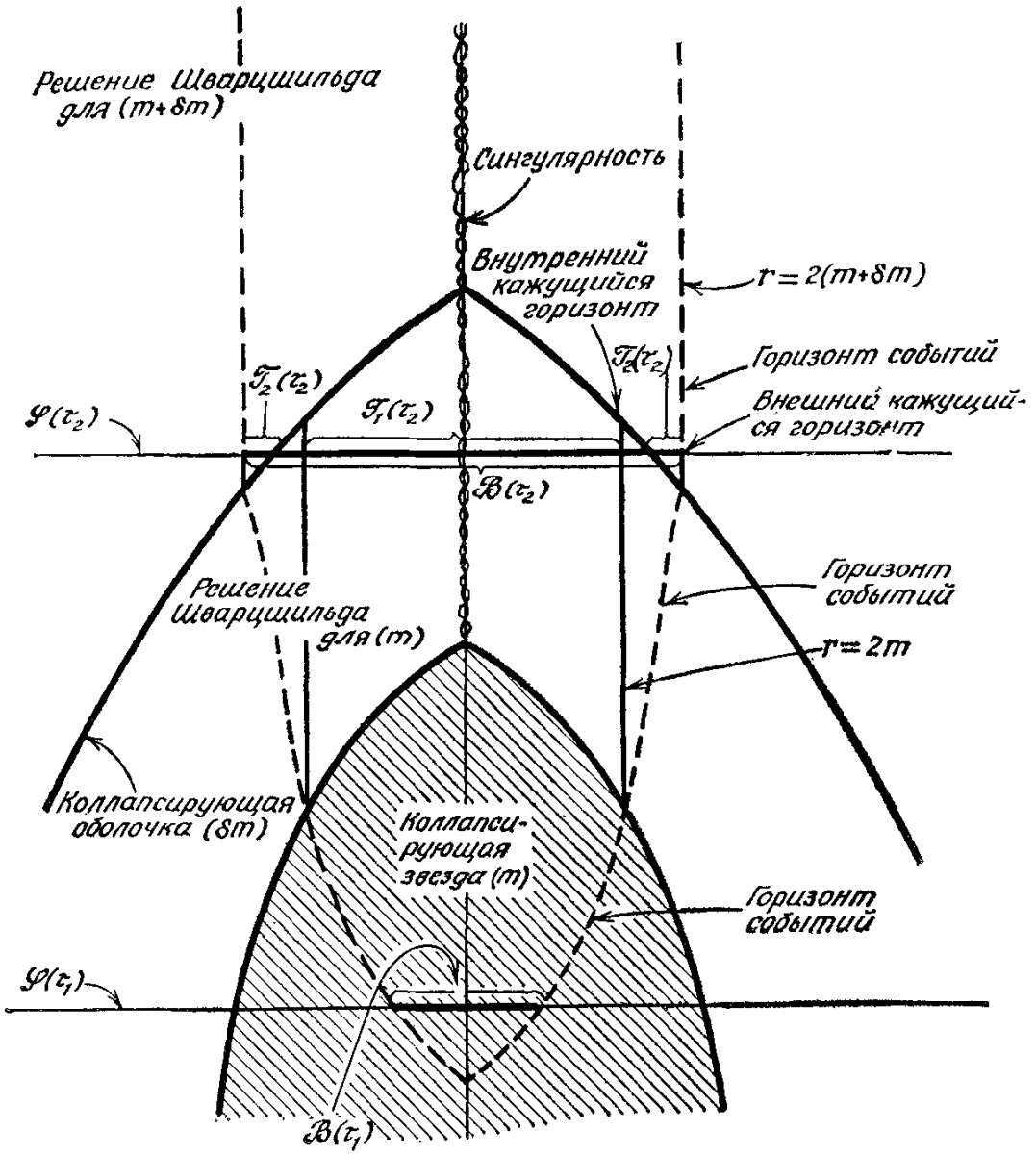


Рис. 59. Сферический коллапс звезды массы m, сопровождаемый сферическим коллапсом оболочки массы δm; внешнее решение будет решением Шварцшильда для массы m после коллапса звезды и решением Шварцшильда для массы $m = \delta m$ после коллапса оболочки. В момент τ_1 имеется горизонт событий, но нет кажущихся горизонтов событий; в момент τ_2 внутри горизонта событий есть два кажущихся горизонта событий.

С другой стороны, внутри одной компоненты $\partial\mathcal{B}_1(\tau)$ горизонта событий может быть более одной связной компоненты ловушечной области $\mathcal{T}(\tau)$. Такая возможность иллюстрируется рис. 59. Похожая ситуация возникает при рассмотрении столкновения и слияния двух черных дыр. На начальной поверхности

$\mathscr{S}(\tau_1)$ имеются две ловушечные области $\mathscr{T}_1(\tau_1)$ и $\mathscr{T}_2(\tau_1)$, находящиеся соответственно в черных дырах $\mathscr{B}_1(\tau_1)$ и $\mathscr{B}_2(\tau_1)$.

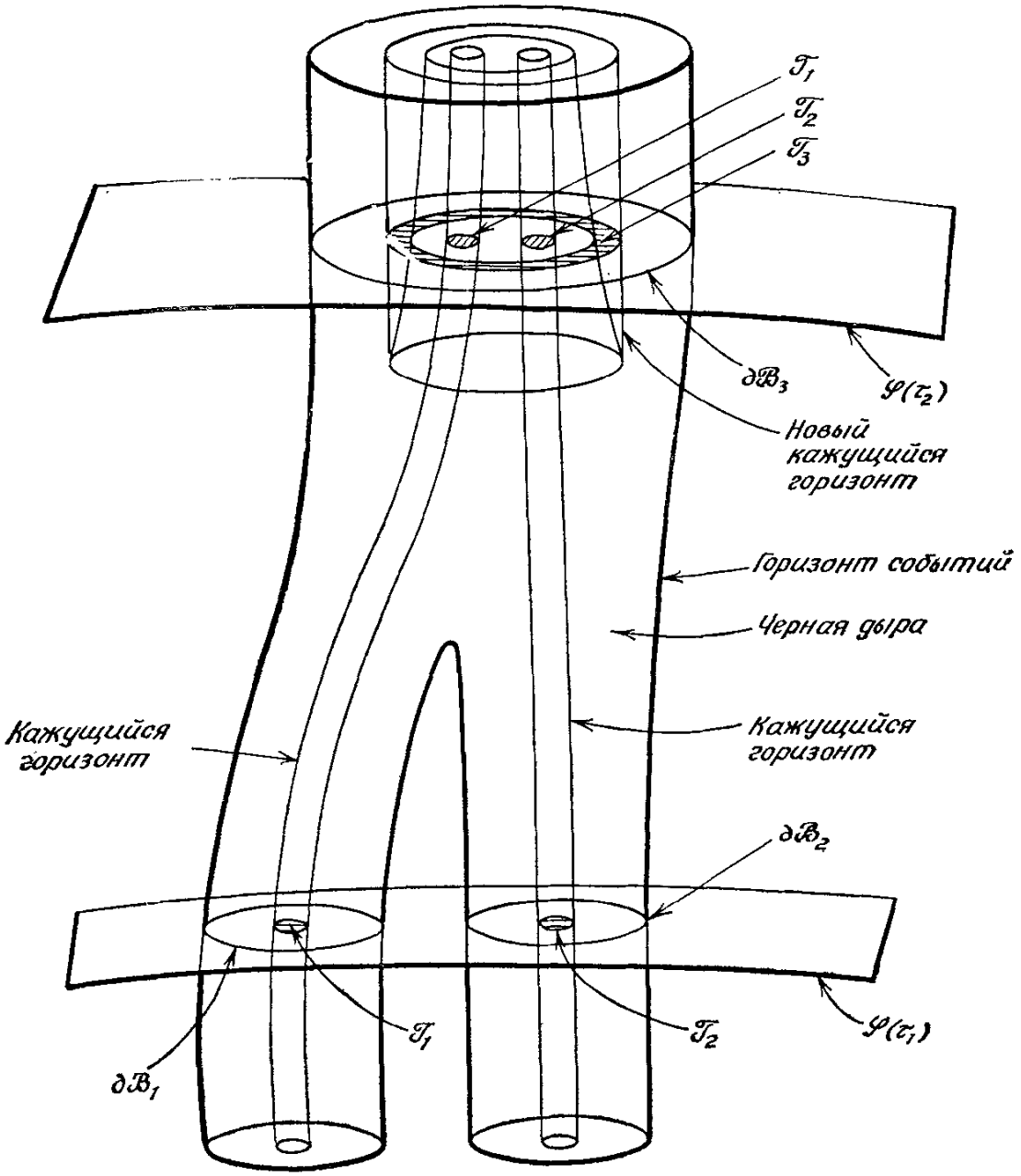


Рис. 60. Столкновение и слияние двух черных дыр. В момент времени τ_1 внутри горизонтов событий $\partial\mathscr{B}(\tau_1)$, $\partial\mathscr{B}(\tau_2)$ существуют кажущиеся горизонты $\partial\mathscr{T}_1$, $\partial\mathscr{T}_2$. К моменту τ эти горизонты событий сливаются и образуют один горизонт событий; к этому моменту образуется третий кажущийся горизонт, окружающий предшествующие кажущиеся горизонты.

По мере их взаимного сближения они сливаются и образуют единую черную дыру $\mathscr{B}_3(\tau_2)$ на более поздней поверхности $\mathscr{S}(\tau_2)$. Однако кажущиеся горизонты $\partial\mathscr{T}_1(\tau)$ и $\partial\mathscr{T}_2(\tau)$ не сольются немедленно, а появится третья ловушечная область $\mathscr{T}_3(\tau)$, окружающая первые две (рис. 60). В некоторый более поздний момент $\mathscr{T}_1$, $\mathscr{T}_2$ и $\mathscr{T}_3$ сольются воедино.

Мы лишь набросаем схему доказательства основных свойств кажущегося горизонта. Прежде всего имеем

Предложение 9.2.9

Каждая компонента $\partial\mathcal{T}(\tau)$ является такой 2-поверхностью, что выходящие ортогональные изотропные геодезические имеют на $\partial\mathcal{T}(\tau)$ нулевое схождение $\hat{\theta}$. (Такую поверхность мы будем называть *маргинальной внешней ловушечной поверхностью.*)

Если бы $\hat{\theta}$ было положительным в некоторой окрестности в $\partial\mathcal{T}(\tau)$ точки $p \in \partial\mathcal{T}(\tau)$, то была бы окрестность $\mathcal{U}$ точки p, такая, что любая внешняя ловушечная поверхность в $\mathcal{S}(\tau)$, которая пересекает $\mathcal{U}$, пересекла бы также и $\partial\mathcal{T}(\tau)$. Следовательно, $\hat{\theta} \leqslant 0$ на $\partial\mathcal{T}(\tau)$.

Если бы $\hat{\theta}$ было отрицательно в некоторой окрестности в $\partial\mathcal{T}(\tau)$ точки $p \in \partial\mathcal{T}(\tau)$, то можно было бы растянуть $\partial\mathcal{T}(\tau)$ в $\mathcal{S}(\tau)$ так, чтобы получить внешнюю ловушечную поверхность снаружи $\partial\mathcal{T}(\tau)$. □

Итак, изотропные геодезические, ортогональные к кажущемуся горизонту $\partial\mathcal{T}(\tau)$ на поверхности $\mathcal{S}(\tau)$, выходят из $\partial\mathcal{T}(\tau)$ с нулевым схождением. Однако если на их пути встретится материя или область, где тензор Вейля удовлетворяет типовому условию (разд. 4.4), они начнут сходиться, поэтому их пересечение с более поздней поверхностью $\mathcal{S}(\tau')$ будет лежать внутри кажущегося горизонта $\partial\mathcal{T}(\tau')$. Иначе говоря, кажущийся горизонт движется наружу по меньшей мере со скоростью света и быстрее света, если на его пути попадется вещество или излучение. Как видно из вышеприведенного примера, это движение может быть еще и разрывным. Вследствие этого с кажущимся горизонтом иметь дело труднее, чем с горизонтом событий, который всегда движется непрерывным образом. Мы покажем в следующем разделе, что горизонт событий и кажущийся горизонт совпадают, когда решение стационарно. Поэтому надо ожидать, что оба горизонта очень близки друг к другу, если решение почти стационарно в течение длительного времени. В частности, их площади в таких случаях, по всей вероятности, одинаковы. Если имеется решение, которое из начального почти стационарного состояния через некоторый период нестационарности переходит к конечному почти стационарному состоянию, то можно использовать предложение 9.2.7, чтобы связать площади начального и конечного горизонтов.

9.3. Конечное состояние черной дыры

В предыдущем разделе мы принимали, что в удалении от коллапсирующей звезды можно предсказать будущее. Как было показано, из этого предположения следует, что звезда уходит

за горизонт событий, который скрывает сингулярность от внешнего наблюдателя. Материя и энергия, пересекающие горизонт событий, будут навсегда потеряны для внешнего мира. Поэтому следует ожидать, что на бесконечность в виде гравитационных волн будет излучено ограниченное количество энергии. Когда подавляющая часть энергии будет излучена, вне горизонта решение, по-видимому, будет приближаться к стационарному состоянию. Ввиду этого в данном разделе мы исследуем строго стационарные решения для черных дыр в предположении, что их внешние области будут довольно точно соответствовать конечным состояниям решений вне сколлапсировавших объектов.

Точнее, мы будем рассматривать пространство $(\mathscr{M}, \mathbf{g})$, которое удовлетворяет следующим условиям:

1) $(\mathscr{M}, \mathbf{g})$ — регулярно предсказуемое пространство, развивающееся из частичной поверхности Коши $\mathscr{S}$;

2) существует группа изометрии $\theta_t: \mathscr{M} \to \mathscr{M}$, у которой вектор Киллинга $\mathbf{K}$ времениподобен вблизи $\mathscr{I}^+$ и $\mathscr{I}^-$;

3) $(\mathscr{M}, \mathbf{g})$ — пустое пространство или содержит поля типа электромагнитного или скалярного, которые удовлетворяют хорошо определенным гиперболическим уравнениям и условию энергодоминантности: $T_{ab}N^aL^b \geqslant 0$ для направленных в будущее времениподобных векторов $\mathbf{N}$, $\mathbf{L}$.

Будем называть пространство, удовлетворяющее этим условиям, *стационарным регулярно предсказуемым пространством.* Мы считаем, что при больших значениях τ область $J^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap \cap J^+(\mathscr{L}(\tau))$ регулярно предсказуемого пространства, содержащая коллапсирующие звезды, будет почти изометрична аналогичной области некоторого стационарного регулярно предсказуемого пространства.

Условие (3) основано на том предположении, что любая материя с ненулевой массой покоя в конечном счете «провалится» за горизонт. Могут остаться лишь поля дальнодействующих сил типа электромагнитных. Из условий (2) и (3) следует, что пространство $(\mathscr{M}, \mathbf{g})$ аналитично вблизи бесконечности, где поле векторов Киллинга времениподобно [110]. Мы примем, что решение в других местах является аналитическим продолжением из этой наружной области. Рассматриваемые здесь стационарные решения не обладают асимптотически простыми прошлыми, так как они соответствуют лишь конечному состоянию системы, и не описывают более раннюю динамическую стадию. Но нас интересуют свойства этих решений только в будущем, а не в прошлом. Свойства в будущем и прошлом могут различаться, поскольку нет никакой априорной причины для того, чтобы они были обратимы во времени; фактически такая обратимость возникает как следствие результатов, которые будут получены ниже.

В стационарном регулярно предсказуемом пространстве площадь двумерных сечений горизонта событий не будет зависеть от времени. Это приводит к следующему фундаментальному результату:

Предложение 9.3.1

Пусть $(\mathscr{M}, \mathbf{g})$ — стационарное регулярно предсказуемое пространство-время. Тогда образующие горизонта событий будущего $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ не имеют конечных точек в прошлом в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Пусть $Y_1{}^a$ — направленные в будущее касательные к этим образующим векторы; тогда в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ вектор $Y_1{}^a$ обладает нулевыми сдвигом $\hat{\sigma}$ и расхождением $\hat{\theta}$ и удовлетворяет равенству

$$R_{ab}Y_1{}^aY_1{}^b = 0 = Y_{1[e}C_{a]bc[d}Y_{1f]}Y_1{}^bY_1{}^c.$$

Чтобы не прерывать изложения, мы отложим доказательство этого и других результатов до конца раздела. Из этого предложения видно, что в стационарном пространстве-времени кажущийся горизонт совпадает с горизонтом событий.

Теперь мы приведем некоторые результаты, которые указывают на то, что семейством решений Керра (разд. 5.6), вероятно, исчерпываются пустые стационарные регулярно предсказуемые пространства. Мы не будем приводить здесь доказательств теорем Израэля и Картера, а только сошлемся на соответствующие статьи. Остальные результаты будут доказаны в конце данного раздела. На основании этих результатов мы полагаем, что для незаряженного коллапсирующего объекта в конце концов устанавливается решение Керра. Если сколлапсировавшее тело обладает ненулевым полным зарядом, решение, как мы полагаем, будет стремиться к одному из решений Керра для заряженного тела.

Предложение 9.3.2

В стационарном регулярно предсказуемом пространстве каждая связная компонента горизонта $\partial\mathscr{B}(\tau)$ в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ гомеоморфна 2-сфере.

Не исключено, что могут существовать несколько связных компонент $\partial\mathscr{B}(\tau)$, соответствующих нескольким черным дырам, расположенным на неизменных расстояниях друг от друга. Такая ситуация может возникнуть в предельном случае, когда каждая из этих черных дыр имеет заряд e, равный ее массе m, и не вращается [69]. По всей вероятности, это единственный случай, в котором могут возникнуть настолько значительные сильные отталкивания, что они уравновесят гравитационное притяжение между черными дырами. Поэтому мы будем

рассматривать решения, в которых $\partial\mathscr{B}(\tau)$ обладает только одной связной компонентой.

Предложение 9.3.3

Пусть $(\mathscr{M}, g)$ — стационарное регулярно предсказуемое пространство. Тогда вектор Киллинга K^a отличен от нуля в односвязной области $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Пусть дано τ_0, при котором $\mathscr{S}(\tau_0) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ содержится в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Если $\partial\mathscr{B}(\tau_0)$ имеет только односвязную компоненту, то область $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap \mathscr{M}$ гомеоморфна $[0, 1] \times S^2 \times R^1$.

При дальнейшем изложении мы будем следовать одному из двух возможных путей в зависимости от того, равен или не равен нулю всюду ротор вектора Киллинга K_a, т. е. $K_{a;b}K_c\eta^{abcd}$. Если он равен нулю, то мы будем называть решение *статическим регулярно предсказуемым пространством-временем.* Грубо говоря, если черная дыра в некотором смысле не вращается, то можно ожидать, что решение будет статическим.

Предложение 9.3.4.

В статическом регулярно предсказуемом пространстве-времени вектор Киллинга **K** времениподобен во внешней области $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, а на

$$\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$$

отличен от нуля и направлен вдоль изотропных образующих поверхности $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$.

Поскольку ротор вектора **K** равен нулю, он ортогонален к некоторой гиперповерхности, т. е. существует некоторая функция ξ, такая, что вектор K_a пропорционален $\xi_{;a}$. Тогда мы можем представить метрику во внешней области в виде $g_{ab} = f^{-1}K_aK_b + h_{ab}$, где $f \equiv K^aK_a$ и h_{ab} — индуцированная метрика на поверхностях $\{\xi = \text{const}\}$, причем она характеризует разделение интегральных кривых поля K^a. Внешняя область, таким образом, допускает изометрию, которая переводит точку на поверхности ξ в точку на поверхности $-\xi$, лежащую на той же интегральной кривой векторного поля **K**. Эта изометрия обращает направление течения времени; пространство, допускающее такого рода изометрию, будем называть *симметричным во времени.* Таким образом, если аналитическое расширение внешней области содержит горизонт событий будущего $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, то оно будет содержать и горизонт событий прошлого $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Эти горизонты событий могут как пересекаться, так и не пересекаться; решения Шварцшильда и Райсснера — Нордстрема при $e^2 < m^2$ служат примерами, когда эти горизонты пересе-

каются, а решение Райсснера — Нордстрема при $e^2 = m^2$ — примером, когда они не пересекаются. В последнем случае $f_{;a} = 0$ на горизонте. Важность этого положения вытекает из того факта, что на горизонте будущего $I^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap I^+(\mathscr{I}^-, \bar{\mathscr{M}})$ выполняется условие $K_{a;b}K^b = {}^1/_2 f_{;a} = \beta K_a$, где $\beta \geqslant 0$ — постоянная вдоль каждой изотропной геодезической образующей поверхности $I^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Пусть v есть направленный в будущее аффинный параметр вдоль каждой такой образующей. Тогда $\mathbf{K} = \alpha\partial/\partial v$, где α — функция вдоль образующей, подчиняющаяся уравнению $d\alpha/dv = \beta$. Если $\beta \neq 0$ и образующая геодезически полна в направлении прошлого, то функция α и вектор Киллинга $\mathbf{K}$ будут равны нулю в какой-то точке. Эта точка не может лежать в $I^+(\mathscr{I}^-, \bar{\mathscr{M}})$ и потому будет точкой пересечения горизонта событий будущего $I^-(\mathscr{I}^+, \bar{\mathscr{M}})$ и горизонта событий прошлого $I^+(\mathscr{I}^-, \bar{\mathscr{M}})$ [14]. Если $\beta = 0$, то вектор $\mathbf{K}$ всюду отличен от нуля, и ни в одной точке горизонт не разветвляется.

Израэль [84] показал, что статическое регулярно предсказуемое пространство-время должно быть шварцшильдовым, если:

а) $T_{ab} = 0$;

б) величина $f \equiv K^a K_a$ для вектора Киллинга K_a всюду в $I^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap I^-(\mathscr{I}^+, \bar{\mathscr{M}})$ обладает ненулевым градиентом;

в) горизонт событий прошлого $I^+(\mathscr{I}^-, \bar{\mathscr{M}})$ пересекает горизонт событий будущего $I^-(\mathscr{I}^+, \bar{\mathscr{M}})$ по компактной 2-поверхности $\mathscr{F}$.

[Из условия (в) и предложения 9.3.2 следует, что $\mathscr{F}$ связна и обладает топологией 2-сферы. Израэль не приводит условия (а)—(в) в точно такой форме, но его и наша формулировки эквивалентны.] Далее, он показал [85], что при замене условия отсутствия материи (а) требованием, чтобы тензор T_{ab} был тензором энергии-импульса электромагнитного поля, пространство-время оказывается пространством Райсснера — Нордстрема. В работе [111] условие (б) было снято для случая вакуума.

На основании этих результатов мы полагаем, что в случае статичности конечного состояния решения вне горизонта событий метрика во внешней области должна быть метрикой Шварцшильда.

Теперь мы рассмотрим случай, когда конечное состояние внешнего решения стационарное, но не статическое. Мы считаем, что это имеет место, когда сколлапсировавший объект в исходном состоянии вращался.

Предложение 9.3.5

В пустом стационарном регулярно предсказуемом пространстве, если оно нестатическое, вектор Киллинга K^a

пространственноподобен в некоторой части внешней области $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$.

Область в $\overline{J^+}((\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$, в которой вектор K^a пространственноподобен, называется *эргосферой*. Из предложения 9.3.4 следует, что эргосферы не существует, если метрика статическая. Значение эргосферы состоит в том, что в ней невозможно движение частицы по интегральной кривой вектора Киллинга K^a, т. е. частица не может оставаться в покое при наблюдении ее с бесконечности. Поскольку эргосфера расположена снаружи от горизонта, частица из нее все же может уйти на бесконечность. Примером стационарного нестатического регулярно предсказуемого пространства с эргосферой служит решение Керра при $a^2 \leqslant m^2$ (разд. 5.6). В работах [129, 133] показано, что из черной дыры с эргосферой можно извлечь определенное количество энергии, бросив частицу из бесконечности в эргосферу. Поскольку частица будет двигаться по геодезической, величина $E_0 \equiv -p_0{}^a K_a > 0$ будет постоянна вдоль ее траектории

$$(p_0{}^a K_a)_{;b}\, p_0{}^b = (p_0{}^a{}_{;b} p_0{}^b) K_a + p_0{}^a K_{a;b} p_0{}^b = 0,$$

так как $p_0{}^a$ — геодезический вектор и K^a — вектор Киллинга; здесь $p_0{}^a = m v_0{}^a$ — вектор импульса частицы, m — ее масса покоя и $\mathbf{v}_0$ — единичный вектор, касательный к мировой линии частицы. Предположим затем, что частица распадается на две частицы с импульсами $p_1{}^a$ и $p_2{}^a$, где $p_0{}^a = p_1{}^a + p_2{}^a$. Поскольку вектор K^a пространственноподобен, $p_1{}^a$ можно выбрать так, чтобы он был направленным в будущее времениподобным вектором, для которого $E_1 \equiv -p_1{}^a K_a < 0$. Тогда будем иметь $E_2 \equiv -p_2{}^a K_a > E_0$. Это означает, что вторая частица может уйти на бесконечность, где она будет иметь бóльшую энергию, чем первоначально брошенная частица. Таким образом, мы извлечем из черной дыры некоторое количество энергии.

Частица с отрицательной энергией не может уйти на бесконечность, а должна оставаться в области, где вектор K^a пространственноподобен. Допустим, что эргосфера не пересекает горизонт событий $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Тогда эта частица должна будет оставаться во внешней области. Повторяя описанный процесс, можно продолжать извлечение энергии из черной дыры. При этом решение, по-видимому, будет постепенно меняться. Однако эргосфера не может исчезнуть, поскольку должно остаться какое-то место для частиц с отрицательной энергией. Как мы видим, возможно одно из двух: или мы можем извлечь бесконечное количество энергии (что представляется невероятным), или эргосфера должна в конце концов пересечь горизонт. Мы покажем, что в последнем случае решение должно спонтанно пре-

вратиться в аксиально симметричное или статическое и при этом теряется возможность извлечения энергии с помощью процесса Пенроуза. Как возможность извлечения бесконечного количества энергии, так и спонтанное изменение решения, по-видимому, говорят о неустойчивости исходного состояния черной дыры. Поэтому представляется разумным предположить, что в любом реалистичном состоянии черной дыры эргосфера пересекает горизонт.

В работе [67] показано, что предельная стационарная поверхность, которая является внешней границей эргосферы, будет содержать по крайней мере две интегральные изотропные геодезические кривые вектора K^a. Если на этих кривых $f_{;a} \neq 0$ и если они геодезически полны в прошлом, будут точки, в которых $K^a = 0$. Но во внешней области таких точек быть не может (см. предложение 9.3.3); следовательно, в этом случае эргосфера пересекает горизонт. Однако предположение о полноте в прошлом интегральных кривых векторного поля K^a, по-видимому, неприемлемо, так как оно представляло бы собой определенное предположение относительно прошлой области решения, которая, как мы отмечали прежде, не является физически значимым; вместе с тем возможно и имеет смысл предполагать, что эти кривые полны в будущем. В статическом случае можно было бы показать, что решение симметрично во времени, но нет никакой априорной причины для того, чтобы такой симметрией обладало стационарное нестатическое решение. Поэтому мы будем оправдывать предположение о пересечении эргосферой горизонта не результатами [67], а рассуждениями, связанными с извлечением энергии.

Смысл соприкосновения эргосферы с горизонтом можно объяснить следующим образом. Пусть $\mathscr{L}_1$ — односвязная компонента области $I^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap I^+(\mathscr{I}^-, \bar{\mathscr{M}})$, и пусть $\mathscr{G}_1$ — фактормножество $\mathscr{L}$ относительно ее образующих. Согласно предложениям 9.3.1 и 9.3.2, оно гомеоморфно 2-сфере. По предложению 9.3.1 пространственное разделение двух соседних образующих постоянно вдоль них и потому может быть описано индуцированной на $\mathscr{G}_1$ метрикой $\mathbf{h}$. Изометрия θ_t переводит образующие в образующие и, таким образом, действует как группа изометрии пространства $(\mathscr{G}_1, \mathbf{h})$. Если эргосфера пересекает горизонт, то вектор K_a где-то на горизонте пространственноподобен и действие θ_t в $(\mathscr{G}_1, \mathbf{h})$ будет нетривиальным. Тогда оно должно соответствовать вращению сферы $\mathscr{G}_1$ вокруг некоторой оси, и орбитами этой группы в $\mathscr{G}_1$ будут две точки, соответствующие полюсам, и семейство окружностей. Частица, движущаяся вдоль одной из образующих горизонта, казалась бы тогда движущейся относительно определяемой вектором K^a системы отсчета, которая стационарна на бесконечности. Поэтому мы

могли бы сказать, что горизонт вращается относительно бесконечности.

Следующий результат показывает, что вращающаяся черная дыра должна быть аксиально симметричной.

Предложение 9.3.6

Пусть $(\mathscr{M}, \mathbf{g})$ — стационарное нестатическое регулярно предсказуемое пространство, в котором эргосфера имеет пересечение с $I^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap I^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Тогда существует однопараметрическая циклическая группа изометрии $\tilde{\theta}_\phi (0 \leqslant \phi < 2\pi)$ этого пространства, которая коммутирует с θ_t и орбиты которой пространственноподобны вблизи $\mathscr{I}^+$ и $\mathscr{I}^-$.

Метод доказательства этого предложения состоит в использовании аналитичности метрики $\mathbf{g}$ для того, чтобы показать существование изометрии $\tilde{\theta}_\phi$ в некоторой окрестности горизонта. Затем эта изометрия распространяется на все пространство путем аналитического продолжения. Этот метод был бы применим, даже если бы метрика была неаналитична в изолированных областях вне горизонта, например, если бы вокруг черной дыры существовало кольцо из вещества или какая-то рамка из стержней. В этом случае возникает следующий кажущийся парадокс. Рассмотрим вращающуюся звезду, окруженную стационарной квадратной рамкой, сделанной из стержней. Если черная дыра приближается к стационарному состоянию, то из предложения 9.3.6 следует, что метрика $\mathbf{g}$ должна быть аксиально симметричной везде, за исключением области, где она неаналитична (на стержнях). Однако гравитационное воздействие стержней помешает метрике стать аксиально симметричной. Разрешение парадокса состоит, по-видимому, в том, что в этом случае черная дыра не может быть стационарной, пока она вращается. Должно произойти следующее. Гравитационное воздействие стержней слегка деформирует черную дыру. Обратное влияние черной дыры на рамку вызовет ее вращение и, таким образом, приведет к потере момента количества движения через излучение. В итоге вращение как черной дыры, так и рамки затормозится, и решение будет приближаться к статическому состоянию. Статическая черная дыра может и не быть аксиально симметричной, если пространство вне ее не пустое, т. е. если условие (а) теоремы Израэля не удовлетворяется.

Это рассуждение свидетельствует о том, что реальная черная дыра никогда не будет строго стационарной, пока она вращается, так как Вселенная вокруг нее не будет строго аксиально симметричной. Однако в большинстве случаев скорость замедления вращения такой черной дыры будет крайне мала [70, 136]. Поэтому в хорошем приближении можно пренебречь

малой асимметрией, которую вызывает материя в удалении от вращающейся черной дыры, и рассматривать такую черную дыру как если бы она находилась в стационарном состоянии. Поэтому мы будем изучать свойства вращающейся аксиально симметричной черной дыры.

Из следующего результата Папапетру [119], обобщенного Картером [25], видно, что векторы Киллинга: K^a, соответствующий временной трансляции, и $\tilde{K}^a$, соответствующий угловому вращению $\tilde{\theta}_\phi$, — оба ортогональны к некоторым семействам 2-поверхностей.

Предложение 9.3.7

Пусть $(\mathscr{M}, \mathbf{g})$ — пространство-время, допускающее 2-параметрическую абелеву группу изометрии с векторами Киллинга ξ_1 и ξ_2. Пусть $\mathscr{V} \subset \mathscr{M}$ — связное открытое множество, и пусть $w_{ab} \equiv \xi_{1[a}\xi_{2b]}$. Тогда, если

а) $w_{ab}R^b{}_c\eta^{cdef}w_{ef} = 0$ на $\mathscr{V}$,

б) $w_{ab} = 0$ в некоторой точке из $\mathscr{V}$, то $w_{[ab;c}w_{d]e} = 0$ на $\mathscr{V}$.

В стационарном аксиально симметричном пространстве условие (б) удовлетворяется на оси симметрии, т. е. на множестве точек, где $\tilde{K}^a = 0$. Условие (а) удовлетворяется в пустом пространстве, а также в присутствии свободного (без источников) электромагнитного поля [25]. По теореме Фробениуса [151] при $w_{ab} \neq 0$ равенство нулю $w_{[ab;c}w_{d]e}$ представляет собой условие локального существования семейства 2-поверхностей, ортогональных w_{ab}, т. е. любой линейной комбинации ξ_1 и ξ_2. В случае аксиально симметричного стационарного пространства-времени это означает, что локально мы можем ввести координаты (t, ϕ, x^1, x^2), в которых $\mathbf{K} = \partial/\partial t$, $\tilde{K} = \partial/\partial\phi$ и $K^a x^m{}_{;a} = 0 = \tilde{K}^a x^m{}_{;a}$ при $m = 1, 2$. Тогда метрика локально допускает изометрию $(t, \phi, x^1, x^2) \to (-t, -\phi, x^1, x^2)$, обращающую направление течения времени, т. е. метрика симметрична во времени. Таким образом, если аналитическое расширение по метрике вблизи бесконечности пустого стационарного регулярно предсказуемого пространства-времени содержит горизонт будущего, то оно содержит и горизонт событий прошлого.

По аналогии с предложением 9.3.4 имеем предложение 9.3.8.

Предложение 9.3.8 (ср. с [28])

Пусть $(\mathscr{M}, \mathbf{g})$ — стационарное аксиально симметричное регулярно предсказуемое пространство-время, в котором $w_{[ab;c}w_{d]e} = 0$, где $w_{ab} \equiv K_{[a}\tilde{K}_{b]}$. Тогда $h \equiv w_{ab}w^{ab} < 0$ в любой точке, лежащей вне оси $\tilde{\mathbf{K}} = 0$ во внешней области $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. На горизонтах $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ и $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ $h = 0$, но w_{ab} нигде не обращается в нуль, кроме оси.

Отсюда видно, что во внешней области в каждой точке вне оси существует некоторая линейная комбинация векторов Киллинга K^a и $\tilde{K}^a$, которая времениподобна. Снаружи эргосферы времениподобен сам вектор K^a, но чтобы получить времениподобный вектор между предельной стационарной поверхностью и горизонтом, нужно к K^a добавить вектор, пропорциональный $\tilde{K}^a$. На горизонте нет никакой времениподобной линейной комбинации, но существует изотропная линейная комбинация, направленная вдоль изотропных образующих горизонта. Вне оси $\tilde{\mathbf{K}} = 0$ горизонт можно локально охарактеризовать как множество точек, на котором $h \equiv \omega_{ab}\omega^{ab} = 0$.

Тогда мы приходим к теореме Картера [28], которая указывает на то, что объекты, описываемые решениями Керра, вероятно, являются единственными стационарными черными дырами в пустом пространстве. Картер рассмотрел стационарное регулярно предсказуемое пространство, удовлетворяющее условиям

а) $T_{ab} = 0$;

б) аксиально симметричное;

в) горизонт событий прошлого $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ пересекает горизонт событий будущего $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ по компактной связной 2-поверхности $\mathscr{F}_1$. (Согласно предложению 9.3.2, она будет 2-сферой.) Картер показал, что такие решения распадаются на два несвязных семейства, каждое из которых зависит от двух параметров. В качестве этих двух параметров можно принять массу m и момент количества движения L, измеренные с бесконечности. Одно из семейств известно — это решения Керра с $m \geqslant 0$, $a^2 \leqslant m^2$, где $a = L/m$. (Решения Керра с $a^2 > m^2$ содержат голую сингулярность и потому не являются регулярно предсказуемыми пространствами.) Представляется маловероятным, чтобы существовали какие-либо другие несвязные семейства. Поэтому было высказано предположение, что вне незаряженного сколлапсировавшего объекта установится решение Керра с $a^2 \leqslant m^2$. Это предположение подтверждается анализом линейных возмущений при сферическом коллапсе [43, 137, 139, 168].

В предположении, что эта гипотеза Картера — Израэля справедлива, можно полагать, что площадь 2-поверхности $\partial\mathscr{B}(\tau)$ на горизонте событий будет стремиться к площади некоторой 2-поверхности на горизонте событий $r = r_+$ в решении Керра при такой же массе и таком же моменте количества движения, измеренных с поверхности $\mathscr{L}(\tau)$ на $\mathscr{I}^+$. Эта площадь равна $8\pi m(m + (m^2 - a^2)^{1/2})$, где m — масса в решении Керра и ma — момент количества движения. (Если коллапсирующее тело имеет электрический заряд e, то следует ожидать, что

окончательно установится решение Керра с учетом заряда. Площадь 2-поверхности на горизонте событий такого решения равна

$$4\pi\left[2m^2 - e^2 + 2m\left(m^2 - a^2 - e^2\right)^{1/2}\right].$$

Используя это выражение, можно обобщить наши результаты на заряженные черные дыры.) Рассмотрим коллапс, в котором к поверхности $\mathscr{S}(\tau_1)$ устанавливается решение Керра с массой m_1 и моментом количества движения $m_1 a_1$. Предположим, что затем в течение конечного времени черная дыра взаимодействует с частицами или излучением. В итоге к поверхности $\mathscr{S}(\tau_2)$ установится иное решение Керра с параметрами m_2, a_2. Из рассуждений в разд. 9.2 следует, что площадь $\partial\mathscr{B}(\tau_2)$ должна быть больше или равна площади $\partial\mathscr{B}(\tau_1)$. В действительности она должна быть строго больше, так как $\hat{\theta}$ может быть равно нулю лишь при условии, что никакое вещество или излучение не пересекает горизонт. Это означает тогда, что

$$m_2\left[m_2 + \left(m_2^2 - a_2^2\right)^{1/2}\right] > m_1\left[m_1 + \left(m_1^2 - a_1^2\right)^{1/2}\right]. \tag{9.4}$$

Если $a_1 \neq 0$, то неравенство (9.4) позволяет, чтобы m_2 была меньше m_1. Поскольку в асимптотически плоском пространстве-времени [122] выполняется закон сохранения полной энергии и импульса, неравенство $m_2 < m_1$ должно означать, что мы каким-то образом извлекли из черной дыры определенное количество энергии. Один из способов извлечения энергии мог бы состоять в том, чтобы построить вокруг черной дыры квадратную рамку из стержней и заставить момент силы, наводимый вращающейся черной дырой на рамку, совершать работу. С другой стороны, можно было бы использовать процесс Пенроуза: бросить в эргосферу частицу, разделить ее там на две частицы, одна из которых уйдет на бесконечность с энергией, большей чем у первоначальной частицы. Другая частица провалится через горизонт событий и уменьшит момент количества движения черной дыры. Таким образом, этот процесс можно рассматривать как процесс извлечения вращательной энергии черной дыры. Кристодулу [32] показал, что при этом можно достичь результата, сколь-угодно близкого к пределу, устанавливаемому неравенством (9.4). Наибольшая энергия извлекается при $a_2 = 0$; в этом случае полученная энергия $m_1 - m_2$ меньше, чем

$$m_1\left\{1 - \frac{1}{\sqrt{2}}\left[1 + \left(1 - \frac{a_1^2}{m_1^2}\right)^{1/2}\right]^{1/2}\right\}.$$

Рассмотрим теперь случай, когда коллапсируют две звезды, расположенные на значительном расстоянии друг от друга.

и возникают две черные дыры. Следовательно, найдется некоторое значение τ', при котором $\partial\mathscr{B}(\tau')$ состоит из двух отдельных 2-сфер $\partial\mathscr{B}_1(\tau')$ и $\partial\mathscr{B}_2(\tau')$. Поскольку они удалены друг от друга, можно пренебречь их взаимодействием и предположить, что решение вблизи каждой из черных дыр почти совпадает с решениями Керра с параметрами m_1, a_1 и m_2, a_2 соответственно. Отсюда площади $\partial\mathscr{B}_1(\tau')$ и $\partial\mathscr{B}_2(\tau')$ будут приблизительно равны

$$8\pi m_1\left[m_1+(m_1^2-a_1^2)^{1/2}\right] \quad \text{и} \quad 8\pi m_2\left[m_2+(m_2^2-a_2^2)^{1/2}\right]$$

соответственно. Теперь допустим, что эти черные дыры сталкиваются и сливаются. При таком столкновении выделится определенное количество гравитационного излучения. В итоге у поверхности $\mathscr{S}(\tau'')$ эта система придет к решению, похожему на одно решение Керра с параметрами m_3, a_3. В силу тех же аргументов, что и выше, площадь $\partial\mathscr{B}(\tau'')$ должна быть больше полной площади $\partial\mathscr{B}(\tau')$, равной сумме площадей $\partial\mathscr{B}_1(\tau')$ и $\partial\mathscr{B}_2(\tau')$. Итак,

$$m_3\left[m_3+(m_3^2-a_3^2)^{1/2}\right] > m_1\left[m_1+(m_1^2-a_1^2)^{1/2}\right]+ \\ + m_2\left[m_2+(m_2^2-a_2^2)^{1/2}\right].$$

Согласно закону сохранения для асимптотически плоских пространств, энергия, уносимая на бесконечность гравитационным излучением, равна

$$m_1+m_2-m_3$$

и ограничена приведенным выше неравенством. Эффективность превращения массы в гравитационное излучение

$$\varepsilon \equiv (m_1+m_2-m_3)(m_1+m_2)^{-1}$$

всегда меньше $1/2$. Если $a_1 = a_2 = 0$, то $\varepsilon < 1-1/\sqrt{2}$; следует подчеркнуть, что это верхние пределы, истинная эффективность может быть много меньше, хотя само существование предельного значения может означать, что достижима заметная его доля.

Мы показали, что доля массы, способная превратиться в гравитационное излучение при слиянии двух черных дыр, ограничена. Если бы, однако, первоначально было большое число черных дыр, то они могли бы слиться попарно, возникшие при этом новые черные дыры снова могли бы слиться и т. д. Из соображений размерности можно ожидать, что на всех этапах эффективность будет одинаковой. Таким образом, в гравитационное излучение могла бы превратиться очень большая часть первоначальной массы. (Этот аргумент выдвинули Мизнер и Рис.) На каждом последующем этапе энергия, излу-

ченная в виде гравитационных волн, будет больше, чем на предыдущем. Этот механизм мог бы объяснить наблюдавшиеся Вебером всплески гравитационного излучения.

Теперь мы приведем доказательства предложений, выдвинутых в этом разделе. Для удобства мы повторим их формулировки.

Предложение 9.3.1

Пусть $(\mathscr{M}, \mathbf{g})$ — стационарное регулярно предсказуемое пространство-время. Тогда образующие горизонта событий будущего $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ не имеют конечных точек в прошлом в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Пусть $Y_1{}^a$ — направленные в будущее касательные к этим образующим векторы; тогда в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ вектор $Y_1{}^a$ обладает нулевыми сдвигом $\hat{\sigma}$ и расхождением $\hat{\theta}$ и удовлетворяет равенству

$$R_{ab}Y_1{}^aY_1{}^b = 0 = Y_{[e}C_{a]bc[d}Y_{1f]}Y_1{}^bY_1{}^c.$$

Пусть $\mathscr{C}$ — пространственноподобная 2-сфера на $\mathscr{I}^-$. Тогда $\mathscr{I}^-$ можно покрыть семейством 2-сфер $\mathscr{C}(t)$, полученных движением $\mathscr{C}$ в обе стороны вдоль образующих поверхности $\mathscr{I}^-$ и под действием θ_t, т. е. $\mathscr{C}(t) = \theta_t(\mathscr{C})$. Введем теперь функцию x в точке $p \in J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ как наибольшее значение t, при котором $p \in J^+(\mathscr{C}(t), \bar{\mathscr{M}})$. Пусть $\mathscr{U}$ есть окрестность $\mathscr{I}^+$ и $\mathscr{I}^-$, которая изометрична соответствующей окрестности асимптотически простого пространства-времени. Тогда функция x будет непрерывна и будет иметь некоторую нижнюю грань x' на $\mathscr{S} \cap \mathscr{U}$. Из этого следует, что x непрерывна в той области множества $\overline{J^-}(\mathscr{I}^+, \mathscr{M})$, где она больше x'. Пусть $p \in J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \cap J^-(\mathscr{I}^+, \mathscr{M})$. Тогда под действием изометрии θ_t точка p будет переведена в область множества $\overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$, где $x > x'$. Однако

$$x|_{\theta_t(p)} = x|_p + t.$$

Следовательно, x непрерывна в p.

Пусть $\tau_0 > 0$ таково, что $\mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ содержится в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Пусть λ — образующая поверхности $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, пересекающей $\mathscr{S}(\tau_0)$. Допустим, что на λ существует некоторая конечная верхняя грань x_0 функции x. Поскольку рассматриваемое пространство асимптотически простое в слабом смысле, $x \to \infty$ по мере приближения к $\mathscr{L}(\tau_0)$ на $\mathscr{S}(\tau_0)$. Следовательно, на

$$\mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$$

функция x должна иметь некоторую нижнюю грань. Под действием группы θ_t образующая λ преобразуется в другую образующую $\theta_t(\lambda)$. Поскольку образующие поверхности $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ не имеют никаких конечных точек в будущем, продолжение в прошлое кривой $\theta_t(\lambda)$ все же будет пересекать $\mathscr{S}(\tau_0) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})}$. Это ведет к противоречию, поскольку тогда верхняя грань x на $\theta_t(\lambda)$ была бы меньше x_1, если $t < x_1 - x_0$.

Пусть x_2 есть верхняя грань x на $\mathscr{S}(\tau_0) \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Тогда каждая образующая λ поверхности $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, пересекающая $\mathscr{S}(\tau_0)$, пересечет $\mathscr{F}(t) \equiv \dot{J}^+(\mathscr{C}(t), \bar{\mathscr{M}}) \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ при $t \geqslant x_2$. Каждая образующая $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, которая пересекает $\mathscr{F}(t')$, будет пересекать $\theta_t(\mathscr{S}(\tau_0))$ при $t \geqslant t' - x_1$. Но поверхность $\theta_t(\mathscr{S}(\tau_0)) \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) = \theta_t(\mathscr{S}(\tau_0) \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}))$ компактна. Следовательно, компактна и $\mathscr{F}(t)$.

Рассмотрим теперь, как меняется площадь $\mathscr{F}(t)$ по мере роста t. Поскольку $\hat{\theta} \geqslant 0$, эта площадь не может сокращаться. Если бы было $\hat{\theta} > 0$ на некотором открытом множестве, площадь увеличивалась бы. Она увеличивалась бы также, если бы образующие горизонта обладали конечными точками в прошлом на $\mathscr{F}(t)$. Однако, так как $\mathscr{F}(t)$ движется под действием изометрии θ_t, площадь должна оставаться неизменной. Поэтому $\hat{\theta} = 0$, и в той области поверхности $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, для которой $x \geqslant x_2$, нет конечных точек в прошлом, а поскольку каждая точка из $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ может быть переведена изометрией θ_t в область, где $x > x_2$, этот результат относится ко всей поверхности $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Тогда из уравнений распространения (4.35) и (4.36) получаем: $\hat{\sigma}_{mn} = 0$, $R_{ab}Y_1{}^aY_1{}^b = 0$ и $Y_{1[e}C_{a]bc[d}Y_{1f]}Y_1{}^bY_1{}^c = 0$, где $Y_1{}^a$ — направленный в будущее касательный вектор к изотропной геодезической образующей горизонта. □

Предложение 9.3.2

В стационарном регулярно предсказуемом пространстве каждая связная компонента горизонта $\partial\mathscr{B}(\tau)$ в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ гомеоморфна 2-сфере.

Рассмотрим, как ведет себя расхождение уходящих ортогональных к $\partial\mathscr{B}(\tau)$ изотропных геодезических при малой деформации $\partial\mathscr{B}(\tau)$ наружу в $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Пусть $Y_2{}^a$ — другой направленный в будущее изотропный вектор, ортогональный к $\partial\mathscr{B}(\tau)$ и нормированный так, что $Y_1{}^aY_{2a} = -1$. При этом остается свобода преобразований $\mathbf{Y}_1 \to \mathbf{Y}_1' = e^y\mathbf{Y}_1$, $\mathbf{Y}_2 \to \mathbf{Y}_2' = e^{-y}\mathbf{Y}_2$. Индуцированная метрика на пространственноподобной 2-поверхности $\partial\mathscr{B}(\tau)$ равна $\hat{h}_{ab} = g_{ab} + Y_{1a}Y_{2b} + Y_{2a}Y_{1b}$. Введем семей-

ство поверхностей $\mathscr{F}(\tau, w)$, сдвигая каждую точку $\partial\mathscr{B}(\tau)$ на расстояние w по параметру вдоль изотропной геодезической с касательным вектором $Y_2{}^a$. Векторы $Y_1{}^a$ будут ортогональны к $\mathscr{F}(\tau, w)$, если они подчиняются уравнениям

$$\hat{h}_{ab}Y_1{}^b{}_{;c}Y_2{}^c = -\hat{h}_a{}^bY_{2c;b}Y_1{}^c \quad \text{и} \quad Y_1{}^aY_{2a} = -1.$$

Тогда

$$(Y_1{}^a{}_{;b}\hat{h}_a{}^c\hat{h}^b{}_d)_{;g}Y_2{}^g\hat{h}_c{}^s\hat{h}^d{}_t =$$

$$= \hat{h}^{sa}p_{a;b}\hat{h}^b{}_t + p^sp_t - \hat{h}^s{}_aY_1{}^a\hat{h}^{ge}Y_{2e;b}\hat{h}^b{}_t + R^a{}_{ceb}Y_2{}^eY_1{}^ch_a{}^sh^b{}_t, \quad (9.5)$$

где $p^a \equiv -\hat{h}^{ba}Y_{2c;b}Y_1{}^c$. Свертывая с $\hat{h}_s{}^t$, получаем

$$\frac{\mathrm{d}\hat{\theta}}{\mathrm{d}w} = (Y_1{}^a{}_{;b}\hat{h}^b{}_a)_{;c}Y_2{}^c =$$

$$= p_{b;d}\hat{h}^{bd} - R_{ac}Y_1{}^aY_2{}^c + R_{adcb}Y_1{}^dY_2{}^cY_2{}^aY_1{}^b +$$

$$+ p_ap^a - Y_1{}^a{}_{;c}\hat{h}^c{}_dY_2{}^d{}_{;b}\hat{h}^b{}_a.$$

На горизонте $Y_1{}^a{}_{;c}\hat{h}^{cd}\hat{h}^b{}_a = 0$, поскольку сдвиг и расхождения горизонта равны нулю. При масштабном преобразовании $\mathbf{Y}_1{}' = e^y\mathbf{Y}_1$, $\mathbf{Y}_2{}' = e^{-y}\mathbf{Y}_2$ вектор p^a заменяется на $p'^a = p^a + \hat{h}^{ab}y_{;b}$ и поэтому изменение $\mathrm{d}\hat{\theta}/\mathrm{d}w\,|_{w=0}$ равно

$$\left.\frac{\mathrm{d}\hat{\theta}'}{\mathrm{d}w'}\right|_{w=0} = p_{b;d}\hat{h}^{bd} + y_{;bd}\hat{h}^{bd} - R_{ac}Y_1{}^aY_2{}^c +$$

$$+ R_{abcb}Y_1{}^dY_2{}^cY_2{}^aY_1{}^b + p'^ap'_a. \quad (9.6)$$

Член $y_{;bd}\hat{h}^{bd}$ есть лапласиан y на 2-поверхности $\partial\mathscr{B}(\tau)$. Согласно теореме Ходжа [78], можно выбрать y так, что сумма первых четырех членов в правой части (9.6) будет постоянной на $\partial\mathscr{B}(\tau)$. Знак этой постоянной будет определяться знаком интеграла от

$$(-R_{ac}Y_1{}^aY_2{}^c + R_{adcb}Y_1{}^dY_2{}^cY_2{}^aY_1{}^b)$$

по $\partial\mathscr{B}(\tau)$ (интеграл от $p_{b;d}\hat{h}^{bd}$ равен нулю как интеграл от дивергенции). Этот интеграл можно вычислить, используя уравнение Гаусса — Кодацци для скалярной кривизны $\hat{R}$ 2-поверхности с метрикой $\hat{h}$:

$$\hat{R} = R_{ijkl}\hat{h}^{ik}\hat{h}^{jl} = R - 2R_{ijkl}Y_1{}^iY_2{}^jY_1{}^kY_2{}^l + 4R_{ij}Y_1{}^iY_2{}^j,$$

так как $\hat{\theta} = \hat{\sigma} = 0$ на $\partial\mathscr{B}(\tau)$. По теореме Гаусса — Бонне [92]

$$\int_{\partial\mathscr{B}(\tau)} \hat{R}\,\mathrm{d}\hat{S} = 2\pi\chi,$$

где $\mathrm{d}\hat{S}$ — элемент площади и χ — эйлерова характеристика поверхности $\partial\mathscr{B}(\tau)$. Итак,

$$\int\limits_{\partial\mathscr{B}(\tau)} (-R_{ab}Y_1{}^aY_2{}^b + R_{abcb}Y_1{}^dY_2{}^cY_2{}^aY_1{}^b)\,\mathrm{d}\hat{S} =$$

$$= -\pi\chi + \int\limits_{\partial\mathscr{B}(\tau)} \left(\frac{1}{2}R - R_{ab}Y_1{}^aY_2{}^b\right)\mathrm{d}\hat{S}\,. \qquad (9.7)$$

В силу уравнений Эйнштейна

$$\frac{1}{2}R + R_{ab}Y_1{}^aY_2{}^b = 8\pi T_{ab}Y_1{}^aY_2{}^b \geqslant 0;$$

последнее неравенство следует из условия энергодоминантности. Эйлерова характеристика χ равна $+2$ для сферы, 0 для тора и отрицательна для любой другой компактной ориентируемой 2-поверхности (поверхность $\partial\mathscr{B}(\tau)$ должна быть ориентируемой, поскольку она является краем). Отсюда правая часть (9.7) может быть отрицательной лишь при условии, что $\partial\mathscr{B}(\tau)$ — сфера.

Допустим, что правая часть (9.7) положительна. Тогда можно выбрать y так, что $\mathrm{d}\hat{\theta}'/\mathrm{d}w'|_{w=0} > 0$ всюду на $\partial\mathscr{B}(\tau)$. При малых отрицательных значениях w' можно получить в $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ такую 2-поверхность, что ортогональные к ней уходящие изотропные геодезические будут сходиться. Это противоречит предложению 9.2.8. Допустим теперь, что $\chi = 0$ и что $T_{ab}Y_1{}^aY_2{}^b = 0$ на $\partial\mathscr{B}(\tau)$. Тогда можно выбрать y так, чтобы сумма первых четырех членов в правой части (9.6) была равна нулю на $\partial\mathscr{B}(\tau)$. При этом на

$$p'^a{}_{;b}\hat{h}^b{}_a + R_{abcd}Y_1{}^aY_2{}^bY_1{}^cY_2{}^d = 0.$$

Если $R_{abcd}Y_1{}^aY_2{}^bY_1{}^cY_2{}^d \neq 0$ где-либо на $\partial\mathscr{B}(\tau)$, то и член $p'^ap'_a$ в (9.6) должен быть где-то отличен от нуля; можно слегка изменить y так, чтобы было $\mathrm{d}\hat{\theta}'/\mathrm{d}w'|_{w=0} > 0$ всюду. Так мы снова приходим к противоречию.

Теперь допустим, что всюду на $\partial\mathscr{B}(\tau)$ $R_{abcd}Y_1{}^aY_2{}^bY_1{}^cY_2{}^d = 0$ и $p'^a = 0$. 2-поверхность $\partial\mathscr{B}(\tau)$ можно сдвинуть назад вдоль $Y_2{}^a$, выбирая y на каждом этапе так, что

$$p'^a{}_{;b}\hat{h}^b{}_a + R_{abcd}Y_1{}^aY_2{}^bY_1{}^cY_2{}^d - \frac{1}{2}R -$$

$$- 2R_{ab}Y_1{}^aY_2{}^b = p'^a{}_{;b}\hat{h}^b{}_a - \frac{1}{2}\hat{R} = 0.$$

Если бы $T_{ab}Y_1{}^aY_2{}^b$ или p'^a не были равны нулю при $w' < 0$, то можно было бы подобрать y так, чтобы получить в $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ 2-поверхность с $\hat{\theta} < 0$. Это противоречило бы предложению

9.2.8. С другой стороны, если бы при $\varpi' < 0$ всюду было $T_{ab}Y_1{}^aY_2{}^b = 0$ и $p'^a = 0$, то мы получили бы 2-поверхность с $\hat{\theta} = 0$ в $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, что также противоречит предложению 9.2.8.

Противоречия не возникает только при $\chi = 2$, т. е. если $\partial\mathscr{B}(\tau)$ есть 2-сфера. □

Предложение 9.3.3

Пусть $(\mathscr{M}, \mathbf{g})$ — стационарное регулярно предсказуемое пространство-время. Тогда вектор Киллинга K^a отличен от нуля в области $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})}$, которая односвязна. Пусть дано τ_0, при котором $\mathscr{S}(\tau_0) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})}$ содержится в $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Если $\partial\mathscr{B}(\tau_0)$ имеет только односвязную компоненту, то область $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})} \cap \bar{\mathscr{M}}$ гомеоморфна $[0, 1) \times S^2 \times R^1$.

Функция x, введенная в предложении 9.3.1, непрерывна на $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})}$ и обладает тем свойством, что $x|_{\theta_t(p)} = x|_p + t$. Отсюда видно, что $\mathbf{K}$ не может равняться нулю в $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})}$. Интегральные кривые $\mathbf{K}$ устанавливают гомеоморфизм между двумя из поверхностей

$$\dot{J}^+(\mathscr{C}(t), \bar{\mathscr{M}}) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})} \cap \mathscr{M} \quad (-\infty < t < \infty).$$

Область $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})} \cap \mathscr{M}$ покрывается этими поверхностями и, следовательно, гомеоморфна $R^1 \times \dot{J}^+(\mathscr{C}(t'), \bar{\mathscr{M}}) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})} \cap \mathscr{M}$ при любом t'. Выберем t достаточно большим, чтобы поверхность $\dot{J}^+(\mathscr{C}(t), \bar{\mathscr{M}})$ имела пересечение с $\mathscr{S}(\tau_0)$ в такой окрестности $\mathscr{U}$ поверхности $\mathscr{I}^+$, которая изометрична соответствующей поверхности в асимптотически простом пространстве. Интегральные кривые $\mathbf{K}$ устанавливают гомеоморфизм между

$$\dot{J}^+(\mathscr{C}(t), \bar{\mathscr{M}}) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})} \cap \mathscr{M} \text{ и } \mathscr{S}(\tau_0) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})}.$$

Последняя область односвязна в силу свойства (α) и предложения 9.3.2. Если далее $\partial\mathscr{B}(\tau)$ имеет только одну связную компоненту, то

$$\mathscr{S}(\tau_0) \cap \overline{J^-(\mathscr{I}^-, \bar{\mathscr{M}})}$$

обладает топологией $[0, 1) \times S^2$. Отсюда топологически $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-(\mathscr{I}^+, \bar{\mathscr{M}})} \cap \mathscr{M}$ есть $[0, 1) \times S^2 \times R^1$. □

Предложение 9.3.4

В статическом регулярно предсказуемом пространстве-времени вектор Киллинга $\mathbf{K}$ времениподобен во внешней области $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, а на $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ отличен от

нуля и направлен вдоль изотропных образующих поверхности $\dot{J}^-(\mathscr{I}^+, \overline{\mathscr{M}})$.

Горизонт событий $\dot{J}^-(\mathscr{I}^+, \overline{\mathscr{M}})$ изометрией θ_t отображается в себя. Следовательно, на $\dot{J}^-(\mathscr{I}^+, \overline{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \overline{\mathscr{M}})$ вектор $\mathbf{K}$ должен быть либо изотропным, либо пространственноподобным. Пусть при $\tau = \tau_0$ $\mathscr{S}(\tau_0) \cap \overline{J}^-(\mathscr{I}^+, \mathscr{M})$ содержится в $J^+(\mathscr{I}^-, \overline{\mathscr{M}})$. Тогда функция $f \equiv K^a K_a$ должна равняться нулю на некотором замкнутом множестве $\mathscr{N}$ в $J^+(\mathscr{S}(\tau_0)) \cap \overline{J}^-(\mathscr{I}^+, \overline{\mathscr{M}})$.

Из того, что K^a есть вектор Киллинга и $\operatorname{rot} \mathbf{K} = 0$, следует

$$f K_{a;b} = K_{[a} f_{;b]}. \tag{9.8}$$

Согласно предложению 9.3.3, $K^a \neq 0$ на односвязном множестве $J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap \overline{J}^-(\mathscr{I}^+, \overline{\mathscr{M}})$. По теореме Фробениуса из условия $\operatorname{rot} \mathbf{K} = 0$ следует, что в этой области найдется функция ξ, такая, что $K_a = -\alpha \xi_{;a}$, где α — некоторая положительная функция.

Пусть p — точка множества $\mathscr{N}$ и $\lambda(v)$ — кривая, проходящая через p и лежащая на поверхности постоянного ξ. Тогда в силу (9.8),

$$\frac{1}{2} K^a \frac{\mathrm{d}}{\mathrm{d}v} \ln f = \frac{\mathrm{D}}{\partial v} K^a.$$

Если $\lambda(v)$ выходит за пределы $\mathscr{N}$, то левая часть этого уравнения будет не ограничена. Но правая часть непрерывна, поэтому кривая $\lambda(v)$ должна лежать в $\mathscr{N}$ и множество $\mathscr{N}$ должно содержать поверхность $\xi = \xi|_p$. Однако на некоторой открытой окрестности точки p функция f должна быть отлична от нуля, потому что иначе она была бы равна нулю всюду. Следовательно, связная компонента $\mathscr{N}$, содержащая p, есть 3-поверхность $\xi = \xi|_p$. Допустим, что $p \in J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \overline{\mathscr{M}})$. Тогда найдется направленная в будущее времениподобная кривая $\gamma(u)$ от $\mathscr{I}^-$ до $\mathscr{I}^+$, проходящая через p. На $\xi = \xi|_p$ вектор K^a должен быть направлен в будущее. Отсюда $(\partial/\partial u)_\gamma \xi > 0$ при $\xi = \xi|_p$. Это ведет к противоречию: $\xi = \xi|_p$ не может иметь пересечения с $\mathscr{I}^+$ или $\mathscr{I}^-$, потому что K^a вблизи бесконечности времениподобен. Таким образом, вблизи $\mathscr{I}^+$ и $\mathscr{I}^-$ либо $\xi > \xi|_p$, либо $\xi < \xi|_p$. □

Предложение 9.3.5

В пустом стационарном регулярно предсказуемом пространстве-времени, если оно нестатическое, вектор Киллинга K^a пространственноподобен в некоторой части внешней области $J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \overline{\mathscr{M}})$.

Функция x, введенная в предложении 9.3.1, непрерывна на $J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \overline{\mathscr{M}})$ и притом $\partial x/\partial t = 1$ вдоль каждой интегральной кривой. Поверхность $x = 0$ в $J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \overline{\mathscr{M}})$ можно аппроксимировать гладкой поверхностью $\mathscr{K}$, которая нигде не касается K^a. Затем на $J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \overline{\mathscr{M}})$ можно ввести гладкую функцию $\bar{x}$, обладающую тем свойством, что $\bar{x} = 0$ на $\mathscr{K}$ и $\bar{x}_{;a}K^a = 1$. Градиент вектора Киллинга можно представить в виде

$$fK_{a;\,b} = \eta_{abcd}K^c\omega^d + K_{[a}f_{;b]},$$

где $f \equiv K^aK_a$ и

$$\omega^a \equiv \frac{1}{2}\eta^{abcd}K_bK_{c;d}.$$

Вторая производная **K** удовлетворяет равенству

$$2K_{a;[bc]} = R_{dabc}K^d.$$

Однако $K_{a;bc} = K_{[a;b]c}$. Следовательно,

$$K_{a;bc} = R_{dcba}K^d,$$

и отсюда

$$K^{a;b}{}_{b} = -R^a{}_dK^d. \tag{9.9}$$

Вектор $q_a \equiv f^{-1}K_a = \bar{x}_{;a}$ ортогонален к K^a. Умножим (9.9) на q_a и проинтегрируем в $J^-(\mathscr{I}^+, \overline{\mathscr{M}})$ по области $\mathscr{L}$, ограниченной поверхностями $\mathscr{N}_1$ и $\mathscr{N}_2$, которые заданы уравнениями $\bar{x} = x_2 + 1$ и $x = x_2 + 2$, где x_2 имеет тот же смысл, что и в приложении 9.3.1; в результате получим

$$\int_{\mathscr{L}} R_{ab}K^aq^b\,dv = -\int_{\mathscr{L}} (K^{a;b}q_a)_{;b}\,dv + \int_{\mathscr{L}} K_{a;b}q^{a;b}\,dv =$$

$$= \int_{\partial\mathscr{L}} K^{a;b}q_a\,d\sigma_b - 2\int_{\mathscr{L}} f^{-2}\omega^a\omega_a\,dv. \tag{9.10}$$

Граница $\partial\mathscr{L}$ области $\mathscr{L}$ состоит из поверхностей

$$\partial\mathscr{L}_1 \equiv \mathscr{N}_1 \cap J^-(\mathscr{I}^+, \overline{\mathscr{M}}), \qquad \partial\mathscr{L}_2 \equiv \mathscr{N}_2 \cap J^-(\mathscr{I}^+, \overline{\mathscr{M}}),$$

куска $\partial\mathscr{L}_3$ поверхности $J^-(\mathscr{I}^+, \overline{\mathscr{M}})$ между $\mathscr{N}_1$ и $\mathscr{N}_2$ и куска $\partial\mathscr{L}_4$ поверхности $\mathscr{I}^-$ между $\mathscr{N}_1$ и $\mathscr{N}_2$. Поверхностный интеграл по $\partial\mathscr{L}_1$ равен поверхностному интегралу по $\partial\mathscr{L}_2$, взятому с обратным знаком, так как эти поверхности переводятся одна в другую изометрией θ_1.

Вблизи $\mathscr{I}^-$ функция $f = -1 + (2m/r) + O(r^{-2})$ и $\omega^a\omega_a = O(r^{-6})$, где r — некоторая подходящая радиальная координата. Поэтому поверхностный интеграл по $\partial\mathscr{L}_4$ на $\mathscr{I}^-$ равен

нулю. Допустим теперь, что вектор K^a времениподобен всюду на $\mathscr{L}$ и изотропен на горизонте. Тогда ω^a, будучи ортогональным к $\mathbf{K}$, пространственноподобен всюду в $\mathscr{L}$. Отсюда, если $\omega \neq 0$, т. е. решение нестатично, последний член в правой части (9.10) будет отрицательным. Если рассматриваемое пространство пустое и интеграл по $\partial\mathscr{L}_3$ равен нулю, то это ведет к противоречию.

Чтобы вычислить этот интеграл, нужно прибегнуть к предельному переходу. Пусть z — функция на поверхности $\mathscr{N}_1$, такая, что на горизонте она равна нулю, но ее градиент в $\mathscr{N}_1$ на горизонте отличен от нуля. Функцию z можно задать на $\overline{\mathscr{L}}$ условием $z_{;a}K^a = 0$. Градиент z можно представить в виде

$$z_{;a} = \bar{x}_{;b}z^{;b}(K_a + fR_a),$$

где R^a — вектор, касательный к поверхностям $\{\bar{x} = \text{const}\}$ и нормированный так, что $R^aK_a = -1$. Возьмем теперь $\int K^{a;b}q_a\,d\sigma_b$ по поверхности $\{z = \text{const}\}$ между $\mathscr{N}_1$ и $\mathscr{N}_2$. Тогда $\mathrm{d}\sigma_b = \mathrm{d}\sigma z_{;b}$, где $\mathrm{d}\sigma$ — некоторая непрерывная мера. Итак,

$$\int K^{a;b}q_a\,d\sigma_b = \int \left(\frac{1}{2}\bar{x}_{;a}(f)^{;a} - \bar{x}_{;a}K^a{}_{;b}R^b f + \frac{1}{2}f_{;b}R^b\right)\bar{x}_{;b}z^{;b}\,d\sigma.$$

Поскольку горизонт представляет собой поверхность $f = 0$ и поскольку K^a направлен вдоль изотропных образующих горизонта, вектор $f_{;a}$ на горизонте пропорционален K^a. Отсюда

$$\int\limits_{\partial\mathscr{L}_3} K^{a;b}q_a\,\mathrm{d}\sigma_b = 0.$$

Это приводит к противоречию, которое доказывает, что K^a должен быть где-то в $\overline{\mathscr{L}}$ пространственноподобным, если пространство пустое. □

Предложение 9.3.6

Пусть $(\mathscr{M}, \mathbf{g})$ — стационарное нестатическое регулярно предсказуемое пространство-время, в котором эргосфера имеет пересечение с $J^-(\mathscr{I}^+, \overline{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \overline{\mathscr{M}})$. Тогда существует однопараметрическая циклическая группа изометрии $\tilde{\theta}_\phi$ $(0 \leqslant \phi \leqslant 2\pi)$ этого пространства, которая коммутирует с θ_t и орбиты которой пространственноподобны вблизи $\mathscr{I}^+$ и $\mathscr{I}^-$.

Пусть $\mathscr{L}_1$ есть односвязная компонента множества $J^-(\mathscr{I}^+, \overline{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \overline{\mathscr{M}})$, и пусть $\mathscr{G}_1$ — фактор-множество $\mathscr{L}_1$ по ее образующим. Тогда орбитами изометрии θ_t в горизонте $\mathscr{L}_1$ будут винтовые линии, многократно пересекающие одни и те же образующие. Пусть при некотором $t_1 > 0$ θ_{t_1} есть один оборот

в $\mathscr{G}_1$. Тогда, если $p \in \mathscr{L}_1$, то $\theta_{t_1}(p)$ будет лежать на той же образующей $\mathscr{L}_1$ и притом в будущем точки p, так как

$$x|_{\theta_{t_1}(p)} = x|_p + t_1.$$

Мы можем выбрать направленный в будущее изотропный вектор $\mathbf{Y}_1$ так, чтобы он был касателен к образующим, и так, чтобы он обладал следующими масштабными свойствами:

1) $Y_{1a;b}Y_1{}^b = 2\varepsilon Y_{1a}$, где $\varepsilon_{;a}Y_1{}^a = 0$,

2) если v — параметр вдоль этих образующих, такой, что $\mathbf{Y}_1 = \partial/\partial v$, то

$$v|_{\theta_{t_1}(p)} = v|_p + t_1.$$

Векторное поле $\mathbf{Y}_1$, определенное таким образом, инвариантно относительно изометрии θ_t, т. е. $L_{\mathbf{K}}\mathbf{Y}_1 = 0$. Затем мы можем задать в $\mathscr{L}_1$ пространственноподобное векторное поле $\mathbf{Y}_3 \equiv \mathbf{K} - \mathbf{Y}_1$; $L_{\mathbf{K}}\mathbf{Y}_3 = 0$ и $L_{\mathbf{Y}_1}\mathbf{Y}_3 = 0$ (отметим, что $\mathbf{Y}_3$ — не единичный вектор, и в действительности он будет равен нулю на образующих γ_1 и γ_2, соответствующих полюсам $\mathscr{G}_1$). Интегральные кривые векторного поля $\mathbf{Y}_3$ в $\mathscr{L}_1$ будут представлять собой окружности, вырождающиеся на γ_1 и γ_2 в точки.

Пусть μ — некоторая кривая в $\mathscr{L}_1$ от γ_1 до γ_2, ортогональная к $\mathbf{Y}_1$ и $\mathbf{Y}_3$, причем такая, что орбиты $\mathbf{Y}_3$, пересекающие μ, образуют в $\mathscr{L}_1$ гладкую пространственноподобную 2-поверхность $\mathscr{P}$. Пусть $\mathscr{P}(v)$ — семейство пространственноподобных 2-поверхностей в $\mathscr{L}_1$, полученных сдвигом каждой точки поверхности $\mathscr{P}$ вдоль образующих $\mathscr{L}_1$ на параметрическое расстояние v. $\mathscr{P}(v)$ совпадает также с $\theta_v(\mathscr{P})$. Введем $\mathbf{Y}_2$ — другой изотропный, ортогональный к $\mathscr{P}(v)$ вектор, нормированный так, что $Y_1{}^a Y_{2a} = -1$ (рис. 61); тогда $L_{\mathbf{K}}\mathbf{Y}_2 = 0$.

Пусть $\mathbf{Y}_4$ — пространственноподобный вектор на μ, касательный к μ. Тогда мы можем определить $\mathbf{Y}_4$ на $\mathscr{L}_1$, увлекая его вдоль векторных полей $\mathbf{K}$ и $\mathbf{Y}_1$, т. е. $L_{\mathbf{K}}\mathbf{Y}_4 = 0 = L_{\mathbf{Y}_1}\mathbf{Y}_4$. (Эти равенства совместны, поскольку $L_{\mathbf{K}}\mathbf{Y}_1 = 0$.) $\mathbf{Y}_4$ будет ортогонален к $\mathbf{Y}_1$ на $\mathscr{L}_1$ ввиду того, что $L_{\mathbf{K}}(Y_4{}^a g_{ab} Y_1{}^b) = 0$; кроме того,

$$(Y_4{}^a Y_{1a})_{;b} Y_1{}^b = Y_1{}^a{}_{;b} Y_4{}^b Y_{1a} + Y_1{}^a{}_{;b} Y_{4a} Y_1{}^b.$$

Первый член равен нулю ввиду изотропности $\mathbf{Y}_1$, а второй член равен $2\varepsilon Y_{1a} Y_4{}^a$. Таким образом, если первоначально $Y_{1a}Y_4{}^a = 0$, то эта величина остается равной нулю. $\mathbf{Y}_4$ ортогонален к $\mathbf{Y}_2$ на $\mathscr{L}_1$, поскольку он лежит на поверхности $\mathscr{P}(v)$, а $\mathbf{Y}_2$ нормален к этой поверхности. На $\mathscr{L}_1$ он будет также ортогонален к $\mathbf{Y}_3$, потому что $L_{\mathbf{K}}(Y_3{}^a g_{ab} Y_4{}^b) = 0$ и

$$(Y_3{}^a Y_{4a})_{;b} Y_1{}^b = Y_1{}^a{}_{;b} Y_3{}^b Y_{4a} + Y_1{}^a{}_{;b} Y_4{}^b Y_{3a} = 0,$$

силу равенства $Y_{1a;b}\hat{h}^{ac}\hat{h}^{bd} = 0$.

В некоторой окрестности $\mathscr{L}_1$ будет существовать единственная изотропная геодезическая λ, ортогональная к поверхности $\mathscr{P}(v)$, проходящей через данную точку r. При этом для точки r

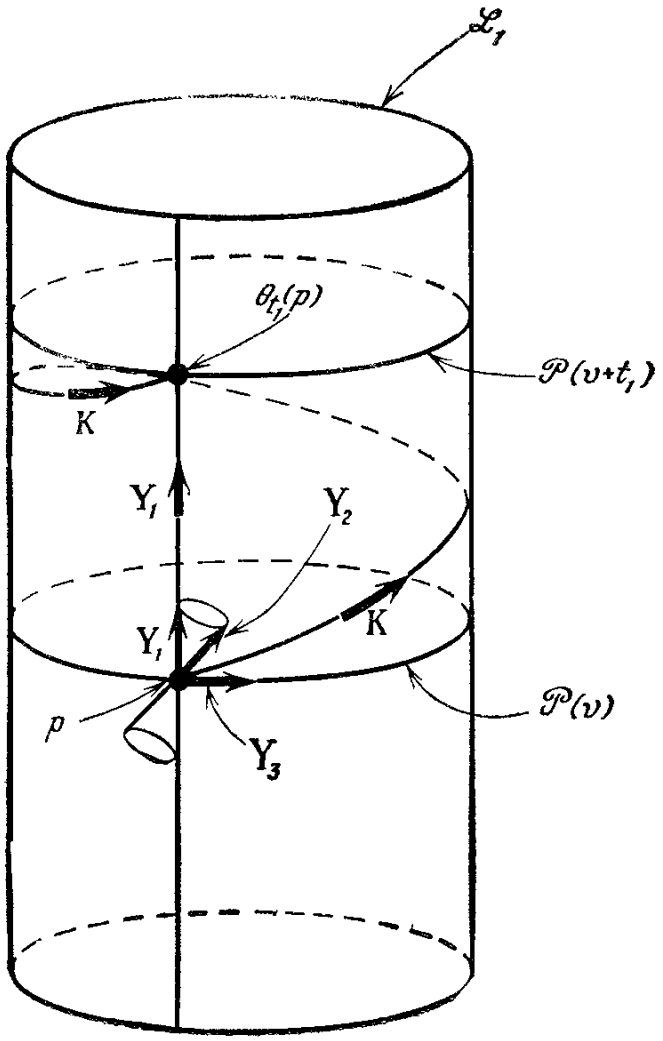


Рис. 61. Изометрия θ_{t_1} сдвигает точку p и поверхность $\mathscr{P}(v)$ в точку $\theta_{t_1}(p)$ и поверхность $\mathscr{P}(v+t_1)$ на горизонте $\mathscr{L}_1$. Вектор $\mathbf{Y}_1$ касателен к изотропной геодезической образующей $\mathscr{L}_1$, $\mathbf{Y}_2$ — изотропный вектор, ортогональный $\mathscr{P}(v)$ и $\mathbf{Y}_3$, лежит в $\mathscr{P}(v)$. $\mathbf{K}$ — поле векторов Киллинга на $\mathscr{L}_1$, порождающее группу изометрии θ_t.

можно ввести координаты (v, w, θ, ϕ), где w — аффинное расстояние вдоль μ, определяемое по $\mathbf{Y}_2$, а значения (v, θ, ϕ) взяты в $\mu \cap \mathscr{L}_1$. Здесь θ и ϕ — такие сферические полярные координаты для образующих поверхности $\mathscr{L}_1$, при которых $Y_3{}^a\theta_{,a} = 0$, $Y_4{}^a\phi_{,a} = 0$. (Другими словами, мы выберем на $\mathscr{L}_1$ $\mathbf{Y}_3 = (2\pi/t_1)\partial/\partial\phi$ и $\mathbf{Y}_4 = \partial/\partial\theta$.) Возьмем базис $\{\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \mathbf{Y}_4\}$, параллельно перенесенный вдоль изотропных геодезических с касательным вектором $\mathbf{Y}_2$. Тогда $\mathbf{Y}_2 = \partial/\partial w$. Определим вектор $\hat{\mathbf{K}}$:

$\hat{\mathbf{K}} = \partial/\partial v$. Это значит, что $L_{\mathbf{Y}_2}\hat{\mathbf{K}} = 0$. Введем вектор Z^a:

$$Z^a = \frac{1}{\sqrt{2}}\left\{\frac{Y_3{}^a}{(Y_3{}^b Y_{3b})^{1/2}} + i\,\frac{Y_4{}^a}{(Y_4{}^b Y_{4b})^{1/2}}\right\}.$$

Тогда

$$Z^a Z_a = 0, \quad Z^a \bar{Z}_a = 1, \quad \bar{Z}^a \bar{Z}_a = 0,$$

где чертой сверху обозначено комплексное сопряжение.

На $\mathscr{L}_1$ можно ввести следующее семейство тензорных полей $\{\mathbf{g}_n\}$:

$$\mathbf{g}_0 = \mathbf{g} \quad \text{и} \quad \mathbf{g}_n = L_{\mathbf{Y}_2}\underbrace{(L_{\mathbf{Y}_2}(\ldots(L_{\mathbf{Y}_2}\mathbf{g})\ldots))}_{n \text{ раз}}.$$

Во введенных выше координатах $g_{nab} = \partial^n(g_{ab})/\partial w^n$. Поскольку рассматриваемое решение аналитично, оно полностью определяется семейством $\mathbf{g}_n$ на $\mathscr{L}_1$. Мы покажем, что на $\mathscr{L}_1$ производные Ли всех $\mathbf{g}_n$ по $\hat{\mathbf{K}}$ равны нулю. Из этого следует, что производные Ли полей $\mathbf{g}_n$ относительно $\tilde{\mathbf{K}} = \hat{\mathbf{K}} - \mathbf{K}$ также равны нулю. Отсюда видно, что пространство-время допускает однопараметрическую группу $\tilde{\theta}_\phi$, порожденную векторным полем $\tilde{\mathbf{K}}$. Для простоты мы ограничимся лишь случаем пустого пространства, но аналогичные аргументы имеют силу и в присутствии материальных полей, подобных электромагнитному и скалярному, которые подчиняются хорошо определенным гиперболическим уравнениям.

При нашем выборе координат компоненты $L_{\hat{\mathbf{K}}}\mathbf{g}$ являются частными производными по v координатных компонент g_{ab}. Последние постоянны на $\mathscr{L}_1$, так что $L_{\hat{\mathbf{K}}}\mathbf{g}\,|_{\mathscr{L}_1} = 0$. Мы покажем ниже, что $L_{\hat{\mathbf{K}}}\mathbf{g}_1\,|_{\mathscr{L}_1} = 0$, и затем применим метод индукции. Предположим, что

$$L_{\hat{\mathbf{K}}}\mathbf{g}_n\,|_{\mathscr{L}_1} = 0, \quad n \geqslant 1.$$

Из самого построения базиса следует, что $L_{\hat{\mathbf{K}}}$ всех n-х ковариантных производных всех базисных векторов $\mathbf{Y}_1$, $\mathbf{Y}_2$, $\mathbf{Z}$, $\bar{\mathbf{Z}}$ равны нулю. Имеем

$$g_{n+1\,ab} = g_{n\,ab;c}Y_2{}^c + g_{n\,cb}Y_2{}^c{}_{;a} + g_{n\,ac}Y_2{}^c{}_{;b}.$$

Производные Ли по $\hat{\mathbf{K}}$ второго и третьего членов равны нулю. Первый член включает в себя ковариантные производные $\mathbf{Y}_2$ порядка $(n+1)$ и более низких порядков. Производная Ли по

$\hat{\mathbf{K}}$ всех членов низших порядков равна нулю. Члены, включающие $(n+1)$-е ковариантные производные, равны

$$(Y_{2a;bef\ldots ghc} + Y_{2b;aef\ldots ghc})\,Y_2{}^e Y_2{}^f \ldots Y_2{}^h Y_2{}^c =$$
$$= (Y_{2a;bc}Y_2{}^e + Y_{2b;ae}Y_2{}^e)_{;f\ldots ghc}\,Y_2{}^f \ldots Y_2{}^c +$$
$$+ (\text{члены более низкого порядка}) =$$
$$= ((Y_{2a;e}Y_2{}^e)_{;b} + R_{pabe}Y_2{}^p Y_2{}^e + (Y_{2b;e}Y_2{}^e)_{;a} + R_{pbae}Y_2{}^p Y_2{}^e)_{;f\ldots gh} \times$$
$$\times Y_2{}^f \ldots Y_2{}^c + (\text{члены более низкого порядка}).$$

Производные Ли по $\hat{\mathbf{K}}$ от этого выражения будут равны нулю, если равны нулю производные Ли по $\hat{\mathbf{K}}$ тензора Римана и его ковариантных производных до порядка $(n-1)$. Тогда получим, что $L_{\hat{\mathbf{K}}}\mathbf{g}_{n+1}|_{\mathscr{L}_1} = 0$.

Чтобы показать, что производные Ли по $\hat{\mathbf{K}}$ тензора $\mathbf{g}_1$ и ковариантных производных тензора Римана равны нулю, удобно воспользоваться некоторыми обозначениями, введенными Ньюменом и Пенроузом [113]. Они состоят в использовании псевдоортонормированного базиса с двумя пространственноподобными векторами $\mathbf{Y}_3$ и $\mathbf{Y}_4$, скомбинированными в один комплексный изотропный вектор $\mathbf{Z}$, во введении для каждой компоненты связности и тензора кривизны особого символа и в явной (без суммирования) записи всех тождеств Бианки и уравнений, определяющих тензор кривизны. Эти соотношения комбинируются в пары, так что они образуют вдвое меньшее число комплексных уравнений. Символы для связности определяются так:

$$\varkappa = Y_{1a;b}Z^a Y_1{}^b, \qquad \pi = -Y_{2a;b}\bar{Z}^a Y_1{}^b,$$
$$\rho = Y_{1a;b}Z^a \bar{Z}^b, \qquad \lambda = -Y_{2a;b}\bar{Z}^a \bar{Z}^b,$$
$$\sigma = Y_{1a;b}Z^a Z^b, \qquad \mu = -Y_{2a;b}\bar{Z}^a Z^b,$$
$$\tau = Y_{1a;b}Z^a Y_2{}^b, \qquad \nu = -Y_{2a;b}\bar{Z}^a Y_2{}^b,$$
$$\varepsilon = \frac{1}{2}(Y_{1a;b}Y_2{}^a Y_1{}^b - Z_{a;b}\bar{Z}^a Y_1{}^b), \qquad \alpha = \frac{1}{2}(Y_{1a;b}Y_2{}^a \bar{Z}^b - Z_{a;b}\bar{Z}^a \bar{Z}^b),$$
$$\beta = \frac{1}{2}(Y_{1a;b}Y_2{}^a Z^b - Z_{a;b}\bar{Z}^a Z^b), \qquad \gamma = \frac{1}{2}(Y_{1a;b}Y_2{}^b Y_2{}^b - Z_{a;b}\bar{Z}^a Y_2{}^b).$$

Символы для тензора Вейля:

$$\Psi_0 = -C_{abcd}Y_1{}^a Z^b Y_2{}^c Z^d,$$
$$\Psi_1 = -C_{abcd}Y_1{}^a Y_2{}^b Y_1{}^c Z^d,$$
$$\Psi_2 = -\frac{1}{2}C_{abcd}(Y_1{}^a Y_2{}^b Y_1{}^c Y_2{}^d - Y_1{}^a Y_2{}^b Z^c \bar{Z}^d),$$
$$\Psi_3 = C_{abcd}Y_1{}^a Y_2{}^b Y_2{}^c \bar{Z}^d,$$
$$\Psi_4 = -C_{abcd}Y_2{}^a \bar{Z}^b Y_2{}^c \bar{Z}^d.$$

Мы рассматриваем пустое пространство, и тензор Риччи равен нулю (т. е. в формализме Ньюмена — Пенроуза $\Phi_{AB} = 0 = \Lambda$). Поскольку базис параллельно переносится вдоль $\mathbf{Y}_2$, имеем $\nu = \gamma = \tau = 0$. Поскольку $\mathbf{Y}_2$ есть градиент координаты v, $\pi = \bar{\beta} + \alpha$ и $\mu = \bar{\mu}$. Более того, на $\mathscr{L}_1$ $\varkappa = \rho = \sigma = 0$, $\varepsilon = \bar{\varepsilon}$, $Y_1(\varepsilon) = 0$ и $\Psi_0 = 0$.

Нужные нам уравнения имеют вид

$$Y_1(\alpha) - \bar{Z}(\varepsilon) = (\rho + \bar{\varepsilon} - 2\varepsilon)\alpha + \beta\bar{\sigma} - \bar{\beta}\varepsilon - \varkappa\lambda + (\varepsilon + \rho)\pi, \quad (9.11а)$$

$$Y_1(\beta) - Z(\varepsilon) = (\alpha + \pi)\sigma + (\bar{\rho} - \bar{\varepsilon})\beta - \mu\varkappa - (\bar{\alpha} - \bar{\pi})\varepsilon + \Psi_1, \quad (9.11б)$$

$$Y_1(\lambda) - \bar{Z}(\pi) = \rho\lambda + \bar{\sigma}\mu + \pi^2 + (\alpha - \bar{\beta})\pi - (3\varepsilon - \bar{\varepsilon})\lambda; \quad (9.11в)$$

$$Y_1(\mu) - Z(\pi) = \bar{\rho}\mu + \sigma\lambda + \pi\bar{\pi} - (\varepsilon + \bar{\varepsilon})\mu - \pi(\bar{\alpha} - \beta) + \Psi_2, \quad (9.11г)$$

$$Z(\rho) - \bar{Z}(\sigma) = \rho(\bar{\alpha} + \beta) - \sigma(3\alpha - \bar{\beta}) - \Psi_1 \quad (9.11д)$$

[эти уравнения получаются из уравнений Ньюмена — Пенроуза (4.2)] и

$$Y_1(\Psi_1) - \bar{Z}(\Psi_0) = -3\varkappa\Psi_2 + (2\varepsilon + 4\rho)\Psi_1 - (-\pi + 4\alpha)\Psi_0, \quad (9.12а)$$

$$Y_1(\Psi_2) - \bar{Z}(\Psi_1) = -2\varkappa\Psi_3 + 3\rho\Psi_2 - (-2\pi + 2\alpha)\Psi_1 - \lambda\Psi_0, \quad (9.12б)$$

$$Y_1(\Psi_3) - \bar{Z}(\Psi_2) = -\varkappa\Psi_4 - (2\varepsilon - 2\rho)\Psi_3 + 3\pi\Psi_2 - 2\lambda\Psi_1, \quad (9.12в)$$

$$Y_1(\Psi_4) - \bar{Z}(\Psi_3) = -(4\varepsilon - \rho)\Psi_4 + (4\pi + 2\alpha)\Psi_3 - 3\lambda\Psi_2, \quad (9.12г)$$

$$Y_2(\Psi_0) - Z(\Psi_1) = -\mu\Psi_0 - 2\beta\Psi_1 + 3\sigma\Psi_2 \quad (9.12д)$$

[эти уравнения получаются из уравнений Ньюмена — Пенроуза (4.5)].

Из (9.11д) $\Psi_1 = 0$ на $\mathscr{L}_1$. Тогда из (9.12б) $Y_1(\Psi_2) = \hat{K}(\Psi_2) = 0$ на $\mathscr{L}_1$. Складывая (9.11а) с уравнением, комплексно-сопряженным (9.11б), получим

$$Y_1(\pi) = Y_1(\alpha + \bar{\beta}) =$$

$$= \bar{Z}(\varepsilon) + Z(\bar{\varepsilon}) + 2\pi\rho + 2\bar{\pi}\bar{\sigma} - \pi(\varepsilon - \bar{\varepsilon}) - \varkappa\lambda - \bar{\varkappa}\bar{\mu} + \overline{\Psi}_1;$$

отсюда на $\mathscr{L}_1$

$$Y_1(\pi) = \bar{Z}(\varepsilon) + Z(\bar{\varepsilon}).$$

Следовательно, на $\mathscr{L}_1$ $Y_1(Y_1(\pi)) = Y_1(\bar{Z}(\varepsilon) + Z(\bar{\varepsilon}))$. Но на $\mathscr{L}_1$ $L_{\mathbf{Y}_1}\bar{\mathbf{Z}} = 0$ и $Y_1(\varepsilon) = 0$. Поэтому на $\mathscr{L}_1$ $Y_1(Y_1(\pi)) = 0$. Отсюда видно, что на $\mathscr{L}_1$ $\pi = A + Bv$, где A и B — постоянные вдоль образующей горизонта $\mathscr{L}_1$. Но $\pi|_p = \pi|_{\theta_{t_1}(p)}$; следовательно, $\pi = \text{const}$ вдоль образующих $\mathscr{L}_1$. Вычитая из (9.11а) комплексно-сопряженное выражение (9.11б), находим, что $(\alpha - \bar{\beta}) = \text{const}$ вдоль указанных образующих.

Затем с помощью подобных же рассуждений относительно (9.11в) и (9.11г) можно показать, что вдоль образующих поверхности $\mathscr{L}_1$ μ и λ постоянны. Поскольку π, μ и λ задают ковариантную производную Y_2, отсюда следует, что на $\mathscr{L}_1$ $L_{\hat{\mathbf{K}}} Y_2{}^a{}_{;b} = 0$, т. е. что $L_{\hat{\mathbf{K}}} \mathbf{g}_1 |_{\mathscr{L}} = 0$.

Таким же способом на основе (9.12в) и (9.12г) доказывается, что на $\mathscr{L}_1$ $Y_1(\Psi_3) = Y_1(\Psi_4) = 0$. Отсюда следует, что на $\mathscr{L}_1$ $L_{\hat{\mathbf{K}}} R_{abcd} = 0$ и соответственно равны нулю производные Ли по $\hat{\mathbf{K}}$ вторых производных базисных векторов. В частности, действие $\mathbf{Y}_1\mathbf{Y}_2$ на любую из компонент связности дает нуль.

Из (9.12д) $\hat{K}(Y_2(\Psi_0)) = Y_1 Y_2(\Psi_0) = 0$ на $\mathscr{L}_1$. Подействуем оператором $\mathbf{Y}_1\mathbf{Y}_2$ на (9.12а). Коммутатор $\mathbf{Y}_1\mathbf{Y}_2 - \mathbf{Y}_2\mathbf{Y}_1$ содержит только первые ковариантные производные базисных векторов. Таким образом,

$$L_{\hat{\mathbf{K}}}(\mathbf{Y}_1\mathbf{Y}_2 - \mathbf{Y}_2\mathbf{Y}_1) = 0 \quad \text{на} \quad \mathscr{L}_1.$$

Отсюда по аналогии с предыдущим доказательством получим, что

$$\hat{K}(Y_2(\Psi_1)) = Y_1(Y_2(\Psi_1)) = 0 \quad \text{на} \quad \mathscr{L}_1.$$

Повторим наши рассуждения в применении к (9.10б — 9.10г), чтобы показать, что $\hat{K}(Y_2(\Psi_2)) = \hat{K}(Y_2(\Psi_3)) = \hat{K}(Y_2(\Psi_4)) = 0$ на $\mathscr{L}_1$. Отсюда видно, что производные Ли относительно $\hat{\mathbf{K}}$ первых ковариантных производных тензора Римана равны нулю. Затем, следуя той же схеме, можно показать, что $\hat{K}(Y_2(Y_2(\Psi_0)))|_{\mathscr{L}_1} = 0$ и т. д. □

Предложение 9.3.7

Пусть $(\mathscr{M}, \mathbf{g})$ — пространство-время, допускающее 2-параметрическую абелеву группу изометрии с векторами Киллинга ξ_1 и ξ_2. Пусть $\mathscr{V} \subset \mathscr{M}$ — связное открытое множество, и пусть $\omega_{ab} = \xi_{1[a}\xi_{2b]}$. Тогда, если:

а) $\omega_{ab} R^b{}_c \eta^{cdef} \omega_{ef} = 0$ на $\mathscr{V}$,

б) $\omega_{ab} = 0$ в некоторой точке из $\mathscr{V}$, то

$$\omega_{[ab;c}\omega_{d]e} = 0 \quad \text{на} \quad \mathscr{V}.$$

Пусть ${}_1\chi = \xi_{1a;b}\omega_{cd}\eta^{abcd}$ и ${}_2\chi = \xi_{2a;b}\omega_{cd}\eta^{abcd}$. Тогда

$$\eta^{abcd}\,{}_1\chi = -4!\xi_1{}^{[a;b}\xi_1{}^{c}\xi_2{}^{d]} =$$
$$= 3!\xi_1{}^{d}\xi_2{}^{[a}\xi_1{}^{b;c]} - 3!\xi_2{}^{d}\xi_1{}^{[a}\xi_1{}^{b;c]} - 2 \times 3!\xi_1{}^{[a}\xi_2{}^{b}\xi_1{}^{c];d}.$$

Отсюда

$$(3!)^{-1}\eta^{abcd}{}_{1}\chi_{;d} = \xi_1{}^{d}{}_{;d}\xi_2{}^{[a}\xi_1{}^{b;c]} + \xi_1{}^{d}\xi_2{}^{[a}{}_{;d}\xi_1{}^{b;c]} +$$
$$+ \xi_1{}^{d}\xi_2{}^{[a}\xi_1{}^{b;c]}{}_{;d} - \xi_2{}^{d}{}_{;d}\xi_1{}^{[a}\xi_1{}^{b;c]} - \xi_2{}^{d}\xi_1{}^{[a}{}_{;d}\xi_1{}^{b;c]} - \xi_2{}^{d}\xi_1{}^{[a}\xi_1{}^{b;c]}{}_{;d} -$$
$$- 2\xi_1{}^{[a}{}_{;d}\xi_2{}^{b}\xi_1{}^{c];d} - 2\xi_1{}^{[a}\xi_2{}^{b}{}_{;d}\xi_1{}^{c];d} - 2\xi_1{}^{[a}\xi_2{}^{b}\xi_1{}^{c];d}{}_{;d}. \quad (9.13)$$

Первый и четвертый члены исчезают, поскольку ξ_1 и ξ_2 — векторы Киллинга; второй и пятый члены взаимно уничтожаются, поскольку ξ_1 и ξ_2 коммутируют. $L_{\xi_1}\xi_{1a;b} = 0$ в силу того, что ξ_1 — вектор Киллинга, и потому равен нулю третий член. Аналогично $L_{\xi_2}\xi_{1a;b} = 0$ вследствие того, что ξ_2 — вектор Киллинга, коммутирующий с ξ_1. Из этого вытекает, что сумма шестого и восьмого членов равна нулю. Седьмой член исчезает в силу симметрии $\xi_1{}^{a}{}_{;d}\xi_1{}^{c;d}$, и соотношение $\xi^{a;d}{}_{;d} = -R^{a}{}_{b}\xi^{b}$, которому удовлетворяет всякий вектор Киллинга. Таким образом, равенство (9.13) приводит к виду

$$\eta^{abcd}{}_{(1)}\chi_{;d} = 2\cdot 3!\xi_1{}^{[a}\xi_2{}^{b}R^{c]}{}_{d}\xi_1{}^{d}.$$

По условию (а) правая часть обращается в нуль на $\mathscr{V}$. Следовательно, ${}_{1}\chi = \text{const}$ на $\mathscr{V}$; на самом деле ${}_{1}\chi = 0$ на $\mathscr{V}$, поскольку эта величина обращается в нуль при $w_{ab} = 0$. Аналогично и ${}_{2}\chi = 0$ на $\mathscr{V}$. Равенство нулю ${}_{1}\chi$ и ${}_{2}\chi$ необходимо и достаточно для того, чтобы

$$w_{[ab;c}w_{d]e} = 0. \qquad \square$$

Предложение 9.3.8.

Пусть $(\mathscr{M}, \mathbf{g})$ — стационарное аксиально симметричное регулярно предсказуемое пространство-время, в котором $w_{[ab;c}w_{d]e} = 0$, где $w_{ab} \equiv K_{[a}\tilde{K}_{b]}$. Тогда $h \equiv w_{ab}w^{ab} < 0$ в любой точке, лежащей вне оси $\tilde{\mathbf{K}} = 0$ внешней области $J^{+}(\mathscr{I}^{-}, \bar{\mathscr{M}}) \cap J^{-}(\mathscr{I}^{+}, \bar{\mathscr{M}})$. На горизонтах $\dot{J}^{-}(\mathscr{I}^{+}, \bar{\mathscr{M}}) \cap J^{+}(\mathscr{I}^{-}, \bar{\mathscr{M}})$ и $\dot{J}^{+}(\mathscr{I}^{-}, \bar{\mathscr{M}}) \cap J^{-}(\mathscr{I}^{+}, \bar{\mathscr{M}})$ $h = 0$, но w_{ab} нигде не обращается в нуль, кроме оси симметрии.

По предложению 9.3.3 вектор K^{a} отличен от нуля в $J^{+}(\mathscr{I}^{-}, \bar{\mathscr{M}}) \cap \bar{J}^{-}(\mathscr{I}^{+}, \bar{\mathscr{M}})$. Пусть λ будет окружностью S^{1}, которая является ненулевой интегральной кривой векторного поля $\tilde{\mathbf{K}}$ в $J^{+}(\mathscr{I}^{-}, \bar{\mathscr{M}}) \cap J^{-}(\mathscr{I}^{+}, \bar{\mathscr{M}})$. Действием изометрии θ_t можно переместить λ в $D^{+}(\mathscr{S})$. Поскольку в $D^{+}(\mathscr{S})$ нет непространственноподобных замкнутых кривых, λ должна быть пространственноподобной кривой, а потому и вектор $\tilde{K}^{a}$ должен быть пространственноподобным всюду в

$$J^{+}(\mathscr{I}^{-}, \bar{\mathscr{M}}) \cap \bar{J}^{-}(\mathscr{I}^{+}, \bar{\mathscr{M}}),$$

за исключением оси, где он равен нулю. Допустим, что существует некоторая точка p, в которой $\tilde{K}^a$ и K^a оба отличны от нуля и одинаково направлены. Поскольку они коммутируют, интегральные кривые $\tilde{K}^a$ и K^a, проходящие через p, должны совпасть. Однако интегральные кривые $\tilde{K}^a$ замкнуты, а интегральные кривые K^a — нет. Следовательно, $\tilde{K}^a$ и K^a там, где они отличны от нуля, линейно независимы. Поэтому $\omega_{ab} \neq 0$ всюду в $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, за исключением оси.

Ось представляет собой двумерную поверхность. Пусть $\mathscr{Y}$ есть множество $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, из которого исключена ось, и пусть $\mathscr{Z}$ есть фактор-множество $\mathscr{Y}$ по $\tilde{\theta}_\phi$. Поскольку интегральные кривые K^a замкнуты и пространственноподобны в $\mathscr{Y}$, фактор-множество $\mathscr{Z}$ будет хаусдорфовым многообразием. На $\mathscr{Z}$ можно ввести лоренцеву метрику $\tilde{h}_{ab} = g_{ab} - (\tilde{K}^c \tilde{K}_c)^{-1} \tilde{K}_a \tilde{K}_b$, спроектировать с ее помощью вектор Киллинга K^a и получить в $\mathscr{Z}$ ненулевое векторное поле $\tilde{h}_{ab} K^b$, которое будет векторным полем Киллинга для метрики $\tilde{h}_{ab}$. Условие $\omega_{[ab;c}\omega_{d]e} = 0$ в $\mathscr{M}$ означает, что в $\mathscr{Z}$ $(K^b \tilde{h}_{b[c})_{|d}\, \tilde{h}_{e]f} K^f = 0$, где вертикальной чертой обозначена ковариантная производная относительно $\tilde{h}$. Оно как раз и означает, что на $\mathscr{Z}$ должна существовать функция ξ такая, что $K^b \tilde{h}_{ba} = -\alpha \xi_{|a}$. Далее будем рассуждать, как при доказательстве предложения 9.3.4. Покажем, что при выполнении условия $K_a K_b \tilde{h}^{ab} = 0$ в точке $p \in \mathscr{Z}$ поверхность $\xi = \xi|_p$ будет поверхностью в $\mathscr{Z}$, изотропной относительно метрики $\tilde{h}$. Функция ξ на $\mathscr{Z}$ индуцирует функцию ξ на $\mathscr{Y}$, обладающую свойством: $\xi_{;a} K^a = 0$. Следовательно, $\xi = \xi|_p$ будет поверхностью в $\mathscr{M}$, изотропной относительно метрики $\mathbf{g}$.

Допустим, что p соответствует интегральной кривой λ векторного поля $\tilde{K}^a$, не лежащей в $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Пусть $q \in \mathscr{M}$ — точка на λ. Тогда должна найтись направленная в будущее времениподобная кривая $\gamma(v)$ от $\mathscr{I}^-$ через q до $\mathscr{I}^+$. Если эта кривая пересечет ось, ее можно будет слегка деформировать, чтобы избежать такого пересечения. Отсюда можно прийти к противоречию, аналогичному противоречию при доказательстве предложения 9.3.4.

Глава 10

Начальная сингулярность во Вселенной

Расширение Вселенной во многих отношениях подобно коллапсу звезды, если не считать того, что направление времени при расширении обратное. В этой главе мы покажем, что условия теорем 2 и 3, по-видимому, выполнены, и это указывает на наличие сингулярности в начале современной фазы расширения Вселенной; в связи с этим мы рассмотрим проявления пространственно-временных сингулярностей.

В разд. 10.1 мы покажем, что при условии частичной термализации микроволнового фонового излучения во Вселенной в результате рассеяния существуют направленные в прошлое замкнутые ловушечные поверхности; альтернативным условием их существования является выполнение принципа Коперника, согласно которому наше положение во Вселенной ничем не выделено. В разд. 10.2 мы рассмотрим возможную природу этой сингулярности и несостоятельность в самой сингулярности современной физической теории.

10.1 Расширение Вселенной

В разд 9.1 мы показали, что многие звезды в конце концов будут коллапсировать и при этом возникнут замкнутые ловушечные поверхности. Переходя к большим масштабам, можно рассматривать расширение Вселенной как обращенный во времени коллапс. Поэтому можно надеяться, что в космологическом масштабе условия теоремы 2 будут выполнены при обратном направлении времени, если материи во Вселенной достаточно для того, чтобы возникли замкнутые ловушечные поверхности и если она в некотором смысле достаточно симметрична. Мы приведем два довода в пользу того, что материи действительно достаточно. Они оба основаны на данных наблюдений микроволнового фонового излучения, но исходные предпосылки несколько различны.

Наблюдения в радиодиапазоне между 20 см и 1 мм указывают на существование фонового излучения, спектр которого (представленный на рис. 62, *а*), по-видимому, близок к спектру излучения абсолютно черного тела при температуре 2,7 К (см.,

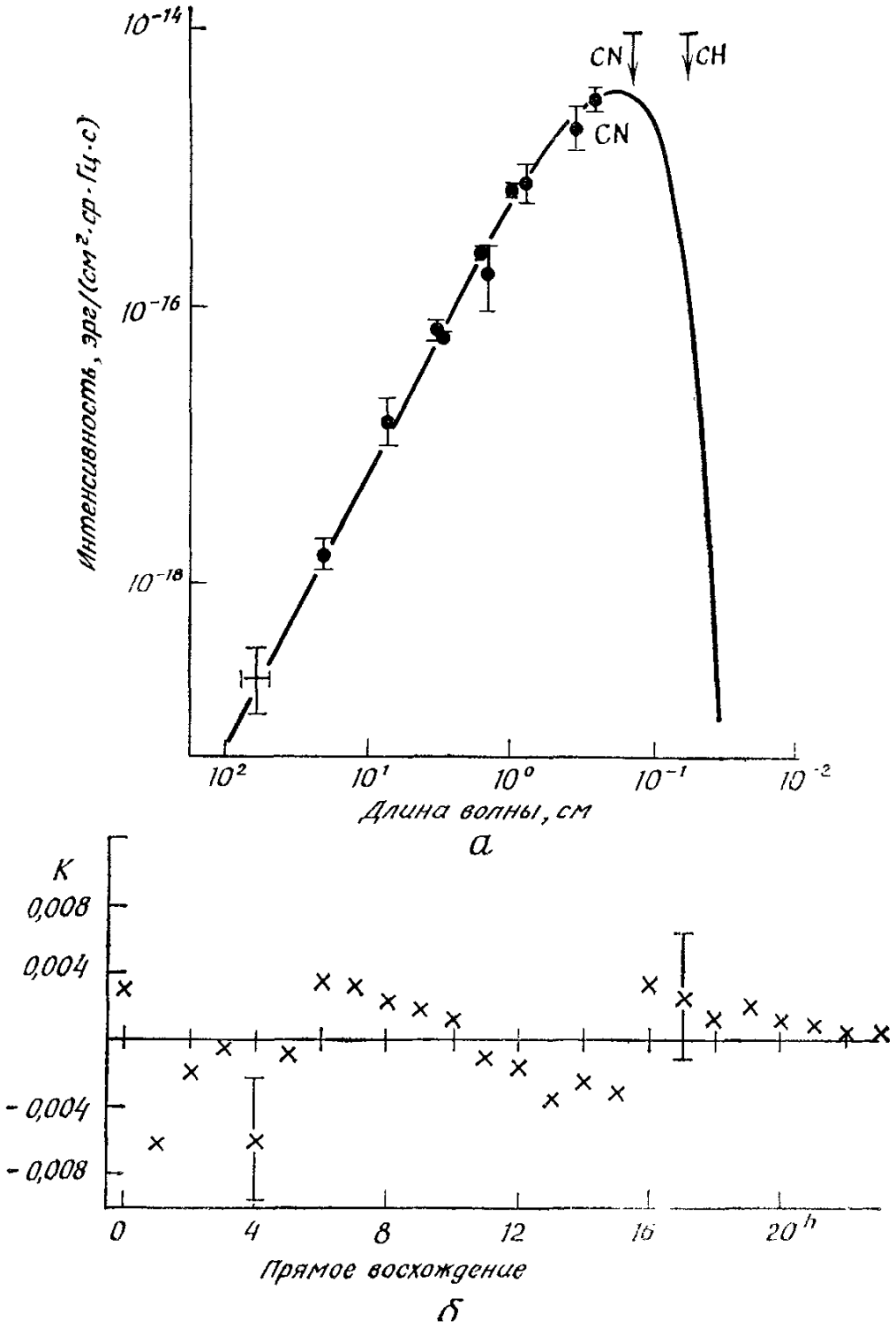


Рис. 62. *а* — спектр микроволнового фонового излучения. Точки соответствуют наблюдаемым значениям «избыточного» фонового излучения. Сплошная линия представляет собой спектр Планка, соответствующий температуре 2,7 К. *б* — изотропия микроволнового фонового излучения. Показано распределение по температуре вдоль небесного экватора; для получения этих точек усреднены данные более чем за 2 года. (Из книги Д. У. Шамы «Современная космология», М., изд-во «Мир». 1973.)

например, [48]). Это фоновое излучение, по-видимому, изотропно в пределах 0,2% (рис. 62, *б*; см., например, [155] и ссылки, приведенные там). Высокая степень изотропии свидетельствует о том, что это излучение не может зарождаться в пределах нашей Галактики (Солнечная система расположена вне плоскости Галактики), а должно быть внегалактического происхождения. В этом диапазоне мы можем наблюдать дискретные источники, причем расстояния до некоторых из них, как это установлено по другим данным, порядка 10^{27} см; следовательно, мы знаем, что на указанных длинах волн Вселенная прозрачна вплоть до таких расстояний. Таким образом, излучение источников, находящихся на расстояниях больше 10^{27} см, должно было свободно распространяться по направлению к нам по крайней мере на таких расстояниях.

Возможные объяснения происхождения этого излучения таковы:

1) оно является излучением абсолютно черного тела, оставшимся от ранней горячей фазы Вселенной;

2) оно является результатом наложения очень большого числа весьма удаленных неразрешенных дискретных источников;

3) оно генерируется межгалактической пылью, которая термализует другие виды излучения (вероятно, в инфракрасной области).

Из этих объяснений наиболее приемлемым представляется первое. Второе кажется невероятным, так как, по-видимому, нет достаточного числа источников с нужным спектром, которые могли бы создать заметную долю наблюдаемого излучения в данном диапазоне частот. Далее, мелкомасштабная изотропия излучения означает, что число дискретных источников должно быть очень велико (порядка числа галактик), а большинство галактик, по-видимому, не имеет заметного излучения в этой области спектра. Третье объяснение также представляется маловероятным, поскольку концентрация межзвездных частиц, которая понадобилась бы для этого, слишком велика. Хотя первое объяснение кажется наиболее вероятным, мы все же не будем исходить из него, так как это означало бы, что мы заранее предполагаем существование ранней горячей фазы в эволюции Вселенной.

Первое объяснение включает в себя принцип Коперника, согласно которому наше местонахождение в пространстве-времени ничем не выделено. По нашей интерпретации из этого принципа следует, что микроволновое фоновое излучение должно выглядеть одинаково изотропным для любого наблюдателя, скорость которого относительно близлежащих галактик мала. Другими словами, мы предполагаем существование

расходящейся геодезической конгруэнции (расходящейся — потому, что галактики удаляются друг от друга, геодезической — потому что они движутся под действием одного лишь тяготения с единичным вектором 4-скорости V^a), представляющей усредненное движение галактик, относительно которых это микроволновое излучение выглядит почти изотропным. Из принципа Коперника следует также, что бо́льшая часть микроволнового излучения свободно движется к нам с больших расстояний ($\approx 3 \cdot 10^{27}$ см). Это видно из того, что вклад в фоновое излучение сферического слоя вокруг нас толщиной dr и радиуса r в определенном приближении не зависит от r: количество излучения, поступающего из этого слоя, будет пропорционально r^2, а уменьшение интенсивности — обратно пропорционально квадрату расстояния r^2. Так будет до тех пор, пока не станет заметным красное смещение источников, не будет сказываться эволюция источников или не станут значительными эффекты кривизны. Однако все эти явления проявятся лишь на расстояниях порядка радиуса Хаббла, $\sim 10^{28}$ см. Следовательно, подавляющая часть излучения будет свободно приходить к нам с расстояния $>10^{27}$ см. Из того факта, что, преодолев такое большое расстояние, излучение остается изотропным, мы можем заключить, что в больших масштабах метрика Вселенной близка к одной из метрик Робертсона — Уокера (разд. 5.3). Этот вывод следует из результатов Элерса, Герена и Сакса [44], к изложению которых мы переходим.

Микроволновое фоновое излучение можно описать функцией распределения $f(u, \mathbf{p})$ ($u \in \mathscr{M}$, $\mathbf{p} \in T_u$), заданной на изотропных векторах в $T(\mathscr{M})$ (их совокупность можно рассматривать как фазовое пространство фотонов). Если функция распределения строго изотропна для наблюдателя, движущегося с 4-скоростью V^a, то она будет иметь вид $f(u, E)$, где $E = -V^a p_a$. Поскольку это излучение распространяется свободно, функция f должна подчиняться уравнению Лиувилля в $T(\mathscr{M})$. Это означает, что функция f постоянна вдоль интегральных кривых горизонтального векторного поля $\mathbf{X}$, т. е. вдоль любой кривой $(u(v), \mathbf{p}(v))$, где $u(v)$ — изотропная геодезическая в $\mathscr{M}$ и $\mathbf{p} = \partial/\partial v$.

Поскольку $f(u, E) \geqslant 0$ и f должна стремиться к нулю при $E \to \infty$ (иначе плотность энергии излучения была бы бесконечной), должен существовать открытый интервал E, в котором $\partial f/\partial E \neq 0$. В этом интервале мы можем выразить E как функцию f: $E = g(u, f)$. Тогда уравнение Лиувилля означает, что на каждой изотропной геодезической

$$\frac{dE}{dv} = g_{;a} p^a, \tag{10.1}$$

где g рассматривается как функция на $\mathscr{M}$ при фиксированном f. Кроме того,

$$\frac{dE}{dv} = -\frac{d(V^a p_a)}{dv} = -V_{a;b}p^a p^b. \tag{10.2}$$

Можно разложить p^a на две составляющие: направленную вдоль V^a и ортогональную к V^a, т. е. $p = E(V^a + W^a)$, где $W^a W_a = 1$, $W^a V_a = 0$. Тогда из (10.1) и (10.2) следует, что при всех единичных векторах W^a, ортогональных к V^a, выполняется равенство

$$\frac{dg}{dt} + \frac{1}{3}\theta g + (g\dot{V}_a + g_{;a})W^a + g\sigma_{ab}W^a W^b = 0,$$

где dg/dt — скорость изменения g вдоль интегральных кривых $\mathbf{V}$. Отделяя сферические гармоники, имеем

$$\sigma_{ab} = 0, \tag{10.3а}$$

$$\dot{V}_a + (\ln g)_{;a} = \alpha V_a, \tag{10.3б}$$

$$\frac{1}{3}\theta = -\frac{d(\ln g)}{dt}. \tag{10.3в}$$

Поскольку мы приняли, что $\dot{V}_a = 0$, из (10.3б) следует, что вектор V_a ортогонален к поверхностям $\{g = \text{const}\}$, и это означает, что завихренность ω_{ab} равна нулю. В силу $\dot{V}^a = 0$ имеем $V_{[a;b]} = 0$. Следовательно, V_a можно записать как градиент функции t: $V_a = -t_{;a}$.

Тензор энергии-импульса излучения будет иметь вид

$$T_{ab} = \frac{4}{3}\mu_r V_a V_b + \frac{1}{3}\mu_r g_{ab},$$

где $\mu_r = \int fE^3\,dE$. Поскольку скорость галактик относительно интегральных кривых вектора V^a мала, их вклад в тензор энергии-импульса можно аппроксимировать тензором энергии-импульса непрерывной жидкости с плотностью μ_G, 4-скоростью V_a и пренебрежимо малым давлением. Тогда мы придем к выводу, что геометрия пространства-времени такая же, как в модели Робертсона — Уокера. Для этого заметим сначала, что

$$(V^a{}_{;b})_{;a} = \frac{1}{3}(\theta(\delta^a{}_b + V^a V_b))_{;a} = (V^a{}_{;a})_{;b} + R^{ca}{}_{ba}V_c = \theta_{;b} + R_{ba}V^a.$$

Умножая это уравнение на $h^b{}_c = g^b{}_c + V^b V_c$, находим

$$h^{bc}R_{ca}V^a = -\frac{2}{3}h^{bc}\theta_c.$$

Левая часть равна нулю в силу уравнений поля, и, следовательно, расхождение θ постоянно на поверхностях $t = \text{const}$

(которые являются также поверхностями постоянного g). По функции $\theta(t)$ из уравнения $\dot{S}/S = \frac{1}{3}\theta$ можно определить функцию $S(t)$. Тогда уравнение Райчаудхури (4.26) принимает форму

$$\frac{3\ddot{S}}{S} + 4\pi\mu - \Lambda = 0,$$

откуда следует, что $\mu = \mu_G + 2\mu_R$ тоже постоянно на поверхностях $\{t = \text{const}\}$. Из определения μ_R следует, что μ_G и μ_R по отдельности постоянны на этих поверхностях.

Из той части равенства (4.27), след которой равен нулю, видно, что $C_{abcd}V^bV^d = 0$. Уравнения Гаусса — Кодацци (разд. 2.7) приводят теперь к следующей формуле для тензора Риччи на 3-поверхностях $\{t = \text{const}\}$:

$$R^3{}_{ab} = h_a{}^c h_b{}^d R_{cd} + R_{acbd}V^cV^d + \theta\theta_{ab} + \theta_{ac} + \theta^c{}_b = \\ = 2h_{ab}\left(-\frac{1}{3}\theta^2 + 8\pi\mu + \Lambda\right).$$

На трехмерном многообразии тензор Римана полностью определяется тензором Риччи:

$$R^3{}_{abcd} = h_{ab}{}^e\left(-R^3{}_{ef} + \frac{1}{2}R^3 h_{ef}\right)\eta^f{}_{cd}.$$

Отсюда каждая 3-поверхность $\{t = \text{const}\}$ есть трехмерное пространство постоянной кривизны $K(t) = \frac{1}{3}\left(8\pi\mu + \Lambda - \frac{1}{3}\theta^2\right)$. Интегрирование уравнения Райчаудхури приводит к формуле

$$K(t) = \frac{1}{3}\left(8\pi\mu + \Lambda - \frac{3\dot{S}^2}{S^2}\right) = \frac{k}{S^2}, \qquad (10.4)$$

где k — постоянная. Нормируя S, можно положить $k = \geqslant 1, 0$ или -1. Рассматриваемое четырехмерное пространственно-временное многообразие есть ортогональное произведение этих 3-пространств и линии t. Итак, наша метрика может быть записана в сопутствующих координатах, как

$$ds^2 = -dt^2 + S^2(t)\,d\gamma^2,$$

где $d\gamma^2$ — метрика 3-пространства постоянной кривизны k, т. е. метрика Робертсона — Уокера (см. разд. 5.3).

Теперь мы покажем, что в любом пространстве Робертсона — Уокера с положительной плотностью энергии и $\Lambda = 0$ имеется замкнутая ловушечная поверхность в каждой 3-поверхности $\{t = \text{const}\}$. Чтобы убедиться в этом, представим $d\gamma^2$ в виде

$$d\gamma^2 = d\chi^2 + f^2(\chi)(d\theta^2 + \sin^2\theta\,d\phi^2),$$

где $f(\chi)$ равно $\sin\chi$, χ или $\operatorname{sh}\chi$ в зависимости от того, чему равно k: $+1$, 0 или -1. Рассмотрим 2-сферу $\mathcal{T}$ радиуса χ_0, рас-

положенную в поверхности $t = t_0$. Два семейства направленных в прошлое ортогональных к $\mathcal{T}$ изотропных геодезических будут пересекать поверхности $\{t = \text{const}\}$ по двум 2-сферам радиуса

$$\chi = \chi_0 \pm \int_{t_1}^{t} \frac{dt}{S(t)}. \tag{10.5}$$

Площадь поверхности 2-сферы радиуса χ равна $4\pi S^2(t) f^2(\chi)$. Таким образом, оба семейства изотропных геодезических будут сходящимися в прошлом, если при $t = t_0$ неравенство

$$\frac{d}{dt}(S^2(t) f(\chi)) > 0$$

выполняется для обоих значений χ из (10.5). Это произойдет при

$$\frac{\dot{S}(t_0)}{S_0(t_0)} > \pm \frac{f'(\chi_0)}{S(t_0) f(\chi_0)},$$

но, согласно (10.4), это неравенство выполняется, если

$$\left[\frac{8}{3}\pi\mu(t_0) S^2(t_0) - k\right]^{1/2} > \pm \frac{f'(\chi_0)}{f(\chi_0)},$$

что в свою очередь справедливо, если $S(t_0)\chi_0$ взять большим, чем $(3/8\pi\mu_0)^{1/2}$ при $k = 0$ или -1, и большим, чем $\min(\sqrt{3/8\pi\mu_0}, \pi/2)$ при $k = +1$.

Этот результат интуитивно можно интерпретировать как требование, чтобы в момент времени t_0 сфера координатного радиуса χ_0 содержала массу порядка $^4/_3 \pi\mu_0 S^3(t_0)\chi_0^3$ и, таким образом, была внутри своего радиуса Шварцшильда, если $S(t_0)\chi_0 < {}^8/_3 \pi\mu_0 S(t_0)^3 \chi_0^3$, т. е. если $S(t_0)\chi$ по порядку величины больше $\sqrt{3/8\pi\mu_0}$. Мы будем называть величину $\sqrt{3/8\pi\mu_0}$ *шварцшильдовой длиной* материи плотностью μ_0.

До сих пор мы полагали, что микроволновое излучение строго изотропно. Это, конечно, не так, поскольку Вселенная не является в точности пространством Робертсона — Уокера. Однако крупномасштабная структура Вселенной должна быть близка к структуре модели Робертсона — Уокера по крайней мере назад во времени, к моменту, когда микроволновое фоновое излучение было испущено или претерпело последнее рассеяние. (Отклонения микроволнового излучения от строгой изотропии можно использовать для оценки того, сколь велики отклонения от геометрии Робертсона — Уокера.) Для достаточно большой сферы существование локальных неоднородностей не должно заметно влиять на количество материи в ней и потому не должно сказываться на существовании замкнутых ловушечных поверхностей вокруг нас в современную эпоху.

Приведенное выше рассуждение не зависит от спектра микроволнового излучения, но подразумевает выполнение принципа Коперника. Рассуждение, которое мы приведем теперь, не связано с принципом Коперника, но в определенной степени зависит от формы спектра. Мы предположим, что приблизительно равновесный характер спектра и высокая степень изотропии фонового излучения являются свидетельством того, что оно, по крайней мере частично, термализовано многократным рассеянием. Иначе говоря, на каждой направленной от нас в прошлое изотропной геодезической должно быть достаточно вещества, чтобы прозрачность в этом направлении была не слишком велика. Мы покажем затем, что этого вещества будет достаточно, чтобы наш световой конус прошлого оказался сходящимся в прошлом.

Рассмотрим точку p, изображающую нас в настоящий момент времени, и пусть W^a — направленный в прошлое единичный вектор, параллельный нашей 4-скорости.

Аффинный параметр v направленных в прошлое изотропных геодезических, проходящих через p, можно нормировать условием $K^a W_a = -1$, где $\mathbf{K} = \partial/\partial v$ — касательный к этим изотропным геодезическим вектор. Расхождение $\hat{\theta}$ в этих изотропных геодезических будет подчиняться уравнению (4.35) с $\hat{\omega} = 0$. Таким образом, если $R_{ab}K^aK^b \geqslant 0$, то $\hat{\theta} < 2/v$. Отсюда следует, что при $v = v_1 > v_0$

$$\int_{v_0}^{v_1} R_{ab}K^aK^b \, dv - \frac{2}{v} > \hat{\theta},$$

и поэтому $\hat{\theta}$ станет отрицательным, если найдется v_0, такое, что

$$\int_{v_0}^{v_1} R_{ab}K^aK^b \, dv > 2/v_0.$$

Используя уравнения Эйнштейна с $\Lambda = 0$, получаем

$$\frac{1}{2} v_0 \int_{v_0}^{v_1} 8\pi T_{ab}K^aK^b \, dv > 1. \tag{10.6}$$

На сантиметровых волнах наибольшее отношение непрозрачности к плотности вещества при разумных значениях последней соответствует томсоновскому рассеянию на свободных электронах в ионизованном водороде. Поэтому оптическая толща на расстоянии v будет меньше

$$\int_0^v \varkappa\rho \, (K^aV_a) \, dv,$$

где $\varkappa$ — непрозрачность вследствие томсоновского рассеяния на единицу массы, ρ — плотность вещества и V_a — локальная скорость газа. Красное смещение вещества определяется формулой $z = K^a V_a - 1$. Поскольку вещество со значительным фиолетовым смещением не наблюдается, мы будем предполагать, что на нашем световом конусе прошлого величина $K^a V_a$ всегда больше единицы вплоть до оптической толщи, равной единице. Поскольку на этих длинах волн наблюдаются галактики с красными смещениями 0,3, большая часть рассеяния должна происходить при красных смещениях $z > 0{,}3$. (В самом деле, если квазары — действительно космологические объекты, то рассеяние должно происходить при красных смещениях $z > 2$.) При постоянной Хаббла порядка 100 км/(с·Мпс) $\sim 10^{10}$ лет$^{-1}$ красное смещение 0,3 соответствует расстоянию около $3 \cdot 10^{27}$ см. Принимая для v_0 это значение для вклада в интеграл (10.6) вещества, которое вызывает рассматриваемое рассеяние, получим

$$3{,}7 \cdot 10^{28} \int_{v_0}^{v_1} \rho (K_a V^a)^2 \, dv,$$

тогда как оптическая толща вещества между v_0 и v_1 меньше

$$6{,}6 \cdot 10^{27} \int_{v_0}^{v_1} \rho (K^a V_a) \, \mathrm{d}v.$$

Поскольку $K^a V_a \geqslant 1$, легко убедиться, что неравенство (10.6) будет выполнено при оптической толще меньше 0,2. При оптической толще Вселенной, меньшей 1, нельзя ожидать ни спектра, близко совпадающего со спектром излучения абсолютно черного тела, ни такой высокой степени изотропии в малых угловых масштабах, если только не существует очень большого числа дискретных источников, покрывающих лишь малую часть неба (по площади), при этом каждый из них должен в грубом приближении иметь спектр излучения абсолютно черного тела с температурой 3 К, но значительно большей интенсивности. Это представляется маловероятным. Поэтому мы уверены, что условие 4. III теоремы 2 удовлетворяется и, следовательно, где-то во Вселенной должна быть сингулярность, если конечно, выполнены и остальные условия.

Ввиду своей общности теорема 2 не говорит нам, находится ли сингулярность в нашем прошлом, или в будущем нашего прошлого. Казалось бы, очевидно, что сингулярность должна быть в нашем прошлом, однако можно построить пример, в котором она находится в будущем: рассмотрим Вселенную Робертсона — Уокера с $k = +1$, которая коллапсирует к сингулярности в некоторый момент $t = t_0$, а при $t \to -\infty$

асимптотически приближается к статической Вселенной Эйнштейна. В ней удовлетворяется энергетическое условие и имеются точки, световые конусы прошлого которых начинают сходиться (поскольку они пересекаются на «противоположной стороне» Вселенной). Однако сингулярность находится в будущем. Конечно, это довольно искусственный пример, но из него видно, что следует проявлять осторожность. Поэтому мы приведем рассуждение, основанное на теореме 3, которое указывает на то, что при выполнении принципа Коперника Вселенная содержит сингулярность в нашем прошлом. Теорема 3 подобна теореме 2, но требует, чтобы вместо всех изотропных геодезических все времениподобные геодезические, направленные в прошлое из некоторой точки, начинали сходиться. Это условие не выполняется в приведенном выше примере, хотя ему удовлетворяют геодезические, направленные в будущее из любой точки.

Рассуждая так же, как в случае изотропных геодезических, получим, что схождение $\theta(s)$ направленных в прошлое из точки p времениподобных геодезических будет меньше

$$\frac{3}{s_0} - \int_{s_0}^{s} R_{ab}V^aV^b\,\mathrm{d}s,$$

где s — собственное расстояние вдоль этих геодезических, $\mathbf{V} = \partial/\partial s$ и $s > s_0$. Пусть $\mathbf{W}$ — направленный в прошлое времениподобный единичный вектор в точке p и $c \equiv -V^aW_a|_p$ (и, следовательно, $c > 1$). Тогда θ станет меньше $-c$ в пределах расстояния R_1/c вдоль любой геодезической, если существует некоторое R_0 ($R_1 > R_0 > 0$), такое, что вдоль этой геодезической

$$\int_{R_1/c}^{R_1/c} R_{ab}V^aV^b\,\mathrm{d}s > c\left(\frac{3}{R_0} + \varepsilon\right). \tag{10.7}$$

Условие (3) теоремы 3 будет удовлетворено при $b = \max(R_1, (3\varepsilon)^{-1})$.

Чтобы привести неравенство (10.7) к виду, подобному (10.6), мы введем вдоль времениподобных геодезических аффинный параметр $v = s/c$; тогда (10.7) приобретает вид

$$\frac{1}{3}R_0\int_{R_1}^{R_1} R_{ab}K^aK^b\,\mathrm{d}v > 1 + R_0\varepsilon, \tag{10.8}$$

где $\mathbf{K} = \partial/\partial v$ и $K^aW_a|_p = -1$. В отличие от (10.6) мы не можем убедиться в выполнении условия (10.8) прямо по его явному виду, поскольку оно записано для времениподобных геодезических. Поэтому мы обратимся к рассуждению, приведенному в первой части настоящего раздела, где мы показали, что

Вселенная близка к модели Робертсона — Уокера по крайней мере в прошлые времена до момента, когда микроволновое фоновое излучение претерпело последнее рассеяние.

Пусть **W** — вектор $-\partial/\partial t$ в модели Робертсона — Уокера. Вдоль направленной в прошлое времениподобной геодезической, проходящей через p,

$$\frac{\mathrm{d}}{\mathrm{d}v}(W_a K^a) = W_{a;b} K^a K^b = \frac{1}{S}\frac{dS}{dt}\left\{(W^a K_a)^2 - \frac{1}{c^2}\right\}.$$

Отсюда, $W_a K^a \leqslant -1$, если $\mathrm{d}S/\mathrm{d}t > 0$. Однако

$$W^a K_a = \frac{\mathrm{d}t}{\mathrm{d}v},$$

и, следовательно, при некотором $\varepsilon > 0$ условие (10.8) будет удовлетворено для любой геодезической, если только существуют моменты времени t_2, t_3, $t_2 < t_3 < t_p$, такие, что

$$\frac{t_p - t_3}{3}\int_{t_2}^{t_3} R_{ab} K^a K^b (-W_c K^c)^{-1}\,\mathrm{d}t > 1. \qquad (10.9)$$

В силу уравнений поля с $\Lambda = 0$

$$R_{ab} K^a K^b = 8\pi\left\{(\mu + p)(W_a K^a)^2 - \frac{1}{2}(\mu - p)c^{-2}\right\}.$$

Поэтому при условии, что $p \geqslant 0$,

$$R_{ab} K^a K^b \geqslant 4\pi\mu (W_a K^a)^2.$$

Следовательно, (10.9) удовлетворяется, если

$$\frac{t_p - t_3}{3}\int_{t_2}^{t_3} 4\pi\mu\,\mathrm{d}t > 1. \qquad (10.10)$$

Если принять, что микроволновое излучение имеет спектр абсолютно черного тела при температуре 2,7 К, то в настоящее время плотность его энергии порядка 10^{-34} г/см^3. Если это излучение — первичное, то плотность его энергии будет пропорциональна S^{-4}. Поскольку $S^{-1} = O(t^{-1/2})$ при стремлении t к нулю, неравенству (10.10) можно удовлетворить, взяв $t_3 = t_p/2$ и достаточно малое t_2. Насколько мало значение t_2, должно зависеть от конкретного поведения S, которое в свою очередь зависит от плотности вещества во Вселенной. Это довольно неопределенная величина, но она, по-видимому, лежит между 10^{-31} и $5\cdot 10^{-29}$ г/см^3. В первом случае t_2 должно быть таким, чтобы $S(t_p)/S(t_2) \geqslant 30$, а в последнем — таким, чтобы $S(t_p)/S(t_2) \geqslant 300$. Микроволновое излучение, по-видимому, заполняет всю Вселенную, поэтому любая направленная в прошлое времениподобная геодезическая должна проходить сквозь

него. Таким образом, оценка, основанная на модели Робертсона — Уокера, должна служить хорошим приближением для вклада фонового излучения в (10.10), если оно возникло позднее момента t_2 и если сама модель Робертсона — Уокера является хорошим приближением для столь далекого прошлого. Из рассуждений, приведенных в начале данного раздела, следует, что последнее справедливо при условии, что с момента t_2 фоновое излучение распространялось к нам свободно. Однако не исключено, что существует ионизованный межгалактический газ плотностью порядка $5 \cdot 10^{-29}$ г/см^3, и в этом случае излучение могло претерпеть последнее рассеяние в момент времени t, в который $S(t_p)/S(t) \approx 5$. Оптическая толща в прошлое к моменту t равна

$$\int_t^{t_p} \varkappa \mu_{\text{газ}}\, \mathrm{d}t, \tag{10.11}$$

где $\varkappa$ не превышает 0,5, если μ измерять в г/см^3 и t в см.

В прошлом не может быть значительной непрозрачности до $t = t_p - 10^{17}$ с, так как нам видны объекты на расстоянии по крайней мере $3 \cdot 10^{27}$ см. Принимая для t_3 это значение, мы видим, что при такой плотности ионизованного газа соотношение (10.10) будет удовлетворяться при значении t_2, которое соответствует оптической толще, не превышающей 0,5.

Итак, сложилось следующее положение. Мы принимаем, что выполняется принцип Коперника и что микроволновое излучение было испущено либо до момента t_2, в который $S(t_p)/S(t_2) \approx \approx 300$, либо до момента, в который оптическая толща Вселенной равна единице, если этот момент более поздний, чем t_2. В первом случае условие (2) теоремы 3 будет удовлетворено за счет плотности излучения, а во втором случае — за счет плотности газа. Следовательно, если выполнены обычные энергетические условия и условия причинности, то мы можем заключить, что в нашем прошлом должна быть сингулярность (т. е. должна существовать направленная от нас в прошлое непространственноподобная геодезическая, которая неполна).

Пусть имеется пространственноподобная поверхность, пересекающая наш изотропный световой конус прошлого, и ряд точек на этой поверхности; можно ли утверждать, что в прошлом каждой из них есть сингулярность? Это было бы так, если бы Вселенная была настолько однородна и изотропна в прошлом, чтобы могло возникнуть схождение всех направленных в прошлое времениподобных геодезических из этих точек. Ввиду тесной связи между схождением времениподобных геодезических и замкнутыми ловушечными поверхностями можно ожидать, что нужная степень однородности и изотропии в прошлом име-

лась, если Вселенная была однородна и изотропна на расстояниях порядка шварцшильдовой длины $(3/8\pi\mu)^{1/2}$.

Измерения Пензиаса, Шрамля и Уилсона [134], обнаруживших, что интенсивность микроволнового фонового излучения изотропна в пределах 4% при телесном угле диаграммы $1 \cdot 4 \cdot 10^{-3}$ квадратных градуса, дают прямые свидетельства однородности Вселенной в прошлом. В предположении, что микроволновое излучение не генерировалось позднее поверхности, соответствующей оптической толще, равной единице, получаем, что наблюдаемая интенсивность пропорциональна $T^4/(1+z)^4$, где T — эффективная температура наблюдаемой точки на этой поверхности и z — ее красное смещение. Вариации наблюдаемой интенсивности могут возникнуть по четырем причинам:

1) из-за доплеровского сдвига, вызванного нашим собственным движением относительно излучения абсолютно черного тела [154, 163];

2) из-за вариаций гравитационного красного смещения, вызванных неоднородностями в распределении вещества между нами и рассматриваемой поверхностью [138, 144];

3) из-за доплеровских сдвигов, вызванных локальными возмущениями скоростей вещества на поверхности;

4) из-за вариаций эффективных температур поверхности. (В действительности деление на пункты (1), (2) и (3) зависит от способа отсчета и имеет только эвристическое значение.) Таким образом, наблюдения свидетельствуют, что неоднородности температуры с характерными угловыми размерами $3'$ имеют относительные амплитуды меньше 1% и что при тех же характерных размерах не существует локальных флуктуаций скорости вещества с 1% скорости света. Область поверхности с угловым диаметром около $3'$ будет соответствовать области, которая сейчас имеет диаметр около 10^7 св. лет. Если поверхность, соответствующая оптической толще, равной единице, имеет красное смещение $z = 1000$ (максимально возможное), то шварцшильдова длина в тот момент соответствовала бы области с современным диаметром около $3 \cdot 10^8$ св. лет. Таким образом, каждая точка на поверхности единичной оптической толщи, по-видимому, должна иметь в своем прошлом сингулярность.

Более косвенные данные о степени однородности Вселенной на ранних стадиях основаны на том, что наблюдения содержания гелия в ряде небесных объектов согласуются с расчетами образования гелия в предположении, что Вселенная была однородна и изотропна в прошлом вплоть до температуры порядка 10^9 К [121, 170]. С другой стороны, расчеты по анизотропным моделям дают совершенно иные количества гелия. Поэтому, если принять, что гелий распределен во Вселенной

довольно равномерно (относительно этого есть некоторые сомнения) и что этот гелий образовался на ранних стадиях Вселенной, то можно сделать вывод, что Вселенная была эффективно изотропна и однородна в эпоху с температурой 10^9 К. Следовательно, можно ожидать, что в прошлом любой точки этого времени имеется сингулярность.

Мизнер [106] показал, что столкновения электронов и нейтрино создают значительную вязкость, если температура достигает $2 \cdot 10^{10}$ К. Эта вязкость могла бы привести к затуханию неоднородностей, размеры которых соответствуют в настоящее время расстоянию 10 св. лет, и снизить анизотропию до сравнительно малой величины. Таким образом, если принять подобное объяснение современной изотропии Вселенной (такое объяснение весьма привлекательно), то можно заключить, что сингулярность должна быть в прошлом любой точки в эпоху с температурой около 10^{10} К.

10.2. Природа и проявление сингулярностей

Есть надежда узнать кое-что о природе сингулярностей, которые, по-видимому, все же появляются, исследуя точные решения с сингулярностями. Но хотя мы и показали, что малые возмущения начальных данных не препятствуют появлению сингулярности, все же не очевидно, что свойства сингулярности будут столь же устойчивы. В разд. 7.5 мы показали, что задача Коши устойчива относительно малых возмущений начальных условий, но эта устойчивость относится только к компактным подобластям области Коши, а область, содержащая сингулярность, не компактна, за исключением случая, когда сингулярность соответствует захваченной неполноте. Мы приведем пример, в котором характер сингулярности действительно неустойчив. Рассмотрим однородное сферически симметричное облако пыли, коллапсирующее к сингулярности. Метрика внутри облака будет подобна метрике части вселенной Робертсона — Уокера, тогда как область вне облака будет описываться метрикой Шварцшильда. И внутри, и вне облака пыли сингулярность пространственноподобна (рис. 63, *а*). Допустим теперь, что к этой пыли мы добавили электрический заряд небольшой плотности. Метрика вне облака пыли станет теперь частью решения Райсснера — Нордстрема для $e^2 < m^2$ (рис. 63, *б*). Внутри облака появится сингулярность, так как достаточно малая плотность заряда не помешает ее возникновению. Характер сингулярности внутри облака пыли, по-видимому, будет зависеть от распределения заряда. Важно, однако, следующее: как только поверхность облака пыли пройдет точку p внутри $r = r_+$, что бы ни произошло внутри облака, это не повлияет на часть sq времениподобной сингулярности.

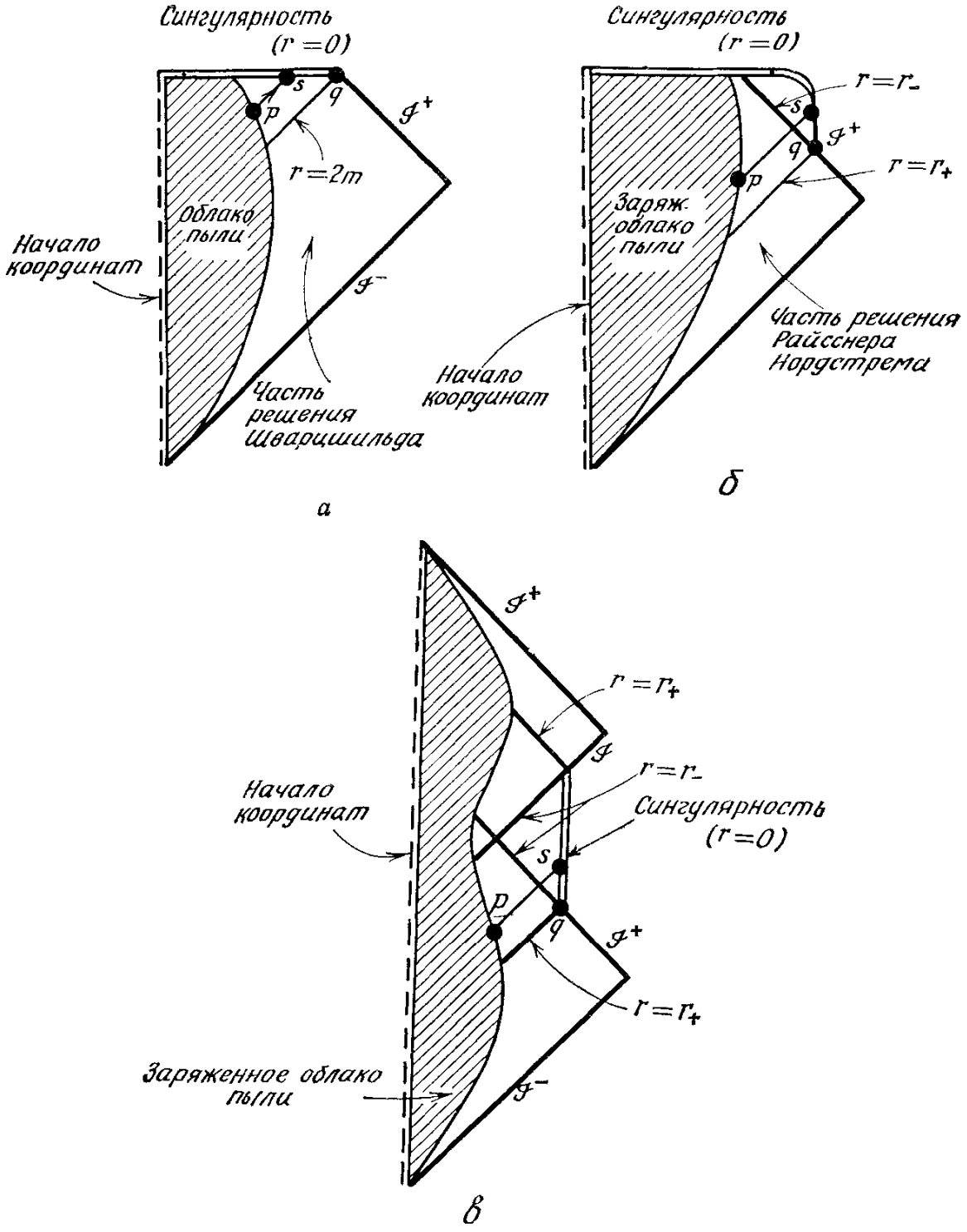


Рис. 63. *а* — коллапс сферического облака пыли; *б* — коллапс заряженного облака пыли, в котором заряд слишком мал, чтобы помешать появлению в облаке пыли сингулярности; *в* — коллапс заряженного облака пыли, в котором имеется заряд, достаточный для того, чтобы воспрепятствовать появлению сингулярности в самом облаке; сингулярность появляется вне облака, которое сжимается и снова расширяется в другое асимптотически плоское пространство

Если теперь увеличить плотность заряда настолько, что она станет больше плотности вещества, то облако может пройти через два горизонта в точках $r = r_+$ и $r = r_-$ и вновь начнет расширяться в другую вселенную, при этом внутри него не появится никакой сингулярности, хотя вне его будет времениподобная сингулярность (*Дж. М. Бардин*, не опубликовано), в полном согласии с теоремой 2 (рис. 63, *в*).

Этот пример очень важен, так как он показывает, что возможны времениподобные сингулярности, что материя может избежать сингулярного состояния и что она может проникнуть сквозь «кротовую нору» в другую область или в другую часть той же области пространства-времени. Конечно, вряд ли коллапсирующая звезда может иметь столь большую плотность заряда, но поскольку решение Керра так похоже на решение Райсснера — Нордстрема, можно ожидать, что момент количества движения способен создать подобную «кротовую нору». Можно немного пофантазировать и предположить, что до современной фазы расширения Вселенная находилась в фазе сжатия, в течение которой росли локальные неоднородности и возникали отдельные сингулярности; однако большая часть материи, избежав этих сингулярностей, перешла к расширению, превратившись в нашу наблюдаемую Вселенную.

Существование сингулярностей в пределах прошлого каждой точки в весьма ранний момент времени, когда плотность была велика, накладывает ограничения на расстояние между сингулярностями. В принципе можно было бы представить, что множество геодезических, которые «упираются» в эти сингулярности (т. е. неполны), имеет меру нуль. Тогда можно было бы утверждать, что эти сингулярности физически несущественны. Однако на самом деле это не так, потому что существование таких сингулярностей приводит к образованию горизонта Коши и, следовательно, к невозможности предсказывать будущее. Мы могли бы получить таким образом путь к решению проблемы энтропии в модели пульсирующей Вселенной, поскольку в каждом цикле сингулярности могли бы вносить отрицательную энтропию.

До сих пор мы исследовали математические следствия предположения о том, что модель пространства-времени представляет собой лоренцево многообразие и что при этом удовлетворяются уравнения поля Эйнштейна с $\Lambda = 0$. Мы показали, что, согласно такой теореме, в нашем прошлом должны быть сингулярности, связанные с коллапсом Вселенной, а в нашем будущем — сингулярности, связанные с коллапсом звезд. Если $\Lambda < 0$, то эти выводы остаются неизменными. Если же $\Lambda > 0$, то наблюдения скорости расширения Вселенной [145, 146] говорят о том, что Λ не может превышать $3 \cdot 10^{-55}$ см$^{-2}$. Это зна-

чение эквивалентно отрицательной плотности энергии $3 \times \times 10^{-27}$ г/см^3. Такое значение Λ могло бы сказаться на расширении Вселенной в целом, но оно бессильно повлиять на положительную плотность вещества коллапсирующей звезды. Таким образом, Λ-член, по-видимому, не дает нам возможности избежать проблемы сингулярности.

Можно, конечно, сказать, что общая теория относительности не дает правильного описания Вселенной. До сих пор эта теория проверялась только при весьма малых отклонениях от плоского пространства (радиусы кривизны порядка 10^{12} см). Поэтому ее применение к явлениям типа коллапса звезды, когда радиус кривизны становится меньше 10^6 см, кажется слишком далеко идущей экстраполяцией. С другой стороны, теоремы о сингулярностях зависят от уравнений Эйнштейна не во всей полноте, а только от того, что $R_{ab}K^aK^b \geqslant 0$ для любого непространственноподобного вектора K^a; поэтому они будут справедливы также и в любой модификации общей теории относительности (такой, как теория Бранса — Дикке), в которой тяготение всегда есть сила притяжения.

По-видимому, можно выдвинуть следующий разумный принцип: если теория предсказывает сингулярность, то это указывает на нарушение теории, т. е. она более не дает правильного описания наблюдений. Возникает вопрос: когда нарушается общая теория относительности? Можно ожидать, что в любом случае она нарушается, когда становятся важны квантовые эффекты; из соображений размерности следует, что этого не должно случиться, пока радиус кривизны не станет порядка 10^{-33} см, что соответствует плотности 10^{94} г/см^3. Однако мы можем спросить: является ли лоренцево многообразие приемлемой моделью для пространства-времени при характерных длинах такого порядка? В настоящее время эксперименты свидетельствуют, что предположение о структуре многообразия при характерных длинах более 10^{-15} см приводит к предсказаниям, согласующимся с наблюдениями [49]; но может оказаться, что нарушение теории происходит при длинах между 10^{-15} и 10^{-33} см. Радиус 10^{-15} см соответствует плотности 10^{58} г/см^3, которую всегда практически можно рассматривать как сингулярность. Поэтому следует построить по методу Шмидта (разд. 8.3) поверхность вокруг областей с радиусом кривизны меньше, скажем, 10^{-15} см. С нашей стороны поверхности будет допустимо рассматривать пространство-время как многообразие, но внутри этой поверхности потребуется еще не известное нам квантовое описание. Материю, пересекающую такую поверхность, можно рассматривать как входящую во Вселенную или покидающую ее, и нет причин, по которым количества входящей и уходящей материи должны быть равны.

Во всяком случае, теоремы о сингулярностях свидетельствуют о том, что по предсказаниям общей теории относительности гравитационное поле должно стать предельно сильным. Что это имело место в нашем прошлом, подтверждается существованием микроволнового фонового излучения со спектром абсолютно черного тела, свидетельствующего о том, что у Вселенной была очень горячая и плотная ранняя фаза.

Вероятно, теоремы о существовании сингулярностей можно несколько усовершенствовать, но на наш взгляд они и в таком виде достаточно хорошо отражают суть дела. Однако они мало что говорят нам о характере сингулярностей. Хотелось бы знать, какого рода сингулярности могут появиться в общей теории относительности в характерных ситуациях. Один из путей к этому состоит в усовершенствовании предложенного Лифшицем и Халатниковым метода разложения в степенные ряды и выяснении пределов его применимости. Может оказаться также, что существует определенная связь между сингулярностями, рассматриваемыми в общей теории относительности и в других областях физики (см., например, теорию элементарных катастроф Тома [165]). С другой стороны, можно было бы попытаться решить проблему в лоб, численным интегрированием уравнений Эйнштейна на вычислительной машине. Однако для этого, вероятно, придется ждать появления нового поколения вычислительных машин. Хотелось бы знать также, будут ли голыми сингулярности, возникающие при коллапсе из несингулярного асимптотически плоского состояния, т. е. будут ли они видны с бесконечности или будут скрыты за горизонтом событий.

Другой важной проблемой является формулировка квантовой теории пространства-времени, применимой к сильным полям. Такого рода теория может исходить из некоторого многообразия или может допускать изменения топологии. Некоторые предварительные попытки в этом направлении уже были предприняты Де Виттом [38], Мизнером [107, 108], Пенроузом [132], Уилером [173] и другими. Однако интерпретация квантовой теории пространства-времени и ее отношение к сингулярностям еще далеки от ясности.

Гипотезы и споры, касающиеся предмета этой книги, не новы. По существу, еще Лаплас предсказал существование черных дыр: «Иные звезды появлялись внезапно и исчезали после нескольких месяцев ярчайшего сияния... . Все эти звезды ... не изменяли своих положений, пока они светили. Поэтому в беспредельных глубинах пространства существуют темные тела, столь же значительные по величине и, возможно, столь же многочисленные, как звезды». (П. С. Лаплас, «Изложение

системы мира»). Как мы видим, современное понимание положения вещей удивительно похоже на лапласовское.

С незапамятных времен высказывалась мысль о возникновении Вселенной из ничего; см., например, первую «Антиномию Чистого Разума» Канта и комментарии к ней ([158], p. 117—123, 145—159 и [117], 389—406). Результаты, которые мы получили, говорят в пользу идеи, что история Вселенной началась некоторое конечное время тому назад. Однако самый момент возникновения — сингулярность — находится за пределами справедливости известных сейчас законов физики.

Приложение А

Пьер Симон Лаплас*)

Доказательство теоремы о том, что сила притяжения небесного тела могла бы быть столь велика, что от него не будет исходить свет **)

1. Если v — скорость, t — время и s — расстояние, которое равномерно проходится в течение этого времени, то, как известно, $v = s/t$.

2. Если движение не является равномерным, то для получения значения v в какой-либо момент нужно поделить пройденное расстояние $\mathrm{d}s$ и интервал времени $\mathrm{d}t$ одно на другое, т. е. $v = \mathrm{d}s/\mathrm{d}t$, поскольку скорость на бесконечно малом интервале постоянна и движение, таким образом, можно принять за равномерное.

3. Непрерывно действующая сила будет стремиться изменить эту скорость. Это изменение скорости, а именно $\mathrm{d}v$, является поэтому наиболее естественной мерой этой силы. Но поскольку любая сила вызывает двойной эффект за двойное время, мы должны разделить изменение скорости $\mathrm{d}v$ на время $\mathrm{d}t$, в течение которого это изменение вызывается силой **P**; таким образом, получается общее выражение для силы **P**, т. е.

$$P = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d} \cdot \frac{\mathrm{d}s}{\mathrm{d}t}}{\mathrm{d}t}.$$

Теперь, если $\mathrm{d}t$ есть постоянная,

$$\mathrm{d} \cdot \frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d} \cdot \mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}\,\mathrm{d}s}{\mathrm{d}t};$$

соответственно

$$P = \frac{\mathrm{d}\,\mathrm{d}s}{\mathrm{d}t^2}.$$

---

*) Allgemeine geographishe Ephemeriden, verfasset von Einer Gesellschaft Gelehrter, Weimer, IV, Bd. Ist, 1799, ed. F. X. von Zach. Мы благодарны Д. У. Дьюхэрсту, снабдившему нас этой ссылкой.

**) Эта теорема о том, что во Вселенной светящееся тело той же плотности, что и Земля, но диаметр которой в 250 раз больше диаметра Солнца, может силой притяжения воспрепятствовать своим лучам света достичь нас и, следовательно, самые большие тела во Вселенной могли бы остаться для нас невидимыми, была приведена Лапласом в его «Изложении системы мира», часть II, стр. 305, без доказательства. См. A. G. E., May 1798, p. 603, v. Z.

4. Пусть сила притяжения тела M; второе тело (например, частица света) находится на расстоянии r; действие силы M на эту частицу света будет $-M/rr$; отрицательный знак появляется из-за того, что действие M противоположно движению света.

5. Теперь согласно (3) эта сила также равна $\mathrm{d}\,\mathrm{d}r/\mathrm{d}t^2$; следовательно,

$$-\frac{M}{rr} = \frac{\mathrm{d}\,\mathrm{d}r}{\mathrm{d}t^2} = -Mr^{-2}.$$

Умножая на $\mathrm{d}r$, получим

$$\frac{\mathrm{d}r\,\mathrm{d}\,\mathrm{d}r}{\mathrm{d}t^2} = -M\,\mathrm{d}r\,r^{-2};$$

после интегрирования

$$\frac{1}{2}\frac{\mathrm{d}r^2}{\mathrm{d}t^2} = C + Mr^{-1},$$

где C — постоянная величина, или

$$\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 = 2C + 2Mr^{-1}.$$

Теперь, в силу (2), $\mathrm{d}r/dt$ — скорость v и соответственно справедливо равенство

$$v^2 = 2C + 2Mr^{-1},$$

где v — скорость частицы света на расстоянии r.

6. Теперь для определения постоянной C допустим, что R — радиус притягивающего тела и a — скорость света на расстоянии R (следовательно, на поверхности притягивающего тела); тогда из (5) получается

$$a^2 = 2C + 2M/R,$$

поэтому

$$2C = a^2 - 2M/R.$$

Подставляя это в предыдущее уравнение, получим

$$v^2 = a^2 - \frac{2M}{R} + \frac{2M}{r}.$$

7. Пусть R' будет радиус другого притягивающего тела; его сила притяжения будет iM, а скорость света на расстоянии r составит v'; тогда, согласно уравнению в (6),

$$v'^2 = a^2 - \frac{2iM}{R'} + \frac{2iM}{r}.$$

8. Если сделать r бесконечно большим, последний член в предыдущем уравнении обращается в нуль, и мы получаем

$$v'^2 = a^2 - \frac{2iM}{R'}.$$

Расстояние до неподвижных звезд столь велико, что это предположение справедливо.

9. Пусть сила притяжения второго тела так велика, что свет не может удалиться от него; это можно выразить аналитически следующим образом: скорость света v' равна нулю. Подставив такое значение v' в уравнение (8) для v', получим уравнение, из которого можно вывести массу iM, при которой это происходит. Следовательно, имеем

$$0 = a^2 - \frac{2iM}{r'} \quad \text{или} \quad a^2 = \frac{2iM}{R'}.$$

10. Чтобы определить a, допустим сначала, что этим притягивающим телом является Солнце; тогда a — скорость света Солнца на поверхности Солнца. Однако сила притяжения Солнца в сравнении со скоростью света столь мала, что можно эту скорость принять за равномерную. Из явления аберрации следует, что Земля проходит $20''\frac{1}{4}$ своего пути, в то время как свет проходит от Солнца до Земли; следовательно: пусть V есть средняя скорость Земли на ее орбите, тогда мы имеем

$a{:}V =$ радиусу (выраженному в секундах): $20''\frac{1}{4} = 1 : \operatorname{tg} 20''\frac{1}{4}$.

11. Мое предположение в «Изложении системы мира», часть II, стр. 305, состоит в том, что $R' = 250\,R$. Далее, масса изменяется как объем притягивающего тела, умноженный на его плотность, объем — как куб радиуса, соответственно масса — как куб радиуса, умноженный на плотность. Пусть плотность Солнца = 1, плотность второго тела = ρ; тогда

$$M{:}iM = 1R^3 : \rho R'^3 = 1R^3{:}\rho 250^3 R^3,$$

или

$$1{:}i = 1 : \rho\,(250)^3,$$

или

$$i = (250)^3\,\rho.$$

12. Подставляем значения i и R' в уравнение $a^2 = 2iM/R$ и, таким образом, имеем

$$a^2 = \frac{2\,(250)^3\,\rho M}{250R} = 2\,(250)^2\,\rho\,\frac{M}{R},$$

или

$$\rho = \frac{a^2 R}{2\,(250)^2\,M}.$$

13. Чтобы получить ρ, необходимо еще определить M. Сила M Солнца на расстоянии D равна M/D^2. Пусть D — среднее расстояние до Земли, V — средняя скорость Земли; тогда эта сила равна также V^2/D (см. «Астрономию» Ланде, III, § 3539).

Отсюда $M/D^2 = V^2/D$, или $M = V^2D$. Подстановка этого в уравнение (12) для ρ дает

$$\rho = \frac{a^2R}{2\,(250)^2\,V^2D} = \frac{8}{(1000)^2}\left(\frac{a}{V}\right)^2\frac{R}{D},$$

$$\frac{a}{V} = \frac{\text{скорость света}}{\text{скорость Земли}} = \frac{1}{\operatorname{tg} 20''\,\frac{1}{4}} \text{ согласно (10),}$$

$$\frac{R}{D} = \frac{\text{абс. радиус } \odot}{\text{среднее расстояние до } \odot} = \operatorname{tg} \text{ среднего видимого радиуса } \odot.$$

Следовательно,

$$\rho = 8\,\frac{\operatorname{tg} 16'2''}{\left(1000 \operatorname{tg} 20''\,\frac{1}{4}\right)^2},$$

откуда следует, что плотность ρ приблизительно равна 4, или такая же по величине, как и плотность Земли.

Приложение Б

Сферически-симметричные решения и теорема Биркхофа

Мы хотим рассмотреть уравнение Эйнштейна в случае сферически-симметричного пространства. Существенным свойством сферически-симметричного пространства-времени можно считать существование мировой линии $\mathscr{L}$, относительно которой пространство-время сферически-симметрично. Тогда эквивалентны все точки на каждой пространственноподобной 2-сфере $\mathscr{S}_d$ с центром в произвольной точке $p \in \mathscr{L}$ и определяемой как поверхность постоянного расстояния d вдоль всех геодезических, ортогональных к $\mathscr{L}$ в точке p. Если изменить направление в точке p действием ортогональной группы $SO(3)$, оставляющей сферу $\mathscr{L}$ инвариантной, пространство-время по определению не меняется и соответствующие точки сферы $\mathscr{S}_d$ отображаются друг в друга; таким образом, пространство-время допускает группу $SO(3)$ в качестве группы изометрий, причем орбитами группы являются сферы $\mathscr{S}_d$. [Могут быть частные значения d, при которых поверхность $\mathscr{S}_d$ превращается в точку p'; тогда p' будет другим центром симметрии. Точек (p' и сама точка p), связанных таким образом, может быть самое большее две.]

Однако может быть так, что в некоторых пространствах, которые хотелось бы рассматривать как сферически-симметричные, мировая линия, подобная $\mathscr{L}$, не будет существовать. Например, в решениях Шварцшильда и Райсснера — Нордстрема пространство-время сингулярно в точках, для которых $r = 0$; иначе эти точки могли бы быть центрами симметрии. Поэтому мы примем существование группы изометрий $SO(3)$, действующей на 2-поверхностях, подобных $\mathscr{S}_d$, за характеристическое свойство сферически-симметричного пространства-времени.

Итак, мы скажем, что пространство-время *сферически-симметрично*, если оно допускает в качестве группы изометрий группу $SO(3)$ с орбитами в виде пространственноподобных 2-поверхностей. Тогда эти орбиты необходимо будут 2-поверхностями постоянной положительной кривизны.

Для каждой точки q на любой орбите $\mathscr{S}(q)$ существует одномерная подгруппа изометрий I_q, оставляющая q инвариантной (если есть центральная ось $\mathscr{S}$, то это — группа вращений вокруг p, оставляющая геодезическую pq инвариантной). Мно-

жество $\mathscr{C}(q)$ всех геодезических, ортогональных к $\mathscr{S}(q)$, в точке q образует локально 2-поверхность, инвариантную относительно I_q [поскольку подгруппа I_q, которая меняет направления в $\mathscr{S}(q)$ относительно q, оставляет неизменными направления, перпендикулярные к $\mathscr{S}(q)$]. В любой другой точке $r \in \mathscr{C}(q)$ группа I_q снова меняет направления, ортогональные к $\mathscr{C}(q)$, оставляя в то же время $\mathscr{C}(q)$ инвариантной; ввиду того что I_q должна действовать в групповой орбите $\mathscr{S}(r)$, проходящей через r, эта орбита ортогональна к $\mathscr{C}(q)$. Таким образом, групповые орбиты $\mathscr{S}$ ортогональны к поверхностям $\mathscr{C}$ [147]. Далее, эти поверхности локально определяют взаимно-однозначное отображение между орбитами группы, в котором $f(q)$, образ точки q в $\mathscr{S}(r)$, является пересечением $\mathscr{C}(q)$ и $\mathscr{S}(r)$. Поскольку это отображение инвариантно относительно действия I_q, векторы равной длины в $\mathscr{S}(q)$ в точке q отображаются в векторы равной длины в $\mathscr{S}(r)$ в точке $f(q)$; и так как все точки $\mathscr{S}(q)$ эквивалентны, при отображении векторов из любой точки $\mathscr{S}(q)$ в ее образ в $\mathscr{S}(r)$ все длины получают один и тот же множитель. Таким образом, ортогональные поверхности $\mathscr{C}$ конформно отображают орбиты $\mathscr{S}$ одну в другую.

Если выбрать координаты $\{t, r, \theta, \phi\}$ так, чтобы групповыми орбитами $\mathscr{S}$ стали поверхности $\{t, r = \text{const}\}$, а ортогональными поверхностями $\mathscr{C}$ — поверхности $\{\theta, \phi = \text{const}\}$, то метрика при этом примет вид $ds^2 = d\tau^2(t, r) + Y^2(t, r)\, d\Omega^2(\theta, \phi)$, далее выбирать функции t, r так, чтобы $\{t = \text{const}\}$, $\{r = \text{const}\}$ где $d\tau^2$ — неопределенная метрика 2-поверхности и $d\Omega^2$ — метрика поверхности постоянной положительной кривизны. Если были ортогональны в 2-поверхностях $\mathscr{C}$ (ср. с [6]), метрику можно записать в виде

$$ds^2 = \frac{-dt^2}{F^2(t, r)} + X^2(t, r)\, dr^2 + Y^2(t, r)(d\theta^2 + \sin^2\theta\, d\phi^2). \quad (Б.1)$$

(Отметим, что на этих поверхностях все еще остается свобода произвольного выбора r или t.)

Допустим, что наблюдатель, движущийся вдоль линий t, измеряет плотность энергии μ, изотропное давление, поток энергии q, причем анизотропных давлений нет. Тогда уравнения поля для метрики (Б. 1) можно записать в виде

$$-8\pi q = \frac{2X}{F}\left(\frac{\dot{Y}'}{Y} - \frac{\dot{X}Y'}{XY} + \frac{\dot{Y}F'}{YF}\right), \quad (Б.2)$$

$$8\pi\mu = \frac{1}{Y^2} + \frac{2}{X}\left(-\frac{Y'}{XY}\right)' - 3\left(\frac{Y'}{XY}\right)^2 + \\ + 2F^2\frac{\dot{X}\dot{Y}}{XY} + F^2\left(\frac{\dot{Y}}{Y}\right)^2, \quad (Б.3)$$

$$-8\pi p = \frac{1}{Y^2} + 2F\left(F\frac{\dot{Y}}{Y}\right)^{\cdot} + 3\left(\frac{\dot{Y}}{Y}\right)^2 F^2 + \\ + \frac{2}{X^2}\frac{Y'F'}{YF} - \left(\frac{Y'}{XY}\right)^2, \quad \text{(Б.4)}$$

$$4\pi(\mu + 3p) = \frac{1}{X}\left(-\frac{F'}{FX}\right)' - F\left(F\frac{\dot{X}}{X}\right)^{\cdot} - 2F\left(F\frac{\dot{Y}}{Y}\right)^{\cdot} - \\ - F^2\left(\frac{\dot{X}}{X}\right)^2 - 2F^2\left(\frac{\dot{Y}}{Y}\right)^2 + \frac{1}{X^2}\left(\frac{F'}{F}\right)^2 - \frac{2}{X^2}\frac{Y'F'}{YF}, \quad \text{(Б.5)}$$

где штрих означает $\partial/\partial r$, а точка — $\partial/\partial t$.

Рассмотрим сначала уравнения поля в пустом пространстве $R_{ab} = 0$; это означает, что в (Б. 2)—(Б. 5) мы должны положить $\mu = p = q = 0$. Локальное решение зависит от рода поверхностей $\{Y = \text{const}\}$; эти поверхности могут быть времениподобны, пространственноподобны или изотропны; кроме того, они могут быть не определены (если Y — постоянная). В исключительном случае, когда $Y^{;a}Y_{;a} = 0$ на некотором открытом множестве $\mathcal{U}$ (сюда входит случай постоянного Y), на $\mathcal{U}$ выполняется равенство

$$\frac{Y'}{X} = F\dot{Y}. \quad \text{(Б.6)}$$

Однако, если справедливо (Б. 6), значение $\dot{Y}'$, получающееся из (Б. 2), не согласуется с (Б. 3). Следовательно, мы можем рассматривать точку p, где $Y^{;a}Y_{;a} < 0$ или $Y^{;a}Y_{;a} > 0$; такое неравенство должно выполняться в некоторой открытой окрестности $\mathcal{U}$ точки p.

Рассмотрим прежде всего случай, когда $Y^{;a}Y_{;a} < 0$. Тогда в $\mathcal{U}$ поверхности $\{Y = \text{const}\}$ времениподобны, и можно взять Y в качестве координаты r. (При этом r будет *координатой, определяемой площадью*, так как площадь 2-поверхности $\{r, t = \text{const}\}$ равна $4\pi r^2$.) Итак, $\dot{Y} = 0$, $Y' = 1$, а из (Б. 2) видно, что $\dot{X} = 0$. Далее, из (Б. 4) следует, что $(F'/F)^{\cdot} = 0$, и поэтому можно выбрать новую временную координату $t'(t)$ таким образом, чтобы положить $F = F(r)$. Тогда имеем $F = F(r)$, $X = X(r)$, $Y = r$ и решение *необходимо статическое*. Из уравнения (Б. 3) после этого получается, что $\mathrm{d}(r/X^2)/\mathrm{d}r = 1$, так что решение имеет вид $X^2 = (1 - 2m/r)^{-1}$, где $2m$ — постоянная интегрирования. Уравнение (Б. 4) можно проинтегрировать; при подходящем выборе постоянной интегрирования получим $F^2 = X^2$, и тогда (Б. 5) тождественно удовлетворяется. При таких F и X метрика имеет вид

$$\mathrm{d}s^2 = -\left(1 - \frac{2m}{r}\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{\left(1 - \frac{2m}{r}\right)} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2); \quad \text{(Б.7)}$$

это — метрика Шварцшильда при $r > 2m$.

Теперь предположим, что $Y^{;a}Y_{;a} > 0$. Тогда поверхности $\{Y = \text{const}\}$ пространственноподобны в $\mathcal{U}$, и мы можем выбрать Y в качестве координаты t. Теперь $\dot{Y} = 1$, $Y' = 0$, и из (Б.2) следует, что $F' = 0$. Можно выбрать координату r так, чтобы было $X = X(t)$; тогда $F = F(t)$, $X = X(t)$, $Y = t$ и решение — *пространственноподобно однородное*. Теперь (Б.4) и (Б.5) можно проинтегрировать и получить следующее решение:

$$ds^2 = -\frac{dt^2}{\left(\frac{2m}{t} - 1\right)} + \left(\frac{2m}{t} - 1\right) dr^2 - t^2 (d\theta^2 + \sin^2\theta\, d\phi^2). \quad (Б.8)$$

Это часть решения Шварцшильда, относящаяся к области внутри шварцшильдова радиуса, ибо преобразование $t \to r'$, $r \to t'$ приводит эту метрику к виду (Б.7) с $r' < 2m$. Наконец, если поверхность $\{Y = \text{const}\}$ в некоторой части открытого множества $\mathcal{V}$ пространственноподобна, а в другой части времениподобна, можно в этих частях получить решения (Б.8) и (Б.7) и затем сшить их на поверхности $Y^{;a}Y_{;a} = 0$, как это было сделано в разд. 5.5. Таким образом, получим ту часть максимального решения Шварцшильда, которая лежит в $\mathcal{V}$. В итоге мы доказали *теорему Биркхофа*: любое C^2-решение уравнений Эйнштейна в пустом пространстве, сферически-симметричное в открытом множестве $\mathcal{V}$, локально эквивалентно в $\mathcal{V}$ части максимально расширенного решения Шварцшильда. (Это верно, даже если пространство класса C^0 и кусочно класса C^1, см. [6].)

Теперь мы рассмотрим сферически-симметричные решения для *статической идеальной жидкости*. Можно подобрать такие координаты $\{t, r, \theta, \phi\}$, в которых метрика имеет вид (Б.1), жидкость движется вдоль линий t (и, следовательно, $q = 0$) и $F = F(r)$, $X = X(r)$, $Y = Y(r)$. Из уравнений поля (Б.3), (Б.4) тогда следует, что $\mu + p = 0$, если $Y' = 0$; такой случай мы исключаем как не имеющий смысла для реальной жидкости и считаем, что $Y' \neq 0$. Поэтому мы снова можем выбрать Y в качестве координаты r; тогда метрика имеет вид

$$ds^2 = -\frac{dt^2}{F^2(r)} + X^2(r)\, dr^2 + r^2 (d\theta^2 + \sin^2\theta\, d\phi^2). \quad (Б.9)$$

Из свернутых тождеств Бианки $T^{ab}{}_{;b} = 0$ следует, что

$$p' - (\mu + p)\frac{F'}{F} = 0. \quad (Б.10)$$

Уравнение (Б.5) удовлетворяется тождественно, если справедливы (Б.3), (Б.4) и (Б.10). Уравнение (Б.3) можно проинтегрировать явно и получить

$$X^{-2} = \left(1 - \frac{2\hat{M}}{r}\right)^{-1}, \quad (Б.11)$$

где

$$\hat{M}(r) \equiv 4\pi \int_0^r \mu r^2 \, dr,$$

и использовать граничное условие $X(0) = 1$ (т. е. эта жидкая сфера имеет регулярный центр). С учетом (Б. 10), (Б. 11) уравнение (Б. 4) приводится к виду

$$\frac{dp}{dr} = -\frac{(\mu + p)(\hat{M} + 4\pi p r^3)}{r(r - 2\hat{M})}, \tag{Б.12}$$

которое определяет p как функцию r, если известно уравнение состояния. Наконец, из (Б. 10) имеем

$$F(r) = C \exp \int_{p(0)}^{p(r)} \frac{dp}{\mu + p}, \tag{Б.13}$$

где C — постоянная. Уравнения (Б. 11)—(Б. 13) задают метрику внутри сферы, т. е. до $r = r_0$, где r_0 — значение r на поверхности жидкости.

Обозначения

Числа указывают на страницы, где даны определения.

$\equiv$ — определение
$\exists$ — существует
$\Rightarrow$ — следует
$\sum$ — знак суммы
$\square$ — конец доказательства

Множества

$\cup$ $A \cup B$ — объединение A и B
$\cap$ $A \cap B$ — пересечение A и B
$\supset$ $A \subset B$, $B \supset A$ — A содержится в B
$-$ $A - B$ — B вычитается из A
$\in$ $x \in A$ — x есть элемент множества A
$\varnothing$ — пустое множество

Отображения

$\phi: \mathcal{U} \to \mathcal{V}$, — ϕ отображает $p \in \mathcal{U}$ в $\phi(p) \in \mathcal{V}$
$\phi(\mathcal{U})$ — образ $\mathcal{U}$ при отображении
ϕ^{-1} — отображение, обратное к ϕ
$f \circ g$ — композиция, вслед за g производится отображение f
ϕ_*, ϕ^* — отображения тензоров, индуцированные отображением ϕ, 31

Топология

$\bar{A}$ — замыкание A
$\dot{A}$ — граница A, 203
$\operatorname{int} A$ — внутренность A, 232

Дифференцируемость

C^0, C^r, C^{r-}, C^∞ — условия дифференцируемости, 19

Многообразия

$\mathcal{M}$ — n-мерное многообразие, 19—20
$(\mathcal{U}_\alpha, \phi_\alpha)$ — локальная карта, задающая локальные координаты x^α, 20

$\partial\mathscr{M}$ — край многообразия $\mathscr{M}$, 20
R^n — евклидово n-мерное пространство, 19
$\frac{1}{2}R^n$ — нижняя половина ($x_1 \leqslant 0$) пространства R^n, 19
S^n — n-сфера, 21
$\times$ — прямое (декартово) произведение, 23

Тензоры

$(\partial/\partial t)_\lambda$, X — векторы, 24
$\boldsymbol{\omega}$, $\mathrm{d}f$ — 1-формы (линейные формы), 25, 30
$\langle \boldsymbol{\omega}, \mathrm{X}\rangle$ — скалярное произведение вектора и 1-формы, 25
$\{\mathbf{E}_a\}$, $\{\mathbf{E}^a\}$ — дуальные базисы векторов и 1-форм, 25
$T^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}$ — компоненты тензора типа (r, s), 26—29
$\otimes$ — тензорное произведение, 27
$\wedge$ — внешнее произведение, 30
$(\ )$ — симметризация (т. е. $T_{(ab)}$), 29
$[\ ]$ — антисимметризация (т. е. $T_{[ab]}$), 29
$\delta^a{}_b$ — символ Кронекера ($+1$, если $a = b$, и 0, если $a \neq b$)
T_p, $T^*{}_p$ — касательное пространство и дуальное пространство в точке p, 24
$T^r_s(p)$ — пространство тензоров типа (r, s) в точке p, 27
$T^r_s(\mathscr{M})$ — расслоение тензоров типа (r, s) на $\mathscr{M}$, 64
$T(\mathscr{M})$ — касательное к $\mathscr{M}$ расслоение, 63
$L(\mathscr{M})$ — расслоение линейных реперов на $\mathscr{M}$, 64

Производные и связность

$\partial/\partial x^i$ — частная производная по координате x^i
$(\partial/\partial t)_\lambda$ — производная вдоль кривой $\lambda(t)$, 24
d — внешняя производная, 26, 34
$L_{\mathrm{X}}\mathbf{Y}$, $[\mathbf{X}, \mathbf{Y}]$ — производная Ли вектора $\mathbf{Y}$ по вектору $\mathbf{X}$, 37
∇, ∇_{X}, $T_{ab;c}$ — ковариантная производная, 40—41
$\mathrm{D}/\partial t$ — ковариантная производная вдоль кривой, 42
$\Gamma^i{}_{jk}$ — компоненты связности, 41
exp — экспоненциальное отображение, 44

Римановы пространства

$(\mathscr{M}, \mathbf{g})$ — многообразие $\mathscr{M}$ с метрикой $\mathbf{g}$ и связностью Кристоффеля
$\boldsymbol{\eta}$ — элемент объема, 60
R_{abcd} — тензор Римана, 46

R_{ab} — тензор Риччи, 47
R — скалярная кривизна, 52
C_{abcd} — тензор Вейля, 52
$O(p, q)$ — ортогональная группа, оставляющая метрику G_{ab} инвариантной, 64
G_{ab} — диагональная метрика

$$\mathrm{diag}\ (\underbrace{+1, +1, \ldots, +1}_{p\ \text{членов}}, \underbrace{-1, \ldots, -1}_{q\ \text{членов}})$$

$O(\mathscr{M})$ — расслоение ортонормированных реперов, 64

Пространство-время

Пространство-время есть 4-мерное риманово пространство $(\mathscr{M}, \mathbf{g})$ с нормальной метрической формой diag $(+1, +1, -1, -1)$. В качестве локальных координат выбраны (x^1, x^2, x^3, x^4)

T_{ab} — тензор энергии-импульса материи, 74
$\Psi_{(i)}{}^{a \ldots b}{}_{c \ldots d}$ — материальные поля, 77
L — лагранжиан, 77

Уравнения поля Эйнштейна имеют вид

$$R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} = 8\pi T_{ab},$$

где Λ — космологическая постоянная.
$(\mathscr{S}, \boldsymbol{\omega})$ — комплект начальных данных, 258

Времениподобные кривые

$\perp$ — ортогональная проекция, 92
$\mathrm{D}_F/\partial s$ — производная Ферми, 93
θ — расхождение, 96
ω^a, ω_{ab}, ω — вращение, 96
σ_{ab}, σ — сдвиг, 96

Изотропные геодезические

$\hat{\theta}$ — расхождение, 101
$\hat{\omega}_{ab}$, $\hat{\omega}$ — вращение, 101
$\hat{\sigma}_{ab}$, $\hat{\sigma}$ — сдвиг, 101

Причинная структура

I^+, I^- — хронологические будущее и прошлое, 203
J^+, J^- — причинные будущее и прошлое, 203
E^+, E^- — контуры будущего и прошлого, 204
D^+, D^- — области Коши будущего и прошлого, 135, 224
H^+, H^- — горизонты Коши в будущем и прошлом, 178, 224

Граница пространства-времени

$\mathscr{M}^* = \mathscr{M} \cup \Delta$, где Δ — c-граница, 224

$\mathscr{I}^+, \mathscr{I}^-, i^+, i^-$ — c-граница асимптотически простого и пустого пространства, 138, 250

$\bar{\mathscr{M}} = \mathscr{M} \cup \partial\mathscr{M}$, — когда $\mathscr{M}$ — асимптотически простое в слабом смысле; граница $\partial\mathscr{M}$ многообразия $\mathscr{M}$ состоит из $\mathscr{I}^+$ и $\mathscr{I}^-$, 246, 250

$\mathscr{M}^+ = \mathscr{M} \cup \partial$, — где ∂ есть b-граница, 315

Литература

1. *Ames W. L., Thorne K. S.*, The optical appearance of a star that is collapsing through its gravitational radius, Astrophys. J., **151**, 659—670 (1968).
2. *Arnett W. D.*, Gravitational collapse and weak interactions, Can. J. Phys., **44**, 2553—2594 (1966).
3. *Auslander L., Markus L.*, Flat Lorentz manifolds, Memoir 30, Amer. Math. Soc. (1958).
4. *Avez A.*, Essais de géométrie Riemannienne hyperbolique globale, Applications à la Relativitè Générale, Ann. Inst. Fourier (Grenoble), **132**, 105—190 (1963).
5. *Белинский В. А., Лифшиц Е. М., Халатников И. М.*, Колебательный режим приближения к особой точке в релятивистской космологии, Успехи физ. наук, **102**, 463—500 (1970).
6. *Bergmann P. G., Cahen M., Komar A. B.*, Spherically symmetric gravitational fields, J. Math. Phys., **6**, 1—5 (1965).
7. *Bianchi L.*, Lezioni sulla teoria dei gruppi continui finiti transformazioni, Spoerri, Pisa, 1918.
8. *Bludman S. A., Ruderman M. A.*, Possibility of the speed of sound exceeding the speed of light in ultradense matter, Phys. Rev., **170**, 1176—1184 (1968).
9. *Bludman S. A., Ruderman M. A.*, Noncausality and instability in ultradense matter, Phys. Rev., **D1**, 3243—3246 (1970).
10. *Bondi H.*, Cosmology, Cambridge Univ. Press, London, 1960.
11. *Bondi H.*, Massive spheres in General Relativity, Proc. Roy. Soc. London, **A282**, 303—317 (1964).
12. *Bondi H., Gold T.*, The steady-state theory of the expanding universe, Mon. Not. Roy. Astron. Soc., **108**, 252—270 (1948).
13. *Bondi H., Pirani F. A. E., Robinson I.*, Gravitational waves in General Relativity, III. Exact plane waves, Proc. Roy. Soc. London, **A251**, 519—533 (1959).
14. *Boyer R. H.*, Geodesic Killing orbits and bifurcate Killing horizons, Proc. Roy. Soc. London, **A311**, 245—252 (1969).
15. *Boyer R. H., Lindquist R. W.*, Maximal analytic extension of the Kerr metric, J. Math. Phys., **8**, 265—281 (1967).
16. *Boyer R. H., Price T. G.*, An interpretation of the Kerr metric in General Relativity, Proc. Camb. Phil. Soc., **61**, 531—534 (1965).
17. *Bruhat Y.*, The Cauchy problem, *in* Gravitation: an introduction to current research, ed. L. Witten, Wiley, New York, 1962, pp. 130—168.
18. *Burkill J. C.*, The Theory of Ordinary Differential Equations, Oliver and Boyd, Edinburgh, 1956.
19. *Calabi E., Marcus L.*, Relativistic space forms, Ann. Math., **75**, 63—76 (1962).
20. *Cameron A. G. W.*, Neutron stars, *in* Ann. Rev. Astron. Astrophys., eds. L. Goldberg, D. Layzer, J. G. Phillips, Ann. Rev. Inc., Palo Alto, California, 1970, pp. 179—208.
21. *Carter B.*, The complete analytic extension of the Reissner — Nordström metric in the special case $e^2 = m^2$, Phys. Lett., **21**, 423—424 (1966).

22. *Carter B.*, Stationary axisymmetric systems in General Relativity, Ph. D. Thesis, Cambridge Univ., 1967.
23. *Carter B.*, Global structure of the Kerr family of gravitational fields, Phys. Rev., **174**, 1559—1571 (1968).
24. *Carter B.*, Hamilton — Jacoby and Schrödinger separable solutions of Einstein's equations, Comm. Math. Phys., **10**, 280—310 (1968).
25. *Carter B.*, Killing horizons and orthogonally transitive groups in space-time, J. Math. Phys., **10**, 70—81 (1969).
26. *Carter B.*, The commutation property of a stationary axisymmetric system, Comm. Math. Phys., **17**, 233—238 (1970).
27. *Carter B.*, Causal structure in space-time, J. General Relativity and Gravitation, **1**, 349—391 (1971).
28. *Carter B.*, Axisymmetric black hole has only two degress of freedom, Phys. Rev. Lett., **26**, 331—332 (1971).
29. *Choquet-Bruhat Y.*, Espace — temps Einsteinies gèneraux, chocs gravitationnels, Ann. Inst. Henri Poincaré, **8**, 327—338 (1968).
30. *Choquet-Bruhat Y.*, Equations aux derivées partielles-solutions C^∞ d'equations hyperboliques non-lineaires, Compt. Rend. Acad. Sci. (Paris), 1971.
31. *Choquet-Bruhat Y.*, *Geroch R. P.*, Global aspects of the Cauchy problem in General Relativity, Comm. Math. Phys., **14**, 329—335 (1969).
32. *Christodoulou D.*, Reversible and irreversible transformation in black hole physics, Phys. Rev. Lett., **25**, 1596—1597 (1970).
33. *Clarke C. J. S.*, On the geodesic completeness of causal space-times, Proc. Camb. Phil. Soc., **69**, 319—324 (1971).
34. *Colgate S. A.*, Mass ejection from supernovae, Astrophys. J., **153**, 335—339 (1968).
35. *Colgate S. A.*, *White R. H.*, The hydrodynamic behaviour of supernovae explosions, Astrophys. J., **143**, 626—681 (1966).
36. *Courant R.*, *Hilbert D.*, Methods of Mathematical Physics, Vol. II: Partial Differential Equations, Interscience, New York, 1962. (Русский перевод: *Р. Курант*, Уравнения с частными производными, изд-во «Мир», М., 1964.)
37. *Demianski M.*, *Newman E.*, A combined Kerr — NUT solution of the Einstein field equations, Bull. Acad. Pol. Sci. (Math. Astron. Phys.), **14**, 653—657 (1966).
38. *De Witt B. S.*, Quantum theory of gravity: I. The canonical theory, Phys. Rev., **160**, 1113—1148 (1967); II. The manifestly covariant theory, Phys. Rev., **162**, 1195—1239 (1967); III. Applications of the covariant theory, Phys. Rev., **162**, 1239—1256 (1967).
39. *Dicke R. H.*, The theoretical significance of Experimental Relativity, Blackie, New York, 1964.
40. *Dionne P. A.*, Sur les problèmes de Cauchy hyperboliques bien posés, J. d'Analyses Math., **10**, 1—90 (1962).
41. *Dirac P. A. M.*, A new basis for cosmology, Proc. Roy. Soc. London, **A165**, 199—208 (1938).
42. *Dixon W. G.*, Dynamics of extended bodies in General Relativity: I. Momentum and angular momentum, Proc. Roy. Soc. London, **A314**, 499—527 (1970); II. Moments of the charge-current vector, Proc. Roy. Soc. London, **A319**, 509—547 (1970).
43. *Дорошкевич А. Г.*, *Зельдович Я. Б.*, *Новиков И. Д.*, Гравитационный коллапс несимметричных и вращающихся масс, ЖЭТФ, **49**, 170—181 (1965).
44. *Ehlers J.*, *Geren P.*, *Sachs R. K.*, Isotropic solutions of the Einstein — Liouville equations, J. Math. Phys., **8**, 1344—1349 (1968).
45. *Ehlers J.*, *Kundt W.*, Exact solutions of the gravitational field equations, *in* Gravitation: an Introduction to Current Research, ed. L. Witten, Wiley, New York, 1962, pp. 49—101.

46. *Ehresmann C.*, Les connexions infinitesimales dans un espace fibre differentiable, *in* Colloque de Topologie (Espaces Fibres) Bruxelles 1950, Masson, Paris, 1957, pp. 29—50.
47. *Ellis G. F. R., Sciama D. W.*, Global and non-global problems in cosmology, *in* Studies in Relativity (Synge Festschrift), ed. L. O'Raiffeartaigh, Oxford Univ. Press, London, 1972.
48. *Fields G. B.*, Cosmic background radiation and its interaction with cosmic matter, Rivista del Nuovo Cimento, **1**, 87—109 (1969).
49. *Foley K. J., Jones R. S., Lindebaum S. J., Love W. A., Ozaki S., Platner E. D., Quarles C. A., Willen E. H.*, Experimental test of the pion-nucleon forward dispersion relations at high energies, Phys. Rev. Lett., **19**, 193—198, 622 (1967).
50. *Geroch R. P.*, Singularities in closed universes, Phys. Rev. Lett., **17**, 445—447 (1966).
51. *Geroch R. P.*, Singularities in the space-time of General Relativity, Ph. D. Thesis, Department of Physics, Princeton Univ., 1967.
52. *Geroch R. P.*, Topology in General Relativity, J. Math. Phys., 8, 782—786 (1967).
53. *Geroch R. P.*, Local characterization of singularities in General Relativity, J. Math. Phys., **9**, 450—465 (1968).
54. *Geroch R. P.*, What is a singularity in General Relativity, Ann. Phys. (New York), **48**, 526—540 (1968).
55. *Geroch R. P.*, Spinor structure of space-times in General Relativity. I, J. Math. Phys., **9**, 1739—1744 (1968).
56. *Geroch R. P.*, Spinor structure of space-times in General Relativity. II, J. Math. Phys., **11**, 343—348 ((1970).
57. *Geroch R. P.*, The domain dependence, J. Math. Phys., **11**, 437—439 (1970).
58. *Geroch R. P.*, Singularities, *in* Relativity, eds. S. Fickler, M. Carmeli and L. Witten, Plenum Press, New York, 1970, pp. 259—291. (Русский перевод в сб. «Квантовая гравитация и топология», «Мир», М., 1973.)
59. *Geroch R. P.*, Space-time structure from a global view point, *in* General Relativity and Cosmology, Proc. of Intern. School in Physics «Enrico Fermi», Course XLVII, ed. R. K. Sachs, Academic Press, New York, 1971, pp. 71—103.
60. *Geroch R. P., Kronheimer E. H., Penrose R.*, Ideal points in space-time, Proc. Roy. Soc. London, **A327**, 545—567 (1972).
61. *Gibbons G., Penrose R.*, в печати.
62. *Gödel K.*, An example of a new type of cosmological solution of Einstein's field equations of gravitation, Rev. Mod. Phys., **21**, 447—450 (1949).
63. *Gold T.* (ed.), The Nature of Time, Cornell Univ. Press, Ithaca, 1967.
64. *Graves J. C., Brill D. R.*, Oscillatory character of Reissner — Nordström metric for an ideal charged wormhole, Phys. Rev., **120**, 1507—1513 (1960).
65. *Грищук Л. П.*, Некоторые замечания об особенностях в космологических решениях уравнений гравитации, ЖЭТФ, **51**, 475—479 (1966).
66. *Hajicek P.*, Causality in non-Hausdorff space-times, Comm. Math. Phys., **21**, 75—84 (1971).
67. *Hajicek P.*, Exact Models of Charged Black Holes, II. Axysymmetric stationary horisons, Comm. Math. Phys., **34**, 53—76 (1973).
68. *Harrison B. K., Thorne K. S., Wakano M., Wheeler J. A.*, Gravitation theory and gravitational collapse, Chicago Univ. Press, Chicago, 1965. (Русский перевод: *Дж. Уилер, Б. Гаррисон, М. Вакано, К. Торн*, Теория гравитации и гравитационный коллапс, изд-во «Мир», М., 1976.)
69. *Hartle J. B., Hawking S. W.*, Solutions of the Einstein-Maxwell equations with many black holes, Comm. Math. Phys., **26**, 87—101 (1972).
70. *Hartle J. B., Hawking S. W.*, Energy and angular momentum flow into a black hole, Comm. Math. Phys., **27**, 283—290 (1972).

71. *Hawking S. W.*, Perturbations of an expanding universe, Astrophys. J., **145**, 544—554 (1966).
72. *Hawking S. W.*, Singularities and the geometry of space-time, Adams Prize Essay, 1966, не опубликовано.
73. *Hawking S. W.*, The occurrence of singularities in cosmology, III. Causality and singularities, Proc. Roy. Soc. London, **A300**, 187—201 (1967).
74. *Hawking S. W., Ellis G. F. R.*, Singularities in homogeneous world models, Phys. Lett., **17**, 246—247 (1965).
75. *Hawking S. W., Penrose R.*, The singularities of gravitational collapse and cosmology, Proc. Roy. Soc. London, **A314**, 529—548 (1970).
76. *Heckmann O., Schücking E.*, Relativistic cosmology, *in* Gravitation: an Introduction to Current Research, ed. L. Witten, Wiley, New York, 1962, pp. 438—469.
77. *Hocking J. G., Young G. S.*, Topology, Addison-Wesley, London, 1961.
78. *Hodge W. V. D.*, The Theory and Application of Harmonic Integrals, Cambridge Univ. Press, London, 1952.
79. *Hogarth J. E.*, Cosmological considerations on the absorber theory of radiation, Proc. Roy. Soc. London, **A267**, 365—383 (1962).
80. *Hoyle F.*, A new model for the expanding universe, Mon. Not. Roy. Astron. Soc., **108**, 372—382 (1948).
81. *Hoyle F., Narlikar J. V.*, Time-symmetric electrodynamics and the arrow of time in cosmology, Proc. Roy. Soc. London, **A277**, 1—23 (1963).
82. *Hoyle F., Narlikar J. V.*, A new theory of gravitation, Proc. Roy. Soc. London. **A282**, 191—207 (1964).
83. *Israel W.*, Singular hypersurfaces and thin shells in General Relativity, Nuovo Cimento, **44B**, 1—14 (1966); erratum Nuovo Cimento, **49B**, 463 (1967).
84. *Israel W.*, Event horizons in static vacuum space-time, Phys. Rev., **164**, 1776—1779 (1967).
85. *Israel W.*, Event horizons in static electrovac space-time, Comm. Math. Phys., **8**, 245—260 (1968).
86. *Jordan P.*, Schwerkraft und Weltall, Friedrich Vieweg, Braunschweig, 1955.
87. *Kantowski R., Sachs R. K.*, Some spatially homogeneous anisotropic relativistic cosmological models, J. Math. Phys., **7**, 443—446 (1967).
88. *Kelley J. L.*, General Topology, van Nostrand, Princeton, 1965.
89. *Khan K. A., Penrose R.*, Scattering of two impulsive gravitational plane waves, Nature, **229**, 185—186 (1971).
90. *Kinnersley W., Walker M.*, Uniformly accelerating charged mass in General Relativity, Phys. Rev., **D2**, 1359—1370 (1970).
91. *Kobayashi S., Nomizu K.*, Foundations of Differential Geometry, Vol. I, Intersci., New York, 1963.
92. *Kobayashi S., Nomizu K.*, Foundations of Differential Geometry, Vol. II, Intersci., New York, 1969.
93. *Kreuzer L. B.*, Experimental measurement of the equivalence of active and passive gravitational mass, Phys. Rev., **169**, 1007—1012 (1968).
94. *Kronheimer E. H., Penrose R.*, On the structure of causal spaces, Proc. Camb. Phil. Soc., **63**, 481—501 (1967).
95. *Kruskal M. D.*, Maximal extension of Schwarzschild metric, Phys. Rev., **119**, 1743—1745 (1960).
96. *Kundt W.*, Trägheitsbahnen in einem von Gödel angegebenen kosmologischen Modell, Z. Phys., **145**, 611—620 (1956).
97. *Kundt W.*, Note on the completeness of space-times, Z. Phys., **172**, 488—489 (1963).
98. *Le Blanc J. M., Wilson J. R.*, A numerical example of the collapse of a rotating magnetized star, Astrophys. J., **161**, 541—552 (1970).
99. *Leray J.*, Hyperbolic differential equations, duplicated notes, Princeton Institute for Advanced Studies, 1952.

100. *Lichnerowicz A.*, Theories Relativistes de la Gravitation et de l'Electromagnétisme, Masson, Paris, 1955.
101. *Лифшиц Е. М., Халатников И. М.*, Investigations in relativistic cosmology, Adv, in Phys. (Phil. Mag. Suppl.), **12**, 185—249 (1963).
102. *Löbell F.*, Beispiele geschlossener drei-dimensionaler Clifford-Kleinsche Räume negativer Krümmung, Ber. Verhandl. Sächs. Akad. Wiss. Leipzig, Math, Phys. Kl., **83**, 167—174 (1931).
103. *Milnor J.*, Morse Theory, Annals of Mathematics Studies No. 51, Princeton Univ. Press, Princeton, 1963. (Русский перевод: *Дж. Милнор*, Теория Морса, изд-во «Мир», М., 1966.)
104. *Misner C. W.*, The flatter regions of Newman, Unti and Tamburino's generalized Schwarzschild space, J. Math. Phys., **4**, 924—937 (1963).
105. *Misner C. W.*, Taub — NUT space as a counterexample to almost anything, *in* Relativity Theory and Astrophysics, I: Relativity and Cosmology, ed. J. Ehlers, Lectures in Applied Mathematics, Vol. 8, American Math. Society, 1967, pp. 160—169.
106. *Misner C. W.*, The isotropy of the universe, Astrophys. J., **151**, 431—457 (1968).
107. *Misner C. W.*, Quantum cosmology. I, Phys. Rev., **186**, 1319—1327 (1969).
108. *Misner C. W.*, Minisuperspace, preprint, Univ. of Maryland, 1971.
109. *Мизнер Ч. В., Тауб А. Х.*, Пустая Вселенная, не имеющая особенностей, ЖЭТФ, **55**, 233—255 (1968).
110. *Müller zum Hagen H.*, On the analyticity of stationary vacuum solutions of Einstein's equations, Proc. Camb. Phil. Soc., **68**, 199—201 (1970).
111. *Müller zum Hagen H., Robinson D. C., Seifert H. J.*, Black holes in static vacuum space-times, *in* General Relativity and Gravitation (1972).
112. *Munkers J. R.*, Elementary Differential Topology, Annals of Math. Studies No. 54, Princeton. Univ. Press, Princeton, 1954.
113. *Newman E. T., Penrose R.*, An aproach to gravitational radiation by a method of spin coefficients, J. Math. Phys., **3**, 566—578 (1962).
114 *Newman E. T., Penrose R.*, New conservation laws for zero-rest mass fields in asymptotically flat space-time, Proc. Roy. Soc. London, **A305**, 175—204 (1968).
115. *Newman E. T., Tamburino L., Unti T. J.*, Empty space generalization of the Schwarzschild metric, J. Math. Phys., **4**, 915—923 (1963).
116. *Newman E. T., Unti T. W. J.*, Behaviour of asymptotically flat empty spaces, J. Math. Phys., **3**, 891—901 (1962).
117. *North J. D.*, The Measure of the Universe, Oxford Univ. Press, London, 1965.
118. *Ozsváth I., Schücking E.*, An anti-Mach metric, *in* Recent Developments in General Relativity, Pergamon Press — PWN, 1962, pp. 339—350.
119. *Papapetrou A.*, Champs gravitationnels stationnares à symmétrie axiale, Ann. Inst. Henri Poincaré, AIV, 83—105 (1966).
120. *Papapetrou A., Hamoui A.*, Surfaces caustiques dégénérées dans la solution de Tolman. La Singularité physique en Relativité Générale, Ann. Inst. Henri Poincaré, VI, 343—364 (1967).
121. *Peebles P. J. E.*, Primordial helium abundance and the primordial fireball. II, Astrophys. J., **146**, 542—552 (1966).
122. *Penrose R.*, Asymptotic properties of fields and space-times, Phys. Rev. Lett., **10**, 66—68 (1963).
123. *Penrose R.*, Conformal treatment of infinity, *in* Relativity, Groups and Topology, eds. C. M. de Witt and B. de Witt, Les Houches Summer School, 1963, Gordon and Breach, New York, 1964. (Русский перевод в сб. «Гравитация и топология», изд-во «Мир», М., 1965.)
124. *Penrose R.*, A remarkable property of plane waves in General Relativity, Rev. Mod. Phys., **37**, 215—220 (1965).
125. *Penrose R.*, Zero rest-mass fields including gravitation: asymptotic behaviour, Proc. Roy. Soc. London, **A284**, 159—203 (1965).

126. *Penrose R.*, Gravitational collapse and space-time singularities, Phys. Rev Lett., **14**, 57—59 (1965).
127. *Penrose R.*, General Relativity energy flux and elementary optics, *in* Perspectives in Geometry and Relativity (Hlavaty Festschrift), ed. B. Hoffmann, Indiana Univ. Press, Bloomington, 1966, pp. 259—274.
128. *Penrose R.*, Structure of space-time, *in* Battelle Rencontres, eds., C. M. de Witt, J. A. Wheeler, Benjamin, New York, 1968, pp. 121—235. (Русский перевод: Р. Пенроуз, Структура пространства-времени, изд-во «Мир», М., 1972.)
129. *Penrose R.*, Gravitational collapse: the role of General Relativity, Rivista del Nuovo Cimento, **1**, 252—276 (1969).
130. *Penrose R.*, The geometry of impulsive gravitational waves, *in* Studies in Relativity (Synge Festschrift), ed. L. O'Raiffeartaigh, Oxford Univ. Press, London, 1972.
131. *Penrose R.*, Techniques of differential topology in relativity, Lectures at Pittsburgh, 1970, A. M. S. Colloquium Publications, 1972.
132. *Penrose R.*, *MacCallum M. A. H.*, A twistor approach to space-time quantization, Phys. Rep. (Phys. Lett. Section C) (1972).
133. *Penrose R.*, *Floyd R. M.*, Extraction of rotational energy from a black hole, Nature, **229**, 177—179 (1971).
134. *Penzias A. A.*, *Schraml J.*, *Wilson R. W.*, Observational constraints on a discrete source model to explain the microwave background, Astrophys. J., **157**, L49—L51 (1969).
135. *Pirani F. A. E.*, On the energy — momentum tensor and the creation of matter in relativistic cosmology, Proc. Roy. Soc., **A228**, 455—462 (1955).
136. *Press W. H.*, Time evolution of a rotating black hole immersed in a static scalar field, Astrophys. J., **175**, 245—252 (1972).
137. *Price R. H.*, Nonspherical perturbations of relativistic gravitational collapse, I. Scalar and gravitational perturbations; II. Integer spin, zero rest-mass fields, Phys. Rev., **5**, 2419—2454 (1972).
138. *Rees M. J.*, *Sciama D. M.*, Large-scaledensity inhomogeneities in the universe, Nature, **217**, 511—516 (1968).
139. *Regge T.*, *Wheeler J. A.*, Stability of a Schwarzschild singularity, Phys. Rev., **108**, 1063—1069 (1957).
140. *Riesz F.*, *Sz-Nagy B.*, Functional Analysis, Blackie and Sons, London, 1955. (Русский перевод: *Ф. Рисс*, *Б. Секефальви-Надь*, Лекции по функциональному анализу, ИЛ, М., 1954.)
141. *Robertson H. P.*, Relativistic cosmology, Rev. Mod. Phys., **5**, 62—90 (1933).
142. *Rosenfeld L.*, Sur le tenseur d'impulsion — energie, Mem. Roy. Acad. Belg. Cl. Sci., **18**, No. 6 (1940).
143. *Ruse H. S.*, On the geometry of Dirac's equations and their expression in tensor form, Proc. Roy. Soc. Edin., **57**, 97—127 (1937).
144. *Sachs R. K.*, *Wolfe A. M.*, Perturbations of a cosmological model and angular variations of the microwave background, Astrophys. J., **147**, 73—90 (1967).
145. *Sandage A.*, The ability of the 200-inch telescope to discriminate between selected world models, Astrophys. J., **133**, 355—392 (1961).
146. *Sandage A.*, Observational cosmology, Observatory, **88**, 91—106 (1968).
147. *Schmidt B. G.*, Isometry groups with surface-orthogonal trajectories, Zs. Naturforsch., **22a**, 1351—1355 (1967).
148. *Schmidt B. G.*, A new definition of singular points in General Relativity, J. Gen. Rel. and Grav., **1**, 269—280 (1971).
149. *Schmidt B. G.*, Local completeness of the *b*-boundary, Comm. Math. Phys., **29**, 49—54 (1972).
150. *Schmidt H.*, Model of an oscillating cosmos which rejuvenates during contraction, J. Math. Phys., **7**, 494—509 (1966).
151. *Schouten J. A.*, Ricci Calculus, Springer, 1954.

152. *Schrödinger E.*, Expanding Universes, Cambridge Univ. Press, London, 1956.
153. *Sciama D. W.*, On the origin of inertia, Mon. Not. Roy. Astron. Soc., **113**, 34—42 (1953).
154. *Sciamo D. W.*, Peculiar velocity of the sun and the cosmic microwave background, Phys. Rev. Lett., **18**, 1065—1067 (1967).
155. *Sciama D. W.*, Astrophysical cosmology, *in* General Relativity and Cosmology, ed. R. K. Sachs, Proc. of the International School of Physics «Enrico Fermi», Course XLVII, Academic Press, New York, 1971, pp. 183—236.
156. *Seifert H. J.*, Global connectivity by timelike geodesics, Zs. Naturforsch., **22a**, 1356—1360 (1967).
157. *Seifert H. J.*, Kausal Lorentzräume, Doctoral Thesis, Hamburg Univ., 1968.
158. *Smart J. J. C.*, Problems of Space and Time, Problems of Philosophy Series, ed. P. Edwards, Collier-Macmillan, London; Macmillan, New York, 1964.
159. *Соболев С. Л.*, Некоторые применения функционального анализа в математической физике, Изд-во Сиб. отделения АН СССР, Новосибирск, 1962.
160. *Spanier E. H.*, Algebraic Topology, McGraw-Hill, New York, 1966.
161. *Spivak M.*, Calculus on Manifolds, Benjamin, New York, 1965. (Русский перевод: *М. Спивак.*, Математический анализ на многообразиях, изд-во «Мир», М., 1968.)
162. *Steenrod N. E.*, The Topology of Fibre Bundles, Princeton Univ. Press, Princeton, 1951. (Русский перевод: *Н. Стинрод*, Топология косых произведений, ИЛ, М., 1953.)
163. *Stewart J. M. S., Sciama D. W.*, Peculiar velocity of the sun and its relation to the cosmic microwave background, Nature, **216**, 748—753 (1967).
164. *Streater R. F., Wightman A. S.*, P. C. T., Spin, Statistics, and All That, Benjamin, New York, 1964.
165. *Thom R.*, Stabilité Structurelle et Morphogenése, Benjamin, New York, 1969.
166. *Thorne K. S.*, The General Relativistic theory of stellar structure and dynamics, *in* High Energy Astrophysics, ed. L. Gratton, Proc. of the International School in Physics «Enrico Fermi», Course XXXV, Academic Press, New York, 1966, pp. 166—280.
167. *Tsuruta S.*, The effects of nuclear forces on the maximum mass of neutron stars, *in* The Crab Nebula, eds. R. D. Davies, F. G. Smith, Reidel, Dordrecht, 1971.
168. *Vishveshwara C. V.*, Generalization of the «Schwarzschild Surface» to arbitrary static and stationary Metrics, J. Math. Phys., **9**, 1319—1322 (1968).
169. *Vishveshwara C. V.*, Stability of the Schwarzschild metric, Phys. Rev., **D1**, 2870—2879 (1970).
170. *Wagoner R. V., Fowler W. A., Hoyle F.*, On the synthesis of elements at very high temperatures, Astrophys. J., **148**, 3—49 (1968).
171. *Walker A. G.*, Completely symmetric spaces, J. Lond. Math. Soc., **19**, 219—226 (1944).
172. *Weymann R. A.*, Mass loss from stars, *in* Ann. Rev. Astron. and Astrophys., Vol. 1, Ann. Rev. Inc., Palo Alto, 1963, pp. 97—141.
173. *Wheeler J. A.*, Superspace and the nature of quantum geometrodynamics, *in* Batelle Rencontres, eds. C. M. de Witt, J. A. Wheeler, Benjamin, New York, 1968, pp. 242—307.
174. *Whitney H.*, Differentiable manifolds, Annals Math., **37**, 645 (1936).
175. *Yano K., Bochner S.*, Curvature and Betti numbers, Annals Math., Studies No. 32, Princeton Univ. Press, Princeton, 1953. (Русский перевод: *К. Яно, С. Бохнер*, Кривизна и числа Бетти, ИЛ, М., 1957.)
176. *Зельдович Я. Б., Новиков И. Д.*, Релятивистская астрофизика, изд-во «Наука», М., 1967

Предметный указатель

Атлас 20, 70
— локально-конечный 22
— ориентированный 35
— полный 20
— совместный с данным атласом 20
Аффинный параметр 43, 287
— — обобщенный 288, 289

Базис 25
— дуальный 26
— координатный 24
— ортонормированный
— псевдоортонормированный 100, 382
Бесконечность будущего (прошлого) времениподобная 137, 138, 145
— — изотропная 137, 145
— — пространственноподобная 137, 138
Бианки тождества 47, 54
Бранса — Дикке теория 72
Будущее асимптотически предсказуемое 346
— сильно предсказуемое 348
— захваченное 216, 217
— причинное 203
— хронологическое 203, 241

Вариация кривой 121, 122
— метрики 78
— поля 77
— связности 79
Вектор вариации 121—123
— времениподобный 49, 69
— изотропный 49, 69
— касательный к C^k-кривой 24
— Киллинга 55
— ковариантный 25
— поток 82
— пространственноподобный 49, 69
Вложение 33, 55
Вращение 95
— изотропных геодезических 101

Геодезическая 44
— изотропная 99
— полная 44, 212
Геодезическая кривая 43
— — максимальная 125, 126
Гиперповерхность 55
— времениподобная 56
— изотропная 56
— пространственноподобная 56
ГНБ (граничное НБ) 243—245, 249, 250
ГНП (граничное НП) 243—245, 249, 250
Горизонт кажущийся 356, 357—359
— Коши 178, 198, 320
— событий 146
— — бесконечности в будущем 185
— частиц 144
Граница ахрональная 208
— пространства-времени 244
Группа изометрии 55, 153, 367
— транзитивная 137, 165, 188

Давление 83
— вырожденных нейтронов 339
— — электронов 337
Девиация кривых 92
— геодезических 101
Диаграмма Пенроуза 138, 139
Диффеоморфизм 33, 34
Длина пути вдоль кривой 47, 48

Задача Коши для материальных полей 72, 73
— — для уравнения Эйнштейна 253—258
— — устойчивость решения 281, 334
Захват 216
Захваченная неполнота 322, 323

Идеальная жидкость 81, 83, 92
— — изоэнтропическая 83, 85
Излучение космическое фоновое 387—389
Изотропный конус 49, 53, 73

Карта локальная 20
Киллинга бивектор 187
— вектор 55
— векторное поле 55, 74, 153, 187

Коллапс звезды 333, 334
Компоненты тензора 28
Комплект (начальных данных) 258, 276, 278
Конгруэнция кривых 81
Контур будущего 204
Конформный множитель 53, 74, 76, 77
Координаты изотропные 133
-- локальные 20
— нормальные 44, 52
Коперника принцип 151, 152, 167, 168, 387, 389
Космологическая постоянная 86, 109, 154, 159, 402
Край многообразия 20, 246
Кривая времениподобная 81
— интегральная 37
— класса C^k (C^k-кривая) 24
— направленная в будущее 205
— непродолжимая в будущее (прошлое) 205
— непространственноподобная 72
— предельная 205
— с захваченным будущим 216, 217
Кристоффеля соотношения 52

Лагранжиан заряженного скалярного поля 77, 81
— идеальной жидкости 81
— скалярного поля 80
— электромагнитного поля 80
Лифт 65
— горизонтальный 66
Локальная причинность (постулат) 72, 131, 210, 265
Локальное сохранение энергии и импульса (постулат) 74

Метрика 47—55, 68
— индуцированная 56
— конформная 53, 73—77, 89, 90, 200
— лоренцева 49—52, 56, 68, 90, 119, 286
— невырожденная 48
— положительно-определенная 49
— пространственно-временная 68, 84
— статическая 85
— фоновая 253—255, 271—276, 280—281
Многообразие 19, 69
— геодезически полное 44, 216, 286, 289, 292, 296
— класса C^r (C^r-многообразие) 20
— — — с краем 20
— накрывающее 228
Многообразие ориентируемое 22, 56, 60, 65
— паракомпактное 22, 44, 49—50
— параллелизуемое 65, 202
— причинно простое 210, 229, 295
— связное 68
— универсальное накрывающее 228
— хаусдорфово 22, 44, 69, 193, 194, 277, 278
Множество ахрональное 207, 225—227, 232, 297
— будущего 207, 208, 209
— вполне упорядоченное 276
— глобально гиперболическое 229, 230—236
— ловушечное для будущего (прошлого) 297, 299
— нарушающее хронологию 210, 211
— непричинное 208, 227
— причинно простое 210, 229
Модель Вселенной Гёделя 187, 188—190
— статическая Эйнштейна 137, 156
— стационарная 141
Модель пространства-времени математическая 68
— — — геодезически полная 33, 165
— — — изометричная 68
— — — — локально C^r-непродолжимая 71
— — — — локально C^r-нерасширяемая 71
— — — — C^r-нерасширяемая 71
НБ (неразложимое множество будущего) 241, 242—245
НП (неразложимое множество прошлого) 241, 242—245

Область Коши 135, 150, 166, 224, 232
— — будущего 224
— ловушечная 356
b-Ограниченность 325, 327
Окрестность выпуклая 44
— локальной причинности 217, 219
— нормальная 44
Отображение инъективное 32
— класса C^r (C^r-отображение) 19, 31
— конформное 53, 157
— собственное 33
— сюръективное 32, 62
— тензоров 32, 34
— экспоненциальное 44, 134

Перенос параллельный 213
— Ферми 94

Плоские волны (гравитационные) 199, 229
Поверхность внешняя ловушечная 355
— — — маргинальная 359
— замкнутая ловушечная 11, 174, 292, 297, 334
— Коши 135, 150, 199, 228, 295
— пространственноподобная 113, 115
— стационарности предельная 185
— транзитивности 151, 152
Погружение 32
Подмногообразие вложенное 33
— погруженное 33
Подпространство вертикальное 66
— горизонтальное 66, 308
Поле линейных элементов 50, 212
— скалярное 80, 109
— — заряженное 81
— электромагнитное 80, 81
— якобиево 110, 125
— — вдоль изотропной геодезической 114
Полнота 289, 308
— геодезическая 286, 289
— метрическая 286, 308
Последовательность сильно сходящаяся 269
— слабо сходящаяся 270
— фундаментальная 286
Произведение внешнее прямое (многообразий) 93, 62
— тензорное 27
Производная внешняя 34, 35, 45
— ковариантная 41, 42
— — вдоль кривой 42
— Ли 37—40, 45
— Ферми 93, 94
Пространство дуальное касательному 25
— евклидово 10
— касательное 25
— компактное 40
— m-полное (метрически полное) 286, 308
— постоянной кривизны 140
— расслоенное 62
— Соболева 259, 260
— Тауба — НУТ 190, 191—198, 229, 322
— типа Бианки I 160
Пространство-время 68
— асимптотически простое 246, 247—250
— — — пустое 246, 247—250
— — — в слабом смысле 251, 345
— — сильно предсказуемое 348, 353
— геодезически полное 286, 289, 292, 296
— де Ситтера 1-го рода 140, 141—148
— — — 2-го рода 148, 149—151, 229
— компактное 189
— Минковского 133, 134—140, 247
— b-ограниченное 325—327
— ориентируемое по времени 201, 202, 212, 246
— b-полное 289, 308, 310, 326
— пространственно-ориентируемое 202
— расслоенное 62, 63
— регулярно предсказуемое 354, 356, 360
— — — статическое 363
— — — стационарное 360, 361, 366, 371
— Робертсона — Уокера 152, 153—159, 390—398
Прошлое асимптотически простое 352

Развитие (начальных данных) 253, 273—277
— максимальное 276—281
Разложение единицы 23, 36
Расхождение 96, 98
— изотропных геодезических 101, 108
— объемное 91, 111
Расслоение 62
— касательное 63
— линейных реперов 64, 194
— ортонормированных базисов 64
— тензорное 64
Расширение многообразия 70, 169, 172, 177, 183, 253, 276—281
C^r-Расширение 70, 171
Решение (уравнений Эйнштейна) 132
— Керра 180
— пространственно-однородное 160
— Райсснера — Нордстрема 175, 176—186, 229
— точное 132
— Шварцшильда 167, 166—174, 363, 412

Свертка тензора 28
Связность 40, 41, 52, 53, 65
— без кручения 45
— класса C^r (C^r-связность) 41
— плоская 47
Сигнатура метрики 48, 49
Сингулярность «голая» 346
— кривизны в параллельно перенесенном базисе 290, 323

— скалярных полиномов кривизны 290
Скалярная кривизна 52
СНБ (СНП) 243, 245
Собственно-разрывное действие группы 193, 194

Тензор 26
— антисимметричный 30
— Вейля 53, 54, 98, 102
— вращения 95
— кручения 45
— метрический 47
— параллельный перенос 43
— поперечного сдвига 96
— расхождения 96
— Римана (кривизны) 46, 47, 52—54, 93, 98, 99
— — на гиперповерхности 59
— Риччи 47, 52, 98
— свертка 28
— симметричный 30
— сложение 27
— умножение на скаляр 27
— фундаментальный второй 113
— — первый 113
— энергии-импульса 78—80, 102—110
— — — идеальной жидкости 83
— — — канонические формы (типы) 102, 103
— — — скалярного поля 80, 81
— — — электромагнитного поля 81
Теорема Биркхофа 413
— Гаусса 62
— Стокса 35
— — обобщенная 36
Теория относительности общая 68, 72, 403
— — специальная 68, 72, 74, 133
Топология 19, 220
— Александрова 219, 220
— лоренцевых метрик 220, 280
— открытая 221
— пространства кривых 230, 231, 238
Точка конечная 205, 209, 226
— предельная 205, 207
— сопряженная поверхности 113, 131
— — точке 111, 112—114, 125, 130

Уравнение Гаусса 59
— Кодацци 59
— Райчаудхури 97, 98, 101, 111
— Якоби (девиации) 93, 110
Уравнения поля (постулат) 90
Уравнения Эйлера — Лагранжа 78
— Эйнштейна 87, 88, 89, 131, 252
— — приведенные 255, 270—276
Условие, выделяющее будущее 215, 216
— времениподобного схождения 108, 319, 398
— изотропного схождения 108, 109, 319, 398
— Липшица 19, 70, 238
— причинности 211, 212—216, 241, 398
— сильной причинности 215, 216—219, 231—233, 302
— типовое 115, 213, 296, 359
— устойчивой причинности 221
— хронологическое 210, 297
— энергетическое сильное 109, 154, 334, 398
— — слабое 102, 103—105, 334
— энергодоминантности 104, 360
Условия калибровочные 255
— — гармонические 255
Устойчивость решения задачи Коши 281, 334

Форма линейная 25, 56, 57
— каноническая 60
— фундаментальная вторая 58
1-форма 30
q-форма 30, 34
Функция на многообразии 23
— — — класса C^r (C^r-функция) 23

Черная дыра 351, 352—359, 368—370
— — вращающаяся 362—370
— — слияние 370—373

Шварцшильда радиус 334
Шварцшильдова длина 393

Эйлера характеристика 51, 65
Эйнштейна оператор 255
Эргосфера 185, 364—366, 369

Оглавление

Предисловие к русскому изданию 5

Предисловие 7

Глава 1. Роль тяготения 9

Глава 2. Дифференциальная геометрия 18

2.1. Многообразия 19
2.2. Векторы и тензоры 23
2.3. Отображение многообразий 31
2.4. Внешнее дифференцирование и производная Ли 34
2.5. Ковариантная производная и тензор кривизны 40
2.6. Метрика 47
2.7. Гиперповерхности 55
2.8. Элемент объема и теорема Гаусса 60
2.9. Расслоенные пространства 62

Глава 3. Общая теория относительности 68

3.1. Пространственно-временное многообразие 68
3.2. Материальные (негравитационные) поля 71
3.3. Лагранжева формулировка 77
3.4. Уравнения поля 84

Глава 4. Физический смысл кривизны 91

4.1. Времениподобные кривые 91
4.2. Изотропные кривые 99
4.3. Энергетические условия 102
4.4. Сопряженные точки 110
4.5. Вариация длины дуги 116

Глава 5. Точные решения 132

5.1. Пространство-время Минковского 133
5.2. Пространство-время де Ситтера 1-го и 2-го рода 140
5.3. Пространство Робертсона — Уокера 151
5.4. Пространственно-однородные космологические модели 160
5.5. Решения Шварцшильда и Райсснера — Нордстрема 167
5.6. Решение Керра 180
5.7. Модель Вселенной Гёделя 187
5.8. Пространство Тауба — НУТ 190
5.9. Прочие точные решения 198

Глава 6. Причинная структура 200

6.1. Ориентируемость 201
6.2. Причинные кривые 202

6.3. Ахрональные границы 207
6.4. Условия причинности 210
6.5. Области Коши 224
6.6. Глобальная гиперболичность 229
6.7. Существование геодезических 236
6.8. Причинная граница пространства-времени 241
6.9. Асимптотически простые пространства 245

Глава 7. Задача Коши в общей теории относительности 251

7.1. Характер задачи 252
7.2. Приведенные уравнения Эйнштейна 253
7.3. Начальные данные 256
7.4. Гиперболические уравнения второго порядка 259
7.5. Существование и единственность решения задачи Коши для уравнений Эйнштейна для пустого пространства 270
7.6. Максимальное развитие и устойчивость 276
7.7. Уравнения Эйнштейна при наличии материи 282

Глава 8. Сингулярности пространства-времени 285

8.1. Определение сингулярностей 285
8.2. Теоремы о сингулярностях 291
8.3. Описание сингулярностей 307
8.4. Характер сингулярностей 316
8.5. Захваченная неполнота 321

Глава 9. Гравитационный коллапс и черные дыры 333

9.1. Коллапс звезды 333
9.2. Черные дыры 343
9.3. Конечное состояние черной дыры 359

Глава 10. Начальная сингулярность во Вселенной 387

10.1. Расширение Вселенной 387
10.2. Природа и проявление сингулярностей 400

Приложение А (Пьер Симон Лаплас) 406

Приложение Б. Сферически-симметричные решения и теорема Биркхофа 410

Обозначения 415

Литература 419

Предметный указатель 426

Уважаемый читатель!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и др. просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., дом 2, изд-во «Мир».

С. Хокинг, Дж. Эллис
КРУПНОМАСШТАБНАЯ СТРУКТУРА ПРОСТРАНСТВА-ВРЕМЕНИ

Редактор Л. В. Самсоненко. Художник П. М. Пучков
Художественный редактор В. М. Сафаров
Технический редактор Т. А. Максимова Корректор Н. И. Баранова
Сдано в набор 28/X 1976 Подписано к печати 5/V 1977. Бумага тип. № 2 60×90¹/₁₆=13,5 бум. л.
27 печ. л. Уч.-изд. л. 26,64. Изд № 27/8446. Цена 3 р. 40 к. Заказ № 381.

Издательство «Мир», Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени Ленинградская типография № 2 имени Евгении Соколовой Союзполиграфпрома при Государственном комитете Совета Министров СССР по делам издательств, полиграфии и книжной торговли
198052, Ленинград, Л-62, Измайловский проспект, 29