САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

Головнев Алексей Валерьевич

Модифицированные теории гравитации в космологическом контексте

Специальность 01.04.02 – "теоретическая физика"

Диссертация на соискание ученой степени
доктора физико-математических наук

Санкт-Петербург – 2017

Оглавление

Введение

Гравитационное взаимодействие знакомо каждому человеку с раннего детства, а связанные с ним закономерности движения планет были детально исследованы уже очень давно. Однако же его изучение и до сих пор представляет собой неиссякаемый источник побед и разочарований для физиков всего мира. Разочарования представлены по большей части безуспешными попытками описания гравитации в квантовом режиме, что впрочем относится больше к вопросам академического характера, поскольку эффекты квантовой гравитации находятся пока ещё бесконечно далеко от доступных экспериментаторам горизонтов. Одним же из самых потрясающих успехов оказалась возможность построения космологических моделей на научной основе, когда гравитационное взаимодействие было описано на языке геометрии пространства-времени в знаменитых работах Альберта Эйнштейна.

Использование общей теории относительности Эйнштейна в космологических масштабах позволяет, с одной стороны, получить блестяще работающую стандартную космологическую модель, но с другой стороны приводит и к новым разочарованиям: мы вынуждены вводить загадочные Тёмные Сектора в теорию – Тёмную Материю, Тёмную Энергию, а также изучать феноменологические модели инфляции без понимания фундаментальной природы инфлатона. Конечно, Тёмную Энергию можно интерпретировать просто как присутствие космологической постоянной, но её чрезвычайно малое значение представляется очень технически неестественным (в смысле квантовой теории поля).

Складывается довольно странная ситация, и положение ещё больше усугубляется тем, что поиски частиц Тёмной Материи (пока?) не приводят к успеху. Сам собой возникает вопрос: может быть, дело не в неполноте нашего знания физики элементарных частиц? Может быть, теория гравитации нуждается в корректировке? Ясно, что ни результаты многочисленных лабораторных экспериментов, ни детальное изучение динамики Солнечной Системы или двойных пульсаров не могут гарантировать, что в галактических и космологических масштабах теория гравитации не должна быть модифицирована для лучшего соответствия с экспериментальными (или в данном контексте лучше сказать – наблюдательными) данными.

Тем самым возникает (весьма обширное) направление в современной теоретической физике – изучение модифицированных теорий гравитации (modified gravity). Его целью является выяснение того, как можно описать гравитацию в рамках модифицированной теории так, чтобы и не войти в противоречие с имеющимися экспериментальными данными, и предложить лучшее описание достаточно широкого круга явлений в космологии. В принципе, подобную программу можно проводить в жизнь и с точки зрения проблем квантования теории гравитационного взаимодействия. Но в данной Диссертации мы будем последовательно придерживаться космологической точки зрения, более близкой к непосредственным наблюдениям и экспериментам.

Заметим, что другой входящий сейчас в моду термин для этого направления – расширенные (extended) теории гравитации. Это связано с тем, что многие классы таких теорий могут быть заменой переменных сведены к стандартной общей теории относительности с дополнительными полями. Материя при этом взаимодействует с модифицированной метрикой, зависящей от вспомогательных полей – например, с метрикой, конформно растянутой функцией дополнительного скалярного поля, в скалярно-тензорном представлении $f(R)$ гравитации. Универсальность взаимодействия всех полей материи с дополнительными полями делает интерпретацию на языке модифицированной гравитации более естественной.

В данной Диссертации рассматривается ряд вопросов, относящихся к теориям модифицированной гравитации и теоретическим аспектам космологии. В их число входят такие темы как инфляция с векторными полями, массивная и биметрическая гравитация, телепараллельные подходы к модифицированной гравитации, а также некоторые другие темы.

В первой главе мы даём краткое введение в основы современной теории гравитации и теоретической космологии, стремясь в первую очередь к обоснованию тезиса о необходимости проведения исследований в области модифицированной гравитации.

Во второй главе представлены работы по векторной инфляции – попыткам построения моделей ускоренного расширения Вселенной с помощью векторных полей. Показано, что эти модели являются неустойчивыми: продольная компонента векторного поля оказывается духом, в широком классе моделей возникает гравитационно-волновая неустойчивость и катастрофический рост анизотропии, существует дополнительная степень свободы, для которой космологический фон является режимом сильной связи. Обсуждается современное состояние вопроса, а также проблемы работы с неканоническими векторными полями. Поскольку в моделях векторной инфляции использовалось неминимальное взаимодействие векторных полей с гравитацией, их тоже можно в некоторой степени считать моделями модифицированной гравитации.

Третья глава посвящена современным моделям массивной и биметрической гравитации – теориям де Рам - Габададзе - Толли. Краеугольным камнем этих моделей является квадратный корень из матрицы, составленной из физической и опорной метрик. Эта конструкция позволяет избежать появления духа Боулвара-Дезера (шестой поляризации массивного гравитона, обладающей отрицательной кинетической энергией) в случае, если потенциал самодействия выбран в виде симметрического полинома по собственным значениям матрицы квадратного корня.

Мы рассматриваем несколько аспектов массивной гравитации. Во-первых, это проблема непертурбативного доказательства отсутствия духа Боулвара-Дезера, к которой мы предложили свой подход на относительно раннем этапе развития этих работ. Во-вторых, нами осуществлён детальный анализ проблем, связанных с извлечением матричного квадратного корня, а также дан новый подход к построению теории возмущений, использующий полиномы от собственных значений вместо самих матриц при извлечении квадратного корня и обладающий большей областью применимости, чем стандартный подход. Наконец, в третьих, показано (на языке космологических возмущений) существование духа Боулвара-Дезера в одном из модифицированных вариантов массивной гравитации – в расширенном квазидилатоне.

В четвёртой главе рассматриваются биметрические модели другого типа. Мы вводим вторую метрику для определения связности, которая тем самым оказывается отличной от связности Леви-Чивита физической метрики. Построен формализм для непертурбативного АДМ анализа подобных моделей, показано, что в случае отсутствия ограничений на вторую метрику неизбежно возникают духи.

Интересным примером нетривиального ограничения, накладываемого на вторую метрику, являются С- (и D-) модели Амендолы - Энквиста - Койвисто. Аффинная структура в этих моделях задаётся метрическим потенциалом (второй метрикой), который отличается от физической метрики, измеряющей расстояния, конформным (C) или дисформным (D) преобразованием, зависящим от кривизны. Нами показано, что эти модели нуждаются в доопределении, но при этом могут быть связаны с интересными классами нелокальных теорий гравитации.

Отметим, что нелокальные теории гравитации являются частным случаем интересного, хотя и потенциально проблематичного класса моделей, с производными бесконечного порядка в действии. Подобные модели активно изучаются в ряде современных работ, привлекая всё большее внимание в теоретических исследованиях по гравитации и космологии, а также теории струн. Предложено множество интересных вариантов, что однако же не исключает сомнений в самосогласованности подобных подходов со стороны многих специалистов. Поэтому изучение фундаментальных аспектов нелокальных теорий представляется на данном этапе весьма интересным, хотя их непосредственное рассмотрение с незивисимой (от биметрического подхода) точки зрения выходит за рамки данной Диссертации.

Наконец, в пятой главе приведены результаты нескольких работ в рамках того же широкого направления исследований, но не подпадающих под узкую тематику предыдущих глав. С нашей точки зрения наиболее значимые из них – это исследование модифицированной телепараллельной гравитации в связи с вопросами локальной лоренц-инвариантности в пространстве тетрад, а также изучение модели, известной в литературе под именем mimetic dark matter, которую можно воспринимать как биметрическую (или скалярно-тензорную – в зависимости от выбранной формулировки) теорию особого вида, не принадлежащую к классам, рассмотренным в третьей и четвёртой главах.

Также в пятой главе, с помощью обращения к парадигме МОНД, приведено обсуждение взаимоотношений стандартной космологической модели, включающей в себя Тёмную Материю, с подходами в рамках модифицированной гравитации. Кроме того, в качестве реакции на некоторые работы, предлагающие использовать в космологии на бранах дополнительные квантовые потенциалы, возникающие в рамках решения задачи о движении квантовой частицы со связями, мы рассматриваем методы квантования систем со связями второго рода, показывая, что без подключения дополнительных соображений они не позволяют дать однозначного ответа для квантового потенциала. Наконец, в последнем разделе приведено краткое обсуждение некоторых фундаментальных проблем квантовой гравитации на примере информационного парадокса в физике Чёрных Дыр.

Актуальность темы исследования

Как уже упоминалось выше, современная теоретическая космология – развитая научная дисциплина, основанная на общей теории относительности. Она находится в прекрасном согласии с наблюдениями. Однако же есть ряд загадок, включающий в себя проблему космологической постоянной, тёмные сектора, природу инфлатона... Это приводит к тому, что исследования в области альтернативных (модифицированных, расширенных) теорий гравитации и их космологических приложений идут очень активно. Каждую неделю появляется целый ряд научных статей по этой теме.

Работы, положенные в основу данной Диссертации, проливают свет на фундаментальные свойства популярных способов модифицирования (расширения) гравитации – биметрических формализмов, теорий с кручением, необычных скалярно-тензорных расширений типа mimetic dark matter, нелокальных теорий. В перспективе это позволит продолжить обсуждение феноменологических аспектов и особенностей построения моделей в рамках космологии с модифицированной гравитацией с новых точек зрения.

Особо следует отметить, что космология – это наука, в которой, несмотря на очевидные сложности, имеется как достаточное количество наблюдательных данных, так и все основания ожидать, что их поток отнюдь не иссякнет в обозримом будущем: изучение поляризации реликтового фона, составление галактических каталогов, применение методов слабого гравитационного линзирования, работа с лесами лайман-альфа и линией сверхтонкой структуры водорода (21 cm) – все эти направления активно развиваются, находясь на разных этапах своего развития. Недавнее открытие гравитационных волн позволяет рассчитывать на относительно скорое развитие гравитационно-волновой астрономии.

Эта уникальная возможность для проверки фундаментальных теорий весьма успешно привлекает к себе внимание исследователей, ранее работавших в других областях теоретической физики, таких как физика элементарных частиц. Достаточно вспомнить Стивена Вайнберга, или отметить, что Эдвард Виттен недавно стал соавтором статьи о возможной природе Тёмной Материи (сверхлёгкие аксионоподобные поля).

Разработанность темы исследования

В современной теоретической физике активно развивается целый ряд направлений, предлагающих описание космологической эволюции в рамках модифицированных (расширенных) теорий гравитации. Общее состояние исследований таково, что на данный момент трудно выделить глобально предпочтительные направления, а конкретные подходы имеют разные степени разработанности и успеха.

Тематика, связанная с векторной инфляцией, практически не существовала до появления наших работ. На сегодняшний день эта тема является весьма разработанной. И хотя исходное предложение оказалось, строго говоря, нежизнеспособным, сейчас существует некоторое количество более изощрённых способов использования векторных полей, возникновение которых было тесно связано с нашими работами.

По массивной и биметрической гравитации де Рам - Габададзе - Толли в последние примерно семь лет существует огромный пласт литературы, исчисляемый многими сотнями наименований. Детально разработаны вопросы гамильтонова анализа в метрическом и тетрадном формализмах, изучены многие обобщения теории и космологические решения вместе с (не всегда полным) анализом возмущений. Однако многие фундаментальные аспекты до сих пор остаются terra incognita. В частности, общие вопросы неоднозначности извлечения квадратных корней были практически не изучены до появления нашей работы.

Степень разработанности других версий модифицированной гравитации, которые мы рассматриваем в данной Диссертации, вариирует от весьма глубокой до совсем недавно появившихся моделей. Мы будем комментировать более подробно каждый случай в началах соответствующих глав и разделов.

Цели работы

Целью данной работы является изучение модифицированных теорий гравитации и их космологических приложений. Это даёт новый взгляд на проблемы и загадки космологии, позволяет лучше понять особенности гравитационного взаимодействия, а также в перспективе приближает нас к решению вопроса о том, какова же вероятная природа тёмных секторов – расширение стандартной модели физики элементарных частиц или же модификация гравитационного взаимодействия (впрочем, как уже упоминалось выше, в буквальном смысле это деление носит условный характер).

Для достижения этой глобальной цели решаются следующие более локальные задачи:

- Построить модели инфляции с векторным инфлатоном и проверить их устойчивость по отношению к возмущениям.
- Убедившись в отсутствии таковой, выяснить свойства возникающих неустойчивостей и исследовать возможности построения более сложных моделей.
- Провести гамильтонов анализ моделей массивной гравитации и установить наличие или отсутствие духа Боулвара-Дезера как в исходной модели, так и в её расширениях, таких как расширенный квазидилатон.
- Установить роль, которую играет неоднозначность извлечения квадратного корня из матрицы в теориях массивной гравитации, и найти удобный метод построения теории возмущений с учётом этих особенностей.
- Построить общие методы гамильтонова анализа биметрических теорий, в которых тензор кривизны порождён связностью Леви-Чивита вспомогательной метрики.
- Найти удобные методы описания биметрических теорий Амендолы - Энквиста - Койвисто, в которых две метрики связаны конформным или дисформным соотношением, зависящим от кривизны.
- Исследовать проблему локальной лоренц-инвариантности в телепараллельной гравитации и выяснить на этой основе роль, которую играет спин-связность в модифицированных теориях типа $f(T)$.
- Дать удобное скалярно-тензорное описание модели миметической тёмной материи (mimetic dark matter).

Основные положения, выносимые на защиту

1. Построены принципиально новые модели векторной инфляции, успешные на уровне фоновых уравнений движения.

2. Установлена неустойчивость векторной инфляции по отношению к возмущениям. Помимо ранее обнаруженного духа в продольной компоненте векторного поля, найдена неустойчивость части моделей по отношению к рождению гравитационных волн и соответствующему росту анизотропии, а также указана дополнительная степень свободы, в космологических решениях находящаяся в режиме сильной связи.

3. Показано, что известные проблемы неканонических векторных полей, связанные с отклонениями от гиперболичности, могут отсутствовать вокруг космологических решений, что представляет интерес для современных космологических моделей с векторными полями.

4. Предложен новый способ доказательства отсутствия духа Боулвара-Дезера в моделях массивной гравитации де Рам - Габададзе - Толли.

5. Дана новая формулировка теории массивной гравитации, работающая непосредственно в терминах симметрических полиномов от собственных значений, а не самих матриц, из которых извлекается квадратный корень. Установлено, что данный подход хорошо работает даже в тех случаях, когда теория возмущений в терминах матриц вообще не определена.

6. Доказано наличие духа Боулвара-Дезера в массивной гравитации с расширенным квазидилатоном при изучении космологических возмущений.

7. Разработан гамильтонов формализм для анализа биметрических теорий со связностью, порождаемой вспомогательной метрикой.

Показано, что теория с двумя полностью независимыми метриками страдает наличием духовых степеней свободы, в то время как модели с наложением различных соотношений между метриками могут быть последовательно построены и представляют интерес.

8. Предложено описание биметрических моделей Амендолы - Энквиста - Койвисто на языке любой из двух метрик. Показано, что эти модели нуждаются в доопределении, но также продемонстрирована их связь с нелокальными теориями гравитации.

9. Дан подробный анализ проблемы локальной лоренц-инвариантности в телепараллельной гравитации и модифицированных телепараллельных теориях. Выяснена роль плоской спин-связности в вариационном принципе телепараллельных теорий гравитации. Получена новая форма уравнений движения для $f(T)$ гравитации.

10. Предложена новая скалярно-тензорная формулировка модели миметической тёмной материи.

Научная новизна и значимость работы

Все перечисленные выше положения, выносимые на защиту, основаны на результатах, полученных впервые. Более подробно положение наших работ в ряду других будет сформулировано (с соответствующими ссылками) в каждой главе. Кратко новизна и значимость могут быть описаны следующим образом.

Идея использования векторных полей для построения моделей инфляции появлялась до нас лишь в нескольких работах, которые были практически малоизвестны. Нам удалось вызвать сильный резонанс в профессиональном сообществе работой [1*] из списка публикаций, приведённого в Заключении. К сожалению, модели векторной инфляции оказались неустойчивыми, причём наши исследования также внесли свой вклад в установление этого факта. Однако общая значимость этих работ не ограничивается одними лишь нежизнеспособными моделями векторной инфляции, поскольку инициированная нашей статьёй [1*] активность теоретиков была непосредственным источником появления моделей инфляции с векторной примесью, кинетически взаимодействующей со скалярным инфлатоном ($f(\phi)F^2$), а также моделей калибровочной инфляции (gauge-flation) с неабелевыми полями. Исследование этих моделей не потеряло своей актуальности.

Модели массивной гравитации оказались очень популярными и многообещающими, хотя и не лишены своих проблем. Наше независимое доказательство отсутствия духа было предложено во времена бурного развития гамильтонова анализа массивной гравитации, и впоследствии с успехом использовалось для исследований на языке полей Штюкельберга другими авторами. Изучение вопросов неоднозначности извлечения квадратного корня является фундаментально важным для понимания оснований теории, и никем до нас в полном объёме проведено не было. Одним из результатов этой работы стало построение новой формулировки теории (симметрические полиномы от собственных значений вместо непосредственно матриц), которая пригодна для использования даже тогда, когда изначальный способ описания становится непригодным.

Работы в области других моделей модифицированной гравитации также проводились в связи с актуальными вопросами, встающими перед исследователями, и предлагали новые результаты. Отдельно стоит отметить изучение вопросов лоренц-инвариантности в телепараллельной гравитации, которые являются источником многих заблуждений и путаницы в имеющейся литературе.

Апробация работы

Основные результаты Диссертации докладывались на международных конференциях: XX Workshop Beyond the Standard Model (Бад Хоннеф, Германия, 2008); 3. Kosmologietag (Билефельд, Германия, 2008); The Jubilee 40th Symposium on Mathematical Physics "Geometry & Quanta"(Торунь, Польша, 2008); Third School and Workshop on "Mathematical Methods in Quantum Mechanics"(Брессаноне, Италия, 2009); Spontaneous Workshop III «New topics in Modern Cosmology» (Кар-

жес, Франция, 2009); International Workshop on "Cosmic Structure and Evolution"(Билефельд, Германия, 2009); 5. Kosmologietag (Билефельд, Германия, 2010); Dual year Russia-Spain, Particle Physics, Nuclear Physics and Astroparticle Physics (Барселона, Испания, 2011); NEB15 – Recent Developments in Gravity (Ханья, Греция, 2012); 7th Mathematical Physics Meeting: Summer School and Conference on Modern Mathematical Physics (Белград, Сербия, 2012); VIIIth Iberian Cosmology Meeting (Гранада, Испания, 2013); The XXI International Workshop on High Energy Physics and Quantum Field Theory (Репино, Санкт-Петербург, 2013); The sixth Petrov International Symposium on High Energy Physics, Cosmology and Gravity (Киев, Украина, 2013); II Russian-Spanish Congress on Particle and Nuclear Physics at all Scales and Cosmology (Санкт-Петербург, 2013); IX International Workshop "Dark Side of the Universe"(Триест, Италия, 2013); 8th Mathematical Physics Meeting: Summer School and Conference on Modern Mathematical Physics (Белград, Сербия, 2014); III Russian-Spanish Congress on Particle, Nuclear, Astroparticle Physics and Cosmology (Сантьяго де Кампостела, Испания, 2015); Geometric Foundations of Gravity (Тарту, Эстония, 2017); 9th Mathematical Physics Meeting: Summer School and Conference on Modern Mathematical Physics (Белград, Сербия, 2017);
на научных семинарах кафедры физики высоких энергий и элементарных частиц физического факультета СПбГУ; а также на научных семинарах соответствующих научных групп в Университете Гранады (Испания), Университете Хельсинки (Финляндия), Институте теоретической физики Nordita (Стокгольм, Швеция), Национальном Автономном Университете Мехико (UNAM, Мексика).

Публикации

Основные результаты диссертации опубликованы в 18 печатных работах в изданиях, индексируемых базами данных "Web of Science" и "SCOPUS". Список работ приведён в Заключении.

Объём и структура работы

Диссертация состоит из Введения, 5 глав и Заключения. Полный объём Диссертации составляет 298 страниц. Диссертация содержит список литературы из 247 наименований, не считая собственных работ.

В начале каждой главы приведено её краткое содержание и указаны работы, в которых опубликованы вошедшие в неё результаты. Основные результаты, полученные в Диссертации, сформулированы в Заключении.

Глава 1

Теория гравитации и теоретическая космология: вводный обзор

В первой главе мы даём краткое введение в теорию гравитации и теоретическую космологию. Разумеется, столь краткое описание ни в коем случае не может претендовать на полноту и глубину. В гравитационной части наша основная задача заключается во введении обозначений и погружении общей теории относительности Эйнштейна в более широкий контекст геометрических теорий с метрической и аффинной структурами, включая телепараллельный эквивалент. В космологической части нашей основной задачей будет представить современную космологию как физическую дисциплину со своими наблюдательными данными, а также мотивировать на этой основе изучение модифицированных теорий гравитации.

Данная глава не содержит выносимых на защиту результатов.

1.1 Геометрические основы теории гравитации

Как известно, в современной теоретической физике гравитация описывается в рамках геометрической теории искривлённого пространства-времени. На идею использования геометрических конструкций наводит факт пропорциональности инертной и гравитационной масс. Пренебрегая сопротивлением воздуха, можно считать, что все тела падают на землю с одинаковым ускорением. Соответственно, естественно предположить, что дело тут не в существовании дополнительной силы, а в поведении геодезических линий, по которым движутся свободные тела.

С математической точки зрения пространство-время моделируется с помощью многообразия, то есть хаусдорфова топологического пространства, обладающего счётной базой топологии и локально гомеорморфного (псевдо)эвклидову пространству. Аксиомы отделимости и счётности используются для исключения заведомо патологических вариантов, и мы их обсуждать не будем, поскольку соответствующие примеры обычно всё-равно не приходят на ум физику-теоретику (см. математические источники по топологии многообразий, например [1]). Локальные же гомеоморфизмы (или карты), по сути, вводят координатные системы в заданных областях на многообразии, а само многообразие, с топологической точки зрения, локально не отличимо от эвклидова пространства.

Поскольку законы физики записываются обычно в виде дифференциальных уравнений, приходится использовать гладкие многообразия. Иными словами, в атлас допускаются только такие карты, которые в случае пересечения носителей на многообразии позволяют пересчитать одни координаты в другие с помощью гладких функций. Соответственно, гладкой функцией на многообразии называется функция, которая может быть записана как гладкая функция координат. Это и есть то свойство, которое важно для нас в практической работе.

Для введения понятия вектора на гладком многообразии естественно положиться на интуицию, связанную с понятием скорости. Вектор можно рассматривать как направление и "быстроту" бесконечно малого сдвига вдоль многобразия. Соответственно, касательный вектор в точке x_0 на гладком многообразии $\mathcal{M}$ можно определить как дифференцирование $\sum\limits_i a^i \frac{\partial}{\partial x^i}$ на ростке гладких функций в точке $x_0 \in \mathcal{M}$. Числа a^i являются привычными компонентами вектора, и их привычный закон преобразования при заменах координат следует из правила замены переменных в дифференциальном операторе.

Множество всех касательных векторов очевидным образом может быть наделено структурой линейного пространства размерности, равной размерности многообразия. Это касательное пространство к многообразию в заданной точке. Линейные функционалы на этом пространстве образуют двойственное пространство, известное как кокасательное пространство. Его элементы называются дифференциальными 1-формами. А дуальный базис может быть определен как $dx^i \cdot \frac{\partial}{\partial x^j} \equiv \delta^i_j$. Как касательные, так и кокасательные пространства, можно рассмотреть одновременно в каждой точке многообразия, и получаемый при этом объект может быть естественным образом наделён структурой гладкого многообразия удвоенной размерности – это касательное, или соответственно кокасательное, расслоение многообразия $\mathcal{M}$. На этом языке векторное поле на многообразии рассматривается как сечение касательного расслоения (гладкий выбор одного из представителей в касательном слое над каждой точкой многообразия).

Аналогично можно построить линейные пространства тензоров более высокой валентности в любой заданной точке, а также и соответствующие расслоения. В частности, можно рассмотреть билинейные формы на касательном пространстве. Очевидно, что невырожденная симметричная билинейная форма $\sum\limits_{i,j} g_{ij} \cdot dx^i \otimes dx^j$ на касательном пространстве может рассматриваться как определение скалярного произведения касательных векторов, и тогда она называется метрикой.

Название "метрика" оправдано тем, что она даёт возможность интуитивно определить длину вдоль данной кривой как $\int \sqrt{\sum\limits_{i,j} g_{ij} dx^i dx^j}$, а формальная подстановка $dx^i = \frac{dx^i}{d\tau} d\tau$, оправдываемая определением кривой как гладкого отображения (отрезка) вещественной прямой в рассматриваемое многообразие, превращает интуитивное выражение в корректно определённый интеграл по параметру кривой τ.

Разумеется, можно рассматривать метрику также и как отображение из касательного пространства в кокасательное, поскольку билинейная форма всегда может быть рассмотрена как линейный функционал, зависящий от одного из аргументов и действующий на другой. Для математика это не что иное, как утверждение о коммутативности диаграммы, в которой можно либо сразу применить симметричную билинейную форму к двум векторам на касательном пространстве, либо сперва преобразовать один из векторов в 1-форму (кокасательный вектор) действием определяемого отображения, а потом применить эту 1-форму ко второму вектору. Для физика же это просто привычное правило опускания и поднимания индексов с помощью метрики (ковариантные и контравариантные компоненты векторов).

Поскольку установлено взаимно-однозначное соответствие между касательными и кокасательными векторами, можно задаться вопросом о нахождении скалярного произведения кокасательных векторов, эквивалентного тому, что задаётся метрикой на касательном пространстве. Легко проверить, что соответствующая билинейная форма определяется матрицей, обратной к g_{ij}, её обычно обозначают g^{ij}.

Рассмотренных выше понятий оказывается достаточно для анализа гладких функций (скалярных полей) на многообразии. Однако многие физические задачи связаны с написанием дифференциальных уравнений для векторных (или других тензорных) полей. И в этом случае необходимо уметь сравнивать векторы, приложенные в двух разных (хотя бы и бесконечно близких) точках пространства. Это требует введения новой структуры – связности. По сути, задав вектор в некоторой точке, мы хотим указать во всех остальных точках многообразия другой вектор (элемент соответствующего касательного слоя), который будет называться параллельным данному. Таким образом, задание параллельного переноса данного вектора вдоль многообразия является выбором некоторого сечения касательного расслоения.

Более того, если потребовать сохранения линейных соотношений между векторами в процессе параллельного переноса, то достаточно указать одно такое сечение. Иными словами, мы рассматриваем расслоение со структурной группой, которая действует на слоях транзитивно, и поэтому коммутативность некоторой диаграммы (сперва перенос, потом преобразование, или наоборот) позволяет определить по одному горизонтальному сечению все остальные.

Общая теория относительности имеет дело с многообразиями с метрикой и (вполне определённой, см. ниже) связностью. Отличие от, возможно, более часто встречающихся в математике задач заключается в том, что метрика выбирается не знакоопределённой, но с лоренцовой сигнатурой квадратичной формы: $(+,-,-,-)$ или $(-,+,+,+)$, в разных работах по-разному. Мы в дальнейшем номера координат на многообразиях с лоренцовой сигнатурой будем обозначать греческими буквами, и как обычно, условимся подразумевать суммирование по повторяющимся буквам, если одна расположена сверху, а другая – снизу.

Разумеется, при попытках построения модифицированных (расширенных) теорий гравитации также вполне естественно широко использовать геометрические идеи. Ниже мы кратко, в том числе с целью фиксации обозначений, приводим основные понятия, необходимые для дальнейшего, но уже в покомпонентной записи без обращения к фундаментальным математическим основам. (С математической точкой зрения на дифференциальную геометрию можно познакомиться, например, по прекрасной монографии [2].)

1.1.1 Аффинная связность

Итак, как мы уже обсудили выше, для построения векторного анализа на многообразиях необходимо умение сравнивать вектора, находящиеся в различных точках.

Разумеется, в эвклидовом пространстве существует естественное понятие параллельного переноса. В координатном представлении оно выглядит просто тольно в случае использования декартовых координат:

компоненты параллельно перенесённого вектора численно совпадают с компонентами исходного. В криволинейных системах координат это уже не так. Однако же легко убедиться, что для гладкой системы координат, и в первом пордяке малости по бесконечно малому параллельному переносу δx^μ, имеем

$$\delta A_\nu = \Gamma^\alpha_{\mu\nu} A_\alpha \delta x^\mu,$$

$$\delta A^\nu = -\Gamma^\nu_{\mu\alpha} A^\alpha \delta x^\mu,$$

где коэффициенты Γ называются коэффиентами связности или символами Кристоффеля.

Тот факт, что формулы для ковариантынх и контравариантных компонент очень похожи, но отличаются знаком, связан конечно же с сохранением скалярного произведения $A_\mu B^\mu$ при параллельном переносе. Данное свойство, разумеется, хочется сохранить и в искривлённом пространстве для касательных векторов (дифференцирований гладких функций в заданной точке) и дифференциальных форм (кокасательных векторов).

Если компоненты двух векторов в бесконечно близких точках многообразия отличаются ровно так, как указано выше (в соответствии с заданными значениями коэффициентов связности), то мы интерпретируем это так, что один вектор получен из другого параллельным переносом, или, допуская известную вольность речи, что два вектора в разных точках многообразия равны. Отличие же от подобного закона преобразования будем связывать с непостоянством векторного поля, представленного в этих точках данными векторами. В соответствии с этим, определим ковариантные производные как

$$\bigtriangledown_\mu A_\nu \equiv \partial_\mu A_\nu - \Gamma^\alpha_{\mu\nu} A_\alpha, \tag{1.1}$$

$$\bigtriangledown_\mu A^\nu \equiv \partial_\mu A^\nu + \Gamma^\nu_{\mu\alpha} A^\alpha. \tag{1.2}$$

Легко найти трансформационные свойства коэффициентов связности из требования, чтобы ковариантные производные были тензорными объектами (уже по двум значкам):

$$\Gamma^{\kappa}_{\mu\nu} = \frac{\partial x^{\kappa}}{\partial x'^{\rho}} \left(\frac{\partial x'^{\alpha}}{\partial x^{\mu}} \frac{\partial x'^{\beta}}{\partial x^{\nu}} \Gamma'^{\rho}_{\alpha\beta} + \frac{\partial^2 x'^{\rho}}{\partial x^{\mu} \partial x^{\nu}} \right). \tag{1.3}$$

Ясно, что симметричные связности, $\Gamma^{\kappa}_{\mu\nu} = \Gamma^{\kappa}_{\nu\mu}$, являются выделенными с точки зрения принципа эквивалентности, поскольку их можно обратить в нуль преобразованием координат (локально инерциальная система отсчёта). С другой стороны, антисимметричная часть является тензором, известным под именем кручения.

Обратим внимание, что закон преобразования связности (1.3) содержит неоднородный член, и с точки зрения абстрактной алгебры он является не линейным, но только аффинным. Отсюда и наименование связности.

Геодезические кривые в пространствах аффинной связности

Понятие аффинной связности позволяет ввести понятие геодезической линии, которое хорошо согласуется с интуитивным представлением о геодезических. А именно, назовём линию $x^{\mu}(\tau)$ геодезической, если и только если вектор касательной $e^{\mu} \equiv \frac{dx^{\mu}}{d\tau}$ ковариантно постоянен вдоль кривой. Иными словами, взяв вектор e^{μ} в одной точке кривой и параллельно перенеся его в другую точку на кривой, мы получаем снова вектор e^{μ} в новой точке. Легко видеть, что такое определение приводит к следующему уравнению геодезической:

$$\frac{d^2 x^{\mu}}{d\tau^2} + \Gamma^{\mu}_{\nu\alpha} \frac{dx^{\alpha}}{d\tau} \frac{dx^{\nu}}{d\tau} = 0. \tag{1.4}$$

Данное уравнение инвариантно по отношению к аффинным преобразованиям параметра τ. При более общих заменах параметра уравнение усложняется, и это определяет предпочтительный класс параметризаций – параметризации аффинным параметром. В свою очередь, последнее позволяет естественным образом определить понятие светоподобной бесконечности – в терминах аффинного параметра.

1.1.2 Кривизна и кручение

Имея некоторое искривлённое пространство, очень полезно знать, действительно ли оно искривлённое, или мы просто ввели криволинейные координаты в эвклидовом пространстве. В римановой геометрии характеристикой, которая позволяет отличить действительно искривлённые пространства, является тензор кривизны Римана (для более общих метрически-аффинных геометрий его надо дополнить тензором кручения и тензором неметричности, см. ниже). Идея его определения очень проста. Возьмём вектор, заданный в некоторой точке, перенесём его по замкнутому контуру и сравним с тем, с чего мы начинали. Если произойдёт ненулевое изменение компонент, то многообразие заведомо искривлённое.

Совершим параллельный перенос вектора ξ^μ вдоль бесконечно малого замкнутого контура $\mathcal{C}$. После одного оборота получаем

$$\delta\xi^\mu = \oint\limits_{\mathcal{C}} d\tau \dot{\xi}^\mu,$$

где можно считать, что $\xi^\mu(x)$ - векторное поле, определённое в некоторой открытой окрестности параллельным переносом исходного вектора. В частности, имеем вдоль контура

$$\dot{\xi}^\mu = -\Gamma^\mu_{\nu\alpha}\xi^\alpha \frac{dx^\nu}{d\tau}.$$

Для простоты записи предположим, что начало координат ($x^\mu = 0$) расположено внутри контура, а далее используем уравнение параллельного переноса, раскладывая все величины в ряд Тейлора с точностью до первого порядка малости по значениям координат: $\delta\xi^\mu =$

$$-\oint\limits_{\mathcal{C}} d\tau \left(\Gamma^\mu_{\nu\alpha}(0) + \Gamma^\mu_{\nu\alpha,\rho}(0)x^\rho + \mathcal{O}(x^2)\right)\left(\xi^\alpha(0) + \xi^\alpha_{,\kappa}(0)x^\kappa + \mathcal{O}(x^2)\right)\frac{dx^\nu}{d\tau},$$

где, как обычно, запятой обозначена простая частная производная по координате.

В низшем порядке интеграл $\sim \oint dx^\mu = 0$, а в первой поправке имеем:

$$\delta\xi^\mu = -\oint\limits_{\mathcal{C}} d\tau \left[\Gamma^\mu_{\nu\alpha}\xi^\alpha_{,\rho}x^\rho + \Gamma^\mu_{\nu\alpha,\rho}\xi^\alpha x^\rho + \mathcal{O}(x^2)\right]\frac{dx^\nu}{d\tau}.$$

В первом слагаемом снова воспользуемся уравнением параллельного переноса $\xi^\alpha_{,\rho} = -\Gamma^\alpha_{\rho\sigma}\xi^\sigma$:

$$\delta\xi^\mu = -\oint\limits_{\mathcal{C}} d\tau \left[-\Gamma^\mu_{\nu\alpha}\Gamma^\alpha_{\rho\sigma}\xi^\sigma x^\rho + \Gamma^\mu_{\nu\sigma,\rho}\xi^\sigma x^\rho + \mathcal{O}(x^2)\right]\frac{dx^\nu}{d\tau}.$$

Как видим, полученное выражение пропорционально

$$\oint x^\rho \frac{dx^\nu}{d\tau}d\tau = -\oint x^\nu \frac{dx^\rho}{d\tau}d\tau = \frac{1}{2}\oint (x^\rho dx^\nu - x^\nu dx^\rho) = \int\int dx^\rho \wedge dx^\nu,$$

антисимметричному элементу площади (заключённому контуром $\mathcal{C}$), который мы обозначим $S^{\rho\nu}$. Антисимметризованный коэффициент при S тоже обязан быть тензором (что можно проверить и напрямую, используя закон преобразования связности (1.3)).

Окончательно мы получаем

$$\delta\xi^\mu = \oint\limits_{\mathcal{C}} d\tau \dot{\xi}^\mu = -\frac{1}{2}R^\mu{}_{\sigma\rho\nu}\xi^\sigma S^{\rho\nu} \tag{1.5}$$

где тензор Римана (тензор кривизны) определён как

$$R^\mu{}_{\sigma\rho\nu} = \Gamma^\mu_{\nu\sigma,\rho} - \Gamma^\mu_{\rho\sigma,\nu} + \Gamma^\mu_{\rho\alpha}\Gamma^\alpha_{\nu\sigma} - \Gamma^\mu_{\nu\alpha}\Gamma^\alpha_{\rho\sigma}. \tag{1.6}$$

и по самому своему определению антисимметричен по последней паре индексов

$$R^\mu{}_{\sigma\rho\nu} = -R^\mu{}_{\sigma\nu\rho}. \tag{1.7}$$

Это общее свойство, справедливое при любой аффинной связности.

Разумеется, можно было рассмотреть и перенос дифференциальной формы по замкнутому контуру. В результате

$$\delta\zeta_\mu = \oint\limits_{\mathcal{C}} d\tau \dot{\zeta}_\mu = \frac{1}{2} R^{\sigma}{}_{\mu\rho\nu} \zeta_\sigma S^{\rho\nu} \tag{1.8}$$

получается тот же самый тензор, что и в соотношении (1.5). Изменение знака снова соответствует сохранению скаляров.

* * *

В случае симметричных связностей можно убедиться, что

$$R^{\mu}{}_{\alpha\beta\gamma} + R^{\mu}{}_{\beta\gamma\alpha} + R^{\mu}{}_{\gamma\alpha\beta} = 0, \tag{1.9}$$

а также (тождества Бьянки)

$$\nabla_\alpha R^{\mu}{}_{\nu\beta\gamma} + \nabla_\beta R^{\mu}{}_{\nu\gamma\alpha} + \nabla_\gamma R^{\mu}{}_{\nu\alpha\beta} = 0. \tag{1.10}$$

Проще всего это сделать в локально геодезической системе координат.

Коммутатор ковариантных производных

Другой способ определения тензора кривизны – это рассмотрение коммутатора ковариантных производных:

$$\begin{aligned}
[\nabla_\mu\,,\ \nabla_\nu]\,\xi^\alpha &= \nabla_\mu \left(\partial_\nu \xi^\alpha + \Gamma^\alpha_{\nu\beta}\xi^\beta\right) - \nabla_\nu \left(\partial_\mu \xi^\alpha + \Gamma^\alpha_{\mu\beta}\xi^\beta\right) = \\
&= \partial_\mu \left(\partial_\nu \xi^\alpha + \Gamma^\alpha_{\nu\beta}\xi^\beta\right) + \Gamma^\alpha_{\mu\rho}\left(\partial_\nu \xi^\rho + \Gamma^\rho_{\nu\beta}\xi^\beta\right) - \Gamma^\rho_{\mu\nu}\left(\partial_\rho \xi^\alpha + \Gamma^\alpha_{\rho\beta}\xi^\beta\right) - \\
&- \partial_\nu \left(\partial_\mu \xi^\alpha + \Gamma^\alpha_{\mu\beta}\xi^\beta\right) - \Gamma^\alpha_{\nu\rho}\left(\partial_\mu \xi^\rho + \Gamma^\rho_{\mu\beta}\xi^\beta\right) + \Gamma^\rho_{\nu\mu}\left(\partial_\rho \xi^\alpha + \Gamma^\alpha_{\rho\beta}\xi^\beta\right) = \\
&= R^{\alpha}{}_{\beta\mu\nu}\xi^\beta - T^{\rho}{}_{\mu\nu}\,\nabla_\rho\,\xi^\alpha,
\end{aligned} \tag{1.11}$$

где мы ввели тензор кручения

$$T^{\rho}{}_{\mu\nu} = \Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\nu\mu}. \tag{1.12}$$

Аналогично для дифференциальных форм получаем

$$[\nabla_\mu \, , \, \nabla_\nu] \, \zeta_\alpha = -R^\beta{}_{\alpha\mu\nu} \zeta_\beta - T^\rho{}_{\mu\nu} \, \nabla_\rho \, \zeta_\alpha. \tag{1.13}$$

Между двумя способами введения кривизны, разумеется, есть глубокая связь. В самом деле, введем понятие производной по направлению вектора $\overrightarrow{u}$ с помощью соотношения $\nabla_{\overrightarrow{u}} \equiv u^\mu \nabla_\mu$. В результате имеем

$$[\nabla_{\overrightarrow{u}} \, , \, \nabla_{\overrightarrow{v}}] \, \xi^\alpha = R^\alpha{}_{\beta\mu\nu} \xi^\beta u^\mu v^\nu + (v^\mu \partial_\mu u^\nu - u^\mu \partial_\mu v^\nu) \, \nabla_\nu \, \xi^\alpha.$$

В правой части получена производная Ли одного вектора по направлению другого $\mathcal{L}_{\overrightarrow{v}} \overrightarrow{u} \equiv \overrightarrow{[v,u]}$, где коммутатор двух векторных полей равен $[v,u]^\nu \equiv v^\mu \partial_\mu u^\nu - u^\mu \partial_\mu v^\nu$. Соответственно, вместо уравнения (1.11) легко получить

$$\left([\nabla_{\overrightarrow{u}} \, , \, \nabla_{\overrightarrow{v}}] - \nabla_{\overrightarrow{[v,u]}} \right) \overrightarrow{\xi} = \hat{R}(\overrightarrow{u}, \overrightarrow{v}) \, \overrightarrow{\xi},$$

где линейный оператор $\hat{R}(\overrightarrow{u}, \overrightarrow{v})$ определён своими компонентами $\hat{R}(\overrightarrow{u}, \overrightarrow{v})^\alpha{}_\beta = R^\alpha{}_{\beta\mu\nu} u^\mu v^\nu$.

Теперь в левой части формулы можно увидеть перенос вектора вдоль почти параллелограмма, составленного из векторов $\overrightarrow{u}$ и $\overrightarrow{v}$ и их параллельных переносов. Незамыкание "параллелограмма" за счет кривизны пространства компенсируется членом с $\nabla_{\overrightarrow{[v,u]}}$. Детальное рассмотрение этого вопроса, а также ещё одну интерпретацию тензора кривизны (с точки зрения расхождения геодезических), можно найти в прекрасном учебнике Мизнера, Торна и Уилера [3].

Тензор Риччи

Путём свёртки двух индексов из тензора Римана (1.6) получается тензор Риччи

$$R_{\mu\nu} \equiv R^\rho{}_{\mu\rho\nu} = \Gamma^\rho_{\mu\nu,\rho} - \Gamma^\rho_{\rho\mu,\nu} + \Gamma^\rho_{\rho\alpha} \Gamma^\alpha_{\nu\mu} - \Gamma^\rho_{\nu\alpha} \Gamma^\alpha_{\rho\mu}. \tag{1.14}$$

Возможны ещё две свёртки, $R^\rho{}_{\mu\nu\rho} = -R_{\mu\nu}$ и $R^\rho{}_{\rho\mu\nu}$. Первая не даёт ничего нового в силу всегда справедливого свойства антисимметрии (1.7), а вторая может использоваться только в моделях с неметрической связностью, ибо в противном случае обращается в нуль, см. ниже ещё одно свойство антисимметрии. Использование такой величины можно встретить в работах по модифицированным теориям гравитации, но нам она в дальнейшем в явном виде не потребуется.

1.1.3 Метрическая связность

На многообразиях также могут заданы и другие структуры помимо аффинной связности. Для нас наиболее важным примером будет метрическая структура. Обычно предполагается, что аффинная и метрическая структуры согласованы друг с другом в следующем смысле. Будем называть связность метрической если и только если

$$\bigtriangledown_\mu g_{\alpha\beta} = 0,$$

то есть метрический тензор является ковариантно постоянным.

Как описано в любом учебнике по общей теории относительности, из этого требования

$$\partial_\mu g_{\alpha\beta} = \Gamma^\rho_{\mu\alpha} g_{\rho\beta} + \Gamma^\rho_{\mu\beta} g_{\alpha\rho},$$

можно однозначно найти связность при условии, что она симметрична:

$$\Gamma^\rho_{\alpha\beta} = \frac{1}{2} g^{\rho\mu} \left(\partial_\alpha g_{\mu\beta} + \partial_\beta g_{\alpha\mu} - \partial_\mu g_{\alpha\beta} \right). \tag{1.15}$$

Полученная связность называется связностью Леви-Чивита.

Впрочем, из стандартного вывода этой формулы очевидно, что произвольная связность может быть представлена в виде связности Леви-Чивита с дополнительными слагаемыми, представляющими вклады кручения $T^\rho{}_{\mu\nu} \equiv \Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\nu\mu}$ и неметричности $Q_{\mu\alpha\beta} \equiv \bigtriangledown_\mu g_{\alpha\beta}$. В разделе 1.1.7 мы увидим это соотношение в частном случае нулевой неметричности.

Если на многообразии задана метрика, то имеется естественное соответствие между векторами и дифференциальными 1-формами, и можно даже просто говорить о ковариантных и контравариантных компонентах векторов и тензоров, опуская и поднимая индексы с помощью метрики, как например $R_{\gamma\beta\mu\nu} = g_{\gamma\alpha} R^{\alpha}{}_{\beta\mu\nu}$ (с помощью $g^{\mu\nu}$ мы обозначаем матрицу, обратную к метрике).

Метричность связности означает, что можно менять характер значков прямо под знаком ковариантной производной, и из этого следует новое свойство симметрии тензора Римана. В самом деле, перепишем уравнение (1.11) в виде

$$[\nabla_\mu \,,\, \nabla_\nu]\, \xi_\alpha = R_{\alpha\beta\mu\nu} \xi^\beta$$

С другой стороны, уравнение (1.13) дает

$$[\nabla_\mu \,,\, \nabla_\nu]\, \zeta_\alpha = -R_{\beta\alpha\mu\nu} \zeta^\beta .$$

И мы заключаем, что

$$R_{\alpha\beta\mu\nu} = -R_{\beta\alpha\mu\nu} . \tag{1.16}$$

* * *

Если связность является одновременно и симметричной и метрической (связность Леви-Чивита), то используя формулу (1.15), можно записать тензор кривизны явно в терминах компонент метрики (для простоты в локально инерциальной системе, где $\partial_\mu g_{\alpha\beta} = 0$ в заданной точке), и увидеть, что

$$R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu} . \tag{1.17}$$

И, в частности, тензор Риччи оказывается симметричным,

$$R_{\mu\nu} = R_{\nu\mu} . \tag{1.18}$$

* * *

Легко проверить, что в случае использования согласованной с метрикой связности данное ранее определение геодезической совпадает с тем, что получается через экстремальность длины кривой.

Также заметим, что имея на многообразии метрику, можно определить скалярную кривизну посредством

$$R = R^{\mu}_{\mu} = g^{\mu\nu} R_{\mu\nu}. \tag{1.19}$$

1.1.4 Уравнения Эйнштейна и действие Эйнштейна-Гильберта

Общая теория относительности может быть получена как простейшая возможная модель гравитации, основанная на (псевдо)римановой геометрии (то есть с использованием связности Леви-Чивита).

Известно, что в нерелятивистском пределе источником гравитационного поля служит масса. В рамках релятивистского обобщения её следует рассматривать как энергию покоя, а следовательно простейшим ковариантным кандидатом на роль источника поля может служить тензор энергии-импульса. Последний должен взаимодействовать с другим симметричным тензорным объектом с двумя значками, описывающим геометрию пространства-времени. В рамках простейших предположений простейшим подходящим объектом, помимо метрики, оказывается тензор Риччи.

Следует, однако, также вспомнить, что тензор энергии-импульса удовлетворяет условию сохранения, которое в ковариантном виде может быть записано как $\bigtriangledown_{\mu} T^{\mu\nu} = 0$. Используя тождества Бьянки (1.10), легко убедиться, что с геометрической стороны аналогичным свойством обладает объект

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}, \tag{1.20}$$

именуемый тензором Эйнштейна.

Поэтому естественно искать уравнения Эйнштейна в виде

$$G_{\mu\nu} \propto T_{\mu\nu},$$

где коэффициент пропорциональности должен быть установлен из соответствия ньютоновскому пределу. Эти детали мы опускаем, отсылая читателя к любому учебнику по общей теории относительности. Кроме того, следует отметить, что к правой части уравнений Эйнштейна можно добавить слагаемое, пропорциональное метрике, не нарушая ковариантного сохранения.

Окончательно получаем

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{8\pi\mathcal{G}}{c^4} \left(T_{\mu\nu} + \Lambda g_{\mu\nu} \right), \tag{1.21}$$

где $\mathcal{G}$ - постоянная Всемирного тяготения, Λ - космологическая постоянная, про чью тяжёлую судьбу мы пока умолчим, а c - скорость света, которую мы почти всюду будем полагать равной единице.

* * *

Существует удивительно простое действие, варированием которого выводятся уравнения Эйнштейна:

$$S = \int d^4x \sqrt{-g} \left(\frac{1}{16\pi\mathcal{G}} R + \Lambda \right) + S_{\mathrm{matter}} \tag{1.22}$$

где g – детерминант матрицы $g_{\mu\nu}$ (выбранный нами знак при гравитационной части действия соответствует выбору сигнатуры метрики $(-,+,+,+)$ при условии положительности кинетической энергии гравитонов). Отметим, что тензор энергии-импульса материи определён как

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta}{\delta g^{\mu\nu}} S_{\mathrm{matter}}. \tag{1.23}$$

Мы не будем останавливаться на хорошо известных способах варьирования детерминантов. Отметим лишь, что по ходу вывода уравнений движения из действия Эйнштейна-Гильберта необходимо убедиться, что вариация тензора Риччи не даёт вклада. Её можно записать (в точном виде) как

$$\delta R_{\mu\nu} = \delta\Gamma^{\rho}_{\mu\nu;\rho} - \delta\Gamma^{\rho}_{\rho\mu;\nu} + \delta\Gamma^{\rho}_{\rho\alpha}\delta\Gamma^{\alpha}_{\nu\mu} - \delta\Gamma^{\rho}_{\nu\alpha}\delta\Gamma^{\alpha}_{\rho\mu} = \delta\Gamma^{\rho}_{\mu\nu;\rho} - \delta\Gamma^{\rho}_{\rho\mu;\nu} + \mathcal{O}\left(\delta\Gamma^2\right), \tag{1.24}$$

где точками с запятой обозначены ковариантные производные (напомним, что вариация коэффициента связности, как и любая разность двух коэффициентов связности, является тензором), например $R_{\mu\nu;\alpha} \equiv \nabla_\alpha R_{\mu\nu}$. Ковариантные производные были получены комбинированием простых частных производных $\partial\Gamma$ с членами вида $\Gamma\delta\Gamma$. Таким образом, данная вариация сводится к чисто поверхностному слагаемому.

Отметим, что это общий факт. Если дана вариация коэффициентов связности, то вариацию тензора Римана легко вычислить точно, причём она квадратична по вариации связности. А если исходная связность была леви-чивитовской, то линейная часть вариации является полной ковариантной (по отношению к метрике g) производной. Это бывает очень полезно иметь в виду при изучении модифицированных теорий гравитации, как мы ещё неоднократно увидим.

1.1.5 Основы АДМ формализма

Для получения гамильтоновой формулировки уравнений Эйнштейна, изучения свойств начально-краевых задач, построения численных методов, а также и при выяснении свойств модифицированных теорий, очень полезной является формулировка в АДМ переменных [4].

В АДМ подходе используется следующее (3+1)-разложение метрики:

$$ds^2 = -\left(N^2 - N_i N^i\right) dt^2 + 2N_i dx^i dt + \gamma_{ij} dx^i dx^j \tag{1.25}$$

где N и N_i известны как шаг (lapse) и сдвиг (shift) соответственно. Пространственные индексы следует поднимать и опускать с помощью метрики γ_{ij}. Иными словами, мы положили

$$g_{00} = -\left(N^2 - N_i N^i\right), \quad g_{0i} = N_i, \quad g_{ij} = \gamma_{ij}.$$

Весьма несложно найти и обратную матрицу:

$$g^{00} = -\frac{1}{N^2}, \quad g^{0i} = \frac{N^i}{N^2}, \quad g^{ij} = \gamma^{ij} - \frac{N^i N^j}{N^2}.$$

Сравнив этот результат с правилом Крамера для вычисления g^{00}, получаем полезное соотношение для детерминантов

$$\sqrt{-g} = N\sqrt{\gamma}.$$

Введём единичный вектор нормали

$$n^{\mu} \equiv \left(\frac{1}{N}, -\frac{N^i}{N}\right)$$

к поверхностям постоянного времени (легко проверить, что он соответствует 1-форме $n_{\mu} \equiv \left(-N, \overrightarrow{0}\right)$, принимающей нулевое значение на касательных к гиперповерхности векторах). Внешние кривизны гиперповерхности $t = const$ определяются как

$$K_{ij} \equiv - \bigtriangledown_i n_j = \Gamma^{\mu}_{ij} n_{\mu} = -N\Gamma^0_{ij}, \tag{1.26}$$

где ковариантные производные $\bigtriangledown_{\mu} n_{\nu} \equiv \partial_{\mu} n_{\nu} - \Gamma^{\alpha}_{\mu\nu} n_{\alpha}$ и символы Кристоффеля

$$\Gamma^{\alpha}_{\mu\nu} = g^{\alpha\beta}\Gamma_{\beta\mu\nu} = \frac{1}{2} g^{\alpha\beta}\left(\partial_{\mu} g_{\beta\nu} + \partial_{\nu} g_{\beta\mu} - \partial_{\beta} g_{\mu\nu}\right)$$

относятся к метрике g.

Очевидно,что $\Gamma_{ijk} = \overset{(3)}{\Gamma}_{ijk}$, где $\overset{(3)}{\Gamma}$ – символы Кристоффеля для метрики γ. Кроме того, получаем $\Gamma_{0ij} = \frac{1}{2}\left(\partial_i N_j + \partial_j N_i - \dot{\gamma}_{ij}\right)$, и соответственно

$$K_{ij} = -N\Gamma^0_{ij} = \frac{1}{2N}\left(\overset{(3)}{\bigtriangledown}_i N_j + \overset{(3)}{\bigtriangledown}_j N_i - \dot{\gamma}_{ij}\right) \tag{1.27}$$

для внешних кривизн (1.26) нашей гиперповерхности. Отметим, что трёхмерные ковариантные производные соответствуют метрике γ. Кстати, отметим также, что

$$g^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & \gamma^{-1} \end{pmatrix} - n \times n^T. \tag{1.28}$$

Для применения АДМ анализа к некоторой гравитационной теории необходимо выразить компоненты тензора кривизны (1.6) в АДМ переменных (1.25). В случае леви-чивитовской связности это задача упрощается в связи с соотношениями симметрии

$$R_{\alpha\beta\mu\nu} = -R_{\beta\alpha\mu\nu} = -R_{\alpha\beta\nu\mu} = R_{\mu\nu\alpha\beta}.$$

По сути, требуется найти три величины: R_{ijkl}, R_{0ijk} и R_{0i0j}. Тензор кривизны другой связности на многообразии с метрикой часто бывает удобно выразить через метрический с поправкой, квадратичной по разности связностей.

Коэффициенты связности

Вычисление кривизны для данной метрики методом грубой силы оказывается весьма громоздким, поэтому удобно начать с получения коэффициентов метрической связности. Нам потребуются в дальнейшем следующие формулы, которые легко можно проверить, всегда заменяя $\dot{\gamma}$ на внешние кривизны по формуле (1.27):

$$\begin{aligned}
\Gamma_{ij0} = \Gamma_{i0j} &= -NK_{ij} + \overset{(3)}{\nabla}_j N_i, && (1.29)\\
\Gamma_{ijk} &= \overset{(3)}{\Gamma}_{ijk}, && (1.30)\\
\Gamma^0_{00} &= \frac{1}{N}\left(\dot{N} + N^i\partial_i N - N^i N^j K_{ij}\right), && (1.31)\\
\Gamma^0_{0i} = \Gamma^0_{i0} &= \frac{1}{N}\left(\partial_i N - N^j K_{ij}\right), && (1.32)\\
\Gamma^i_{0j} = \Gamma^i_{j0} &= -\frac{N^i\partial_j N}{N} - N\left(\gamma^{ik} - \frac{N^i N^k}{N^2}\right)K_{kj} + \overset{(3)}{\nabla}_j N^i, && (1.33)\\
\Gamma^0_{ij} &= -\frac{1}{N}K_{ij}, && (1.34)\\
\Gamma^i_{jk} &= \overset{(3)}{\Gamma}{}^i_{jk} + \frac{N^i}{N}K_{jk}. && (1.35)
\end{aligned}$$

Компоненты тензора Римана

Используя формулы (1.29) – (1.35), легко получить компоненты тензора Римана (1.6). Во-первых:

$$
\begin{aligned}
R_{ijkl} = g_{i\rho}\partial_k\Gamma^\rho_{lj} - g_{i\rho}\partial_l\Gamma^\rho_{kj} + \Gamma_{ik\rho}\Gamma^\rho_{lj} - \Gamma_{il\rho}\Gamma^\rho_{kj} \\
= -N_i\partial_k\left(\frac{1}{N}K_{jl}\right) + \gamma_{im}\partial_k\left(\overset{(3)}{\Gamma}{}^m_{jl} + \frac{N^m}{N}K_{jl}\right) \\
-\frac{1}{N}K_{jl}\left(-NK_{ik} + \overset{(3)}{\nabla}_k N_i\right) + \overset{(3)}{\Gamma}{}_{ikm}\left(\overset{(3)}{\Gamma}{}^m_{lj} + \frac{N^m}{N}K_{lj}\right) - (k\leftrightarrow l) \\
= \overset{(3)}{R}{}_{ijkl} + K_{ik}K_{jl} - K_{il}K_{jk}.
\end{aligned}
\tag{1.36}
$$

Далее проще вычислить величины

$$
n_\mu R^\mu{}_{i\alpha j} = n^\mu R_{\mu i\alpha j} = \frac{1}{N}R_{0i\alpha j} - \frac{N^k}{N}R_{ki\alpha j} \tag{1.37}
$$

вместо $R_{0i\alpha j}$. Получаем

$$
\begin{aligned}
n_\mu R^\mu{}_{ijk} = -N\left(\partial_j\Gamma^0_{ki} + \Gamma^0_{j\rho}\Gamma^\rho_{ki}\right) + N\left(\partial_k\Gamma^0_{ji} + \Gamma^0_{k\rho}\Gamma^\rho_{ji}\right) \\
= \partial_j K_{ki} + \overset{(3)}{\Gamma}{}^m_{ki}K_{jm} - (j\leftrightarrow k)\,,
\end{aligned}
$$

или окончательно

$$
n_\mu R^\mu{}_{ijk} = \overset{(3)}{\nabla}_j K_{ki} - \overset{(3)}{\nabla}_k K_{ji}. \tag{1.38}
$$

Наконец, после простых алгебраических преобразований, имеем

$$
n_\mu R^\mu{}_{i0j} = \dot{K}_{ij} + \overset{(3)}{\nabla}_i\overset{(3)}{\nabla}_j N + NK_i{}^kK_{kj} - \overset{(3)}{\nabla}_j\left(K_{ik}N^k\right) - K_{kj}\overset{(3)}{\nabla}_i N^k.
$$

Используя (1.38) и (1.37), последнее соотношение легко приводится к более симметричному виду

$$
n^\mu n^\nu R_{\mu i\nu j} = \frac{1}{N}\left(\dot{K}_{ij} + \overset{(3)}{\nabla}_i\overset{(3)}{\nabla}_j N + NK_i{}^kK_{kj} - \mathcal{L}ie_{\vec{N}}K_{ij}\right) \tag{1.39}
$$

где производная Ли определена как

$$\mathcal{L}ie_{\vec{N}} K_{ij} \equiv N^k \partial_k K_{ij} + K_{ik} \partial_j N^k + K_{jk} \partial_i N^k,$$

а частные производные можно заменить ковариантными, не меняя значения выражения.

Скалярная кривизна и действие Эйнштейна-Гильберта

Полученные формулы позволяют анализировать любую теорию со связностью Леви-Чивита. В случае же общей теории относительности воспользуемся формулой (1.28), свойствами симметрии тензора Римана (1.7, 1.9, 1.16, 1.17), а также очевидным соотношением

$$\gamma^{ij} \dot{K}_{ij} = \partial_0 K^i_i + K^{ij} \dot{\gamma}_{ij}$$

вместе с определением (1.27), чтобы получить

$$\begin{aligned} R = g^{\mu\nu} g^{\alpha\beta} R_{\mu\alpha\nu\beta} &= \gamma^{ik} \gamma^{jl} R_{ijkl} - 2 n^\mu n^\nu \gamma^{ij} R_{\mu i \nu j} \\ &= \overset{(3)}{R} + K^i_i K^j_j - 3 K^{ij} K_{ij} - \frac{2}{N} \gamma^{ij} \dot{K}_{ij} \\ &+ \frac{4}{N} K^{ij} \overset{(3)}{\nabla}_j N_i + 2 \frac{N^j}{N} \overset{(3)}{\nabla}_j K^i_i - \frac{2}{N} \overset{(3)}{\triangle} N \\ &= \overset{(3)}{R} + K^{ij} K_{ij} + K^i_i K^j_j - \frac{2}{N} \dot{K}^i_i + 2 \frac{N^j}{N} \overset{(3)}{\nabla}_j K^i_i - \frac{2}{N} \overset{(3)}{\triangle} N \end{aligned}$$

Наконец, используя $\partial_0 \sqrt{\gamma} = \frac{\sqrt{\gamma}}{2} \gamma^{ij} \dot{\gamma}_{ij}$ и исключая $\dot{\gamma}_{ij}$ с помощью формулы (1.27), получаем лагранжеву плотность для действия Эйнштейна-Гильберта:

$$\sqrt{-g} R = \sqrt{\gamma} N \left(\overset{(3)}{R} + K^{ij} K_{ij} - K^i_i K^j_j \right) - 2 \sqrt{-g} \, \nabla_\mu \left(K^i_i n^\mu \right) - 2 \sqrt{\gamma} \overset{(3)}{\triangle} N$$

где $-2\sqrt{-g} \, \nabla_\mu \left(K^i_i n^\mu \right) \equiv -2 \partial_0 \left(\sqrt{\gamma} K^i_i \right) + 2 \sqrt{\gamma} \overset{(3)}{\nabla}_j \left(K^i_i N^j \right)$.

Пренебрегая поверхностными членами и откидывая полную производную по времени, получаем окончательно действие общей теории относительности в АДМ переменных

$$S = \int dt d^3x \sqrt{\gamma} N \left(\overset{(3)}{R} + K^{ij} K_{ij} - K^i_i K^j_j \right). \tag{1.40}$$

1.1.6 Гравитация в терминах тетрад

В тетрадном представлении (которое необходимо, например, для введения спинорных полей) метрика задаётся с помощью базиса касательных векторов или отвечающих им дифференциальных форм, гладко заданных в каждой точке:

$$g_{\mu\nu} = e^a_\mu e^b_\nu \eta_{ab}. \tag{1.41}$$

Очевидно, что поле тетрад e^a_μ определено с точностью до локальных преобразований Лоренца. Поскольку метрика должна быть невырожденна, предполагается, что матрица e^a_μ невырожденна (имеем базис касательных векторов), и обратная тетрада e^μ_a определяется как обратная матрица: $e^a_\mu e^\mu_b \equiv \delta^a_b$ и $e^a_\mu e^\nu_a \equiv \delta^\nu_\mu$, а обратная метрика при этом равна

$$g^{\mu\nu} = e^\mu_a e^\nu_b \eta^{ab}.$$

Обратную тетраду можно рассматривать как поле систем отсчета (frame field), задаваемых базисом касательных векторов. Это базис, в котором метрика как квадратичная форма принимает свой канонический вид: $g_{\mu\nu} e^\mu_a e^\nu_b = \eta_{ab}$. Разумеется, локально такой базис всегда существует, хотя глобально могут иметься топологические препятствия.

Будем с помощью латинских индексов из начала алфавита обозначать компоненты тензоров, отнесённые к базису тетрад:

$$\mathcal{T}^{a_1,\ldots,a_n}_{b_1,\ldots,b_m} \equiv e^{a_1}_{\alpha_1} \cdots e^{a_n}_{\alpha_n} \mathcal{T}^{\alpha_1,\ldots,\alpha_n}_{\beta_1,\ldots,\beta_m} e^{\beta_1}_{b_1} \cdots e^{\beta_m}_{b_m}.$$

Это можно также интерпретировать как построение вспомогательной копии касательного пространства в каждой точке с метрикой Минковского. При этом тетрада осуществляет изоморфизм двух линейных (псевдо)нормированных пространств: $A^\mu \to A^a = e^a_\mu A^\mu$.

В принципе, можно ввести два типа коэффициентов связности – для тензоров с разными типами индексов. Для латинских индексов обозначим связность через ${\omega^a}_{\mu b}$ и будем называть её спин-связностью. Геометрически это компоненты дифференциальной 1-формы, принимающей значения в алгебре группы Лоренца.

Однако, естественно потребовать, чтобы две связности описывали по сути один и тот же процесс параллельного переноса. Это означает, что операция ковариантного дифференцирования (с Γ для греческих индексов и ω для латинских) должна коммутировать со сменой типа индекса при помощи тетрады. Иными словами, мы требуем обращения в нуль "полной ковариантной производной" тетрады:

$$\partial_\mu e^a_\nu + {\omega^a}_{\mu b} e^b_\nu - \Gamma^\alpha_{\mu\nu} e^a_\alpha = 0. \tag{1.42}$$

Теперь мы можем ковариантно дифференцировать любые тензоры со смешанными наборами индексов, например

$$\nabla_\mu T^{a\alpha} = \partial_\mu T^{a\alpha} + \Gamma^\alpha_{\mu\beta} T^{a\beta} + {\omega^a}_{\mu b} T^{b\alpha},$$

и свободно менять природу этих индексов прямо под знаком ковариантного дифференцирования. Легко проверить, что ценой такого соглашения является невозможность ввести неметричность в теорию:

$$\begin{aligned}\nabla_\alpha\, g_{\mu\nu} = \eta_{ab} \left(\partial_\alpha \left(e^a_\mu e^b_\nu\right) - \Gamma^\beta_{\alpha\mu} e^a_\beta e^b_\nu - \Gamma^\beta_{\alpha\nu} e^a_\mu e^b_\beta\right)& \\ = -e^b_\mu e^c_\nu \left(\eta_{ab} {\omega^a}_{\alpha c} + \eta_{ac} {\omega^a}_{\alpha b}\right) = 0.&\end{aligned}$$

В последнем равенстве мы учли, что ω принимает значения в алгебре группы Лоренца.

Условие (1.42) легко разрешается и даёт для связи двух связностей

$$\Gamma^\alpha_{\mu\nu} = e^\alpha_a \left(\partial_\mu e^a_\nu + {\omega^a}_{\mu b} e^b_\nu\right) \equiv e^\alpha_a\, \mathfrak{D}_\mu e^a_\nu \tag{1.43}$$

или

$$\omega^a{}_{\mu b} = e^a_\alpha \Gamma^\alpha_{\mu\nu} e^\nu_b - e^\nu_b \partial_\mu e^a_\nu,$$

где $\mathfrak{D}_\mu$ – лоренц-ковариантная (по отношению только к латинскому индексу) производная. В частности, можно найти спин-связность $\overset{(0)}{\omega}$, которая отвечает связности Леви-Чивита $\overset{(0)}{\Gamma}(g)$ заданной на многообразии метрики g.

Как уже обсуждалось выше, условие (1.43) означает, что связности $\Gamma^\alpha_{\mu\beta}$ и $\omega^a{}_{\mu b}$ представляют собой один и тот же математический объект в разных одеяниях. Разумеется, это подтверждается также и вычислением тензора кривизны: дифференциальной 2-формы

$$R^a{}_{b\mu\nu}(\omega) = \partial_\mu \omega^a{}_{\nu b} - \partial_\nu \omega^a{}_{\mu b} + \omega^a{}_{\mu c}\omega^c{}_{\nu b} - \omega^a{}_{\nu c}\omega^c{}_{\mu b} \tag{1.44}$$

для спин связности и стандартного метрического объекта (1.6)

$$R^\alpha{}_{\beta\mu\nu}(\Gamma) = \partial_\mu \Gamma^\alpha_{\nu\beta} - \partial_\nu \Gamma^\alpha_{\mu\beta} + \Gamma^\alpha_{\mu\rho}\Gamma^\rho_{\nu\beta} - \Gamma^\alpha_{\nu\rho}\Gamma^\rho_{\mu\beta}.$$

Простые вычисления показывают, что

$$R^\alpha{}_{\beta\mu\nu}(\Gamma) = e^\alpha_a R^a{}_{b\mu\nu}(\omega) e^b_\beta \tag{1.45}$$

в соответствии с общим правилом замены типа значка у тензора.

1.1.7 Телепараллельный эквивалент

Телепараллельная гравитация основана на использовании кручения $T^\alpha{}_{\mu\nu} \equiv \Gamma^\alpha_{\mu\nu} - \Gamma^\alpha_{\nu\mu}$ вместо кривизны. Итак, мы рассматриваем геометрию, в которой нет кривизны и неметричности, но существует кручение.

Предполагая $\bigtriangledown_\alpha g_{\mu\nu} = 0$, легко убедиться, что

$$\Gamma^\alpha_{\mu\nu} = \overset{(0)}{\Gamma}{}^\alpha_{\mu\nu}(g) + K^\alpha{}_{\mu\nu}, \tag{1.46}$$

где $\overset{(0)}{\Gamma}{}^{\alpha}_{\ \mu\nu}(g)$ – леви-чивитовская связность метрики g, а тензор конторсии определён посредством

$$K_{\alpha\mu\nu} = \frac{1}{2}\left(T_{\alpha\mu\nu} + T_{\nu\alpha\mu} + T_{\mu\alpha\nu}\right) = \frac{1}{2}\left(T_{\mu\alpha\nu} + T_{\nu\alpha\mu} - T_{\alpha\nu\mu}\right). \tag{1.47}$$

У него есть очевидное свойство антисимметрии

$$K_{\alpha\mu\nu} = -K_{\nu\mu\alpha}. \tag{1.48}$$

В соответствии с тем, что уже обуждалось ранее, тензор кривизны связности отличается от обычного римановского на квадратичное по конторсии выражение:

$$R^{\alpha}{}_{\beta\mu\nu}(\Gamma) = R^{\alpha}{}_{\beta\mu\nu}(\overset{(0)}{\Gamma}) + \overset{(0)}{\nabla}{}_{\mu}K^{\alpha}{}_{\nu\beta} - \overset{(0)}{\nabla}{}_{\nu}K^{\alpha}{}_{\mu\beta} + K^{\alpha}{}_{\mu\rho}K^{\rho}{}_{\nu\beta} - K^{\alpha}{}_{\nu\rho}K^{\rho}{}_{\mu\beta}, \tag{1.49}$$

где $\overset{(0)}{\nabla}{}_{\mu}$ – ковариантная производная связности $\overset{(0)}{\Gamma}{}^{\alpha}_{\ \mu\nu}(g)$. Отсюда получаем скалярную кривизну

$$R(\Gamma) = R(\overset{(0)}{\Gamma}) + 2\overset{(0)}{\nabla}{}_{\mu}T^{\mu} + \mathbb{T}, \tag{1.50}$$

где вектор кручения определён как

$$T_{\mu} = T^{\alpha}{}_{\mu\alpha} = -T^{\alpha}{}_{\alpha\mu}, \tag{1.51}$$

а скаляр кручения может быть записан в нескольких эквивалентных формах:

$$\begin{aligned}\mathbb{T} &= \frac{1}{2}K_{\alpha\beta\mu}T^{\beta\alpha\mu} - T_{\mu}T^{\mu} \\ &= \frac{1}{2}T_{\alpha\beta\mu}S^{\alpha\beta\mu} \\ &= \frac{1}{4}T_{\alpha\beta\mu}T^{\alpha\beta\mu} + \frac{1}{2}T_{\alpha\beta\mu}T^{\beta\alpha\mu} - T_{\mu}T^{\mu}\end{aligned} \tag{1.52}$$

с "суперпотенциалом"

$$S^{\alpha\mu\nu} \equiv K^{\mu\alpha\nu} + g^{\alpha\mu}T^{\nu} - g^{\alpha\nu}T^{\mu}, \tag{1.53}$$

удовлетворяющим тому же свойству антисимметрии $S^{\alpha\mu\nu} = -S^{\alpha\nu\mu}$, что и само кручение.

В телепараллельной гравитации традиционно используется связность Вайтценбёка, опеределяемая как $\overset{\mathfrak{W}}{\omega}{}^{a}{}_{\mu b} = 0$ или

$$\overset{\mathfrak{W}}{\Gamma}{}^{\alpha}_{\ \mu\nu} = e^{\alpha}_{a}\partial_{\mu}e^{a}_{\nu}, \tag{1.54}$$

кривизна которой, очевидно, обращается в нуль, $R^{\alpha}{}_{\beta\mu\nu}(\overset{\mathfrak{W}}{\Gamma}) = 0$.

Разумеется, условие обращения спин-связности в нуль не является локально лоренц-инвариантным. В ковариантном виде можно потребовать, чтобы спин-связность была плоской.

Обозначая детерминант матрицы e^{a}_{μ} через $\|e\|$, мы видим из соотношения (1.50), что действие

$$S_{\mathfrak{W}} = -\int d^4x\|e\| \cdot \mathbb{T} \tag{1.55}$$

эквивалентно действию общей теории относительности $\int d^4x\sqrt{-g}\cdot R(\overset{(0)}{\Gamma})$, с точностью до отбрасывания поверхностных слагаемых. Разумеется, с той же точностью на уровне действия восстанавливается локальная лоренц-инвариантность даже при явном выборе нулевой спин-связности.

Уравнения движения

Как обычно, уравнения движения выводятся путем варьирования действия по отношению к полям тетрад. Это нетрудно проделать даже для случае ненулевой фиксированной спин-связности. В частности, в первом порядке по вариациям имеем

$$\begin{aligned} \delta e_a^\mu &= -e_b^\mu e_a^\nu \delta e_\nu^b, & (1.56)\\ \delta\|e\| &= \|e\| \cdot e_a^\mu \delta e_\mu^a, & (1.57)\\ \delta g_{\mu\nu} &= \eta_{ab}\left(e_\mu^a \delta e_\nu^b + e_\nu^a \delta e_\mu^b\right), & (1.58)\\ \delta g^{\mu\nu} &= -\left(g^{\mu\alpha} e_a^\nu + g^{\nu\alpha} e_a^\mu\right)\delta e_\alpha^a, & (1.59)\\ \delta_e T^\alpha{}_{\mu\nu} &= -e_a^\alpha T^\beta{}_{\mu\nu}\delta e_\beta^a + e_a^\alpha\left(\mathfrak{D}_\mu \delta e_\nu^a - \mathfrak{D}_\nu \delta e_\mu^a\right). & (1.60) \end{aligned}$$

Для квадратичных инвариантов тензора кручения это даёт:

$$\delta_e(T_\mu T^\mu) = -2\left(T^\beta T^\alpha{}_{\beta\mu} + T^\alpha T_\mu\right) e_a^\mu \delta e_\alpha^a \\ + 2\left(T^\alpha e_a^\mu - T^\mu e_a^\alpha\right) \cdot \mathfrak{D}_\alpha \delta e_\mu^a, \quad (1.61)$$

$$\delta_e(T_{\alpha\mu\nu}T^{\mu\alpha\nu}) = 2\left(T^{\beta\mu\alpha} - T^{\alpha\mu\beta}\right) T_{\mu\alpha\nu} e_a^\nu \delta e_\beta^a \\ + \left(T^\alpha{}_\mu{}^\beta - T^\beta{}_\mu{}^\alpha\right) e_a^\mu \cdot \mathfrak{D}_\alpha \delta e_\beta^a, \quad (1.62)$$

$$\delta_e(T_{\alpha\mu\nu}T^{\alpha\mu\nu}) = 4T_\alpha{}^{\mu\nu} e_a^\alpha \cdot \mathfrak{D}_\mu \delta e_\nu^a - 4T^{\alpha\mu\nu} T_{\alpha\mu\beta} e_a^\beta \delta e_\nu^a. \quad (1.63)$$

В частности, для телепараллельного эквивалента общей теории относительности (1.55) получаем

$$\delta_e S = -\int d^4x \|e\| \cdot \left(-2S^{\alpha\mu\nu} T_{\alpha\beta\nu} e_a^\beta \delta e_\mu^a + \mathbb{T} e_a^\mu \delta e_\mu^a - 2S_\beta{}^{\mu\alpha} e_a^\beta \mathfrak{D}_\alpha \delta e_\mu^a\right),$$

где лоренц-ковариантная производная $\mathfrak{D}$ в нашем случае совпадает с простой частной производной, поскольку $\omega^b{}_{\alpha a} = 0$ по определению геометрии Вайтценбёка.

Интегрирование по частям в последнем слагаемом даёт

$$2\delta e_\mu^a \cdot \left(\partial_\alpha\left(\|e\| \cdot S_\beta{}^{\mu\alpha} e_a^\beta\right) - \|e\| \cdot \omega^b{}_{\alpha a} S_\beta{}^{\mu\alpha} e_b^\beta\right) \\ = 2\|e\| \cdot \left(\overset{(0)}{\nabla}_\alpha S_\beta{}^{\mu\alpha} - K^\nu{}_{\alpha\beta} S_\nu{}^{\mu\alpha}\right) \cdot e_a^\beta \delta e_\mu^a$$

где мы учли антисимметрию S, а также различие между Γ и $\overset{(0)}{\Gamma}$ с помощью второго слагаемого в правой части. В самом деле, в силу антисимметрии тензора S имеем

$$\overset{(0)}{\nabla}{}_{\nu} S_a{}^{\mu\nu} = \frac{1}{\|e\|}\partial_\nu \left(\|e\| S_a^{\mu\nu}\right) - \overset{(0)}{\omega}{}^b{}_{\nu a} S_b{}^{\mu\nu},$$

а разницу в связностях учитываем посредством $\omega^b{}_{\nu a} - \overset{(0)}{\omega}{}^b{}_{\nu a} = K^b{}_{\nu a}$.

Используя невырожденность тетрад, уравнения движения можно записать в виде

$$\overset{(0)}{\nabla}{}_{\alpha} S_\beta{}^{\mu\alpha} - S^{\alpha\mu\nu}\left(T_{\alpha\beta\nu} + K_{\alpha\nu\beta}\right) + \frac{1}{2}\mathbb{T}\delta^\mu_\beta = 0, \tag{1.64}$$

эквивалентность которого общей теории относительности можно установить, напрямую подставив тождество

$$R^\alpha{}_{\beta\mu\nu}(\overset{(0)}{\Gamma}) = -\left(\overset{(0)}{\nabla}{}_{\mu} K^\alpha{}_{\nu\beta} - \overset{(0)}{\nabla}{}_{\nu} K^\alpha{}_{\mu\beta} + K^\alpha{}_{\mu\rho}K^\rho{}_{\nu\beta} - K^\alpha{}_{\nu\rho}K^\rho{}_{\mu\beta}\right)$$

в уравнения Эйнштейна $G^\mu_\beta = 0$.

Взаимодействие с материей можно ввести самосогласованным образом. Для бозонных полей всё просто: они обычным образом взаимодействуют с метрикой (1.41). Для фермионных полей необходимо вводить спин-связность $\overset{(0)}{\omega}$, что не столь естественно для телепараллельного подхода, но тоже возможно.

1.2 Основы теоретической космологии

Представление гравитации как кривизны пространства-времени позволило сформулировать космологические вопросы на научном языке, на языке теоретической физики. Современной космологии посвящено много хороших учебников и монографий, например двухтомник [5,6]. Мы кратко изложим современную космологическую картину мира, чтобы описать современное положение дел и дать (надеемся) убедительную мотивировку для занятий модифицированной гравитацией. При этом, поскольку фактический материал в основном представлен общеизвестными фактами, мы позволим себе ограничиться лишь избранными ссылками на литературу, отсылая заинтересованного читателя, например, к двухтомнику [5,6] и ссылкам в нём.

1.2.1 Релятивистская космология

В основе космологических моделей лежит предположение о том, что Вселенная пространственно однородна и изотропна, и соответственно описывается метрикой Фридмана-Робертсона-Уокера. Мы не будем повторять несложных общеизвестных выкладок, которые при подстановке этой метрики в уравнения Эйнштейна приводят к уравнениям Фридмана

$$3\left(H^2+\frac{K}{a^2}\right)=8\pi\mathcal{G}\rho, \tag{1.65}$$

$$2\dot{H}+3H^2+\frac{K}{a^2}=-8\pi\mathcal{G}p, \tag{1.66}$$

в которых $H\equiv\frac{\dot{a}}{a}$ – постоянная Хаббла, ρ – плотность материи, p – её давление, а K равно 0 или ±1, причём в последних случаях радиус пространственной кривизны положен равным единице.

Мы будем практически всегда рассматривать пространственно плоский случай $K=0$, метрика которого может быть записана как

$$ds^2=-dt^2+a^2(t)\overrightarrow{dx}^2.$$

Видно, что для пространственно плоского мира заданному значению постоянной Хаббла $H\equiv\frac{\dot{a}}{a}$ соответствует вполне определённая плотность материи, наполняющей Вселенную – "критическая плотность"

$$\rho_{\mathrm{cr}}=\frac{3H^2}{8\pi\mathcal{G}}. \tag{1.67}$$

Отметим, что численное значение константы Хаббла в современную эпоху известно отнюдь не столь хорошо, как хотелось бы. В связи с этим для удобства и минимизации погрешностей при анализе данных вводят параметр

$$h \equiv \frac{H}{100\frac{\mathrm{km}}{\mathrm{s\cdot Mpc}}}, \tag{1.68}$$

который равен примерно 0.7. Для критической плотности имеем

$$\rho_{\mathrm{cr}} = 1.878 \cdot h^2 \cdot 10^{-29} \frac{\mathrm{g}}{\mathrm{cm}^3}.$$

По современным данным суммарная плотность вещества в нашей Вселенной весьма (с процентной точностью) близка к критической (впрочем, 95% этой плотности имеет непонятную природу). Отметим, что в замкнутой вселенной $\rho > \rho_{\mathrm{cr}}$, а в открытой – наоборот $\rho < \rho_{\mathrm{cr}}$.

Чтобы решить уравнения Фридмана, требуется ещё знание уравнения состояния вещества, наполняющего Вселенную. В космологии часто бывает достаточно предположить следующий простой вид зависимости:

$$p = w\rho, \tag{1.69}$$

с постоянным параметром уравнения состояния w.

Если присутствует лишь одна компонента материи с постоянным параметром $w > -1$, то легко найти решение и получить степенной закон расширения

$$a(t) = A \cdot (t - t_0)^{\frac{2}{3(w+1)}} \tag{1.70}$$

с двумя постоянными интегрирования, A и t_0. Последнюю можно без ограничения общности положить равной нулю, договорившись отсчитывать время от начальной сингулярности.

Случай $w = -1$ отвечает космологической постоянной и (экспоненциальному) решению де Ситтера. Если $w < -1$ (так называемые фантомные поля), то вселенная заканчивает своё существование в сингулярности Большого Разрыва.

Рассмотрим наиболее важные примеры.

Доминирование материи (пылевидная стадия)

В случае нерелятивистской материи $w = 0$. Доминирует энергия покоя, и давление пренебрежимо мало в сравнении с ней. В космологии такое вещество обычно называется пылью. Решение принимает вид

$$a(t) \propto t^{\frac{2}{3}}.$$

Отметим, что

$$\rho \propto H^2 \propto \frac{1}{t^2} \propto \frac{1}{a^3}.$$

Это является простым отражением того факта, что предоставленный объём растет как a^3 при неизменном полном числе частиц, которое определяет массу.

Доминирование излучения

Как известно, для газа фотонов (и любых других ультрарелятивистских частиц) $p = \frac{1}{3}\rho$, то есть $w = \frac{1}{3}$. Тогда имеем

$$a(t) \propto \sqrt{t},$$

то есть излучение замедляет Вселенную больше, чем пыль. Этот, возможно контринтуитивный, результат является проявлением того, что давление тоже служит источником гравитационного поля. При этом плотность энергии убывает с ростом масштабного фактора быстрее, чем в случае пыли

$$\rho \propto H^2 \propto \frac{1}{t^2} \propto \frac{1}{a^4},$$

поскольку помимо уменьшения плотности числа частиц, энергия каждой частицы убывает как $\frac{1}{a}$ вследствие красного смещения.

Сегодня вклад пыли в энергетический баланс Вселенной намного превосходит таковой от излучения, но очевидно, что так было не всегда. Ранняя Вселенная должна была быть радиационно-доминированной.

Параметр замедления

Уже сравнительно давно был введён в обиход параметр замедления

$$q = -\frac{a\ddot{a}}{\dot{a}^2}, \tag{1.71}$$

который в обезразмеренном виде показывает, сколь существенен эффект замедления Вселенной. Сегодня мы знаем, что Вселенная расширяется ускоренно, что отвечает случаю $q < 0$. По историческим причинам в литературе предпочитают рассматривать отрицательный параметр замедления, а не вводить положительный параметр ускорения.

Из закона расширения (1.70) легко видеть, что для ускоренного расширения требуется вещество с отрицательным давлением, причём с $w < -\frac{1}{3}$.

1.2.2 Параметры космологической модели

Реальная Вселенная заполнена веществом, состоящим из нескольких различных компонент. Этот состав традиционно описывается параметрами

$$\Omega_i = \frac{\rho_i}{\rho_{cr}} = \frac{8\pi\mathcal{G}\rho_i}{3H^2}, \tag{1.72}$$

которые всегда относят к текущей эпохе. Очевидно, $\sum\Omega_i = 1$ в плоской вселенной, $\sum\Omega_i > 1$ в замкнутой, и $\sum\Omega_i < 1$ в открытой. Формально можно ввести и вклад пространственной кривизны, тогда сумма всех Ω будет тождественно равна единице.

Вклад излучения можно оценить с помощью планковского распределения для температуры реликтового фона, которая равна примерно 2.7K. Результат соответствует примерно 410 фотонам в кубическом сантиметре, или $\Omega_r < 10^{-4}$. Отметим сразу, что излучение реликтового фона с хорошей точностью чернотельно и изотропно.

Самая большая анизотропия связана с движением Земли по отношению к реликтовому фону и определяется эффектом Допплера

$$T = T_0\frac{\sqrt{1-v^2}}{1-v\cos\theta},$$

причём, поскольку скорость нашего движения много меньше скорости света, $v \ll 1$, анизотропия в основном носит дипольный характер:

$$\frac{\Delta T}{T_0} \approx v \cos\theta.$$

Кинематический диполь имеет порядок $\Delta T \approx 3.35\text{mK}$, или $\frac{\Delta T}{T_0} \sim 10^{-3}$, в то время как первичные флуктуации, регистрируемые в следующих мультиполях, – порядка $\sim 10^{-5}$.

Что касается пыли, то её суммарный вклад в плотность энергии составляет около 30%. Однако же обычное (как говорят в астрономии, барионное) вещество составляет не более 5% (а реально видимое и того меньше). Оставшиеся же примерно 25% приходятся на долю так называемой Тёмной Материи.

Вклад Тёмной Энергии оказывается около 70%. Она вызывает ускоренное расширение Вселенной, не склонна к кластеризации (скучиванию), и её уравнение состояния совместимо с гипотезой о космологической постоянной, $w = -1$.

Вклад пространственной кривизны не найден, и по-видимому не превосходит процента, или даже половины процента. Отметим, что из уравнений Фридмана следует, что вклад кривизны меняется как $\frac{1}{a^2}$, то есть убывает медленнее, чем вклады обычной материи. Следовательно, ранняя Вселенная была ещё более плоской.

История Вселенной в терминах красного смещения

Поскольку абсолютное измерение расстояний в космологических масштабах – дело весьма тонкое, то в космологии принято (везде, где возможно) в качестве переменной времени использовать красное смещение

$$z = \frac{a(t_0)}{a(t)} - 1,$$

которое является непосредственно (и весьма точно) измеримой величиной (здесь t_0 - возраст Вселенной на сегодняшний момент, а t - рассматриваемый момент времени). Использование физического времени и/или красного смещения можно сравнить с абсолютной и стратиграфической шкалами времени в геологических науках.

Если же фиксирована конкретная космологическая модель, то все данные, полученные в терминах красного смещения, можно пересчитать к любой удобной переменной времени. В самом деле, достаточно получить константу Хаббла как явную функцию красного смещения. Покажем, как это делается.

Параметры космологической модели равны

$$\Omega = \frac{8\pi\mathcal{G}\rho_0}{3H_0^2},$$

где индекс 0 указывает на сегодняшнее значение величины. Если данная компонента меняется с расширением Вселенной как $\rho \propto \frac{1}{a^n}$, то имеем

$$\rho(t) = \rho_0 \left(\frac{a(t_0)}{a(t)}\right)^n = \Omega\rho_{cr}(1+z)^n$$

Соответственно, константа Хаббла на красном смещении z равна

$$3H^2(z) = 8\pi\mathcal{G}\rho = 8\pi\mathcal{G}\rho_{cr}\sum_i \Omega_i(1+z)^{n_i} = 3H_0^2\sum_i \Omega_i(1+z)^{n_i},$$

или окончательно

$$H(z) = H_0\sqrt{\sum_i \Omega_i(1+z)^{n_i}}\ . \tag{1.73}$$

Разумеется, при желании можно написать обобщение для любых других уравнений состояния.

1.2.3 Тепловая история Вселенной

Современная Вселенная является на больших масштабах однородной и изотропной. Как мы уже упоминали, она заполнена Тёмной Энергий, пылевидной материей (состоящей из Тёмной Материей и обычного вещества), а также излучением, вклад которого в плотность энергии сильно подавлен, но по энтропии оно доминирует, поскольку плотность числа барионов весьма мала. Отношение барионов к фотонам равно

$$\eta_B \equiv \frac{n_B}{n_\gamma} \approx 6 \cdot 10^{-10} \tag{1.74}$$

Учитывая плотность фотонов реликтового фона, это даёт примерно $n_B \sim 2 \cdot 10^{-7}\text{cm}^{-3}$. Отметим, что для удобства часто вводится параметр

$$\eta_{10} \equiv \eta_B \cdot 10^{10} \approx 6.$$

На термодинамику вещества, разумеется, оказывает влияние расширение Вселенной. Для понимания этого эффекта необходимо знать, как сказывается изменение масштабного фактора на распределении частиц вещества по импульсам.

Пространственный импульс в "пространственно-инвариантном" виде можно записать как

$$p = m\sqrt{g_{ij}\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}},$$

где τ – собственное время частицы. В самом деле, приведя метрику в данной точке к метрике Минковского, имеем $g_{ij} = \delta^i_j$ и $d\tau = dt \cdot \sqrt{1-v^2}$, где $v^i \equiv \frac{dx^i}{dt}$. Следовательно $p = \frac{mv}{\sqrt{1-v^2}}$, как и должно быть.

В пространственно-плоской модели Фридмана (1.4) уравнение геодезической принимает вид

$$\frac{d^2x^i}{d\tau^2} = -\Gamma^i_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = -\frac{2}{a}\frac{da}{dt}\frac{dx^i}{d\tau}\frac{dt}{d\tau}.$$

Умножая обе части на $\frac{d\tau}{dt}$, имеем

$$\frac{d}{dt}\left(\frac{dx^i}{d\tau}\right) = -\frac{2}{a}\frac{da}{dt}\frac{dx^i}{d\tau},$$

а следовательно $\frac{dx^i}{d\tau} \propto \frac{1}{a^2}$. Учитывая, что в космологических координатах $g_{ij} \propto a^2$, получаем

$$p \propto \frac{1}{a}.$$

Как видим, результат не зависит от массы, и можно убедиться, что он справедлив и для безмассовых частиц, приводя к обычному ответу для красного смещения.

Вспоминая, что равновесное распределение (планковское) для газа фотонов зависит от отношения $\frac{h\nu}{kT}$, видим, что термодинамическое равновесие фотонного газа, заполняющего Вселенную, не нарушается её расширением, лишь температура его убывает как $T \propto \frac{1}{a}$.

Аналогичный результат справедлив для пылевидной среды. В самом деле, пусть есть нерелятивистский газ с нулевым (или, в релятивистском смысле, равным массе частицы) химическим потенциалом μ. Тогда имеем $E - \mu \approx \frac{p^2}{2m}$. И поскольку равновесное распределение зависит от величины $\frac{E-\mu}{kT}$, то оно остаётся при расширении Вселенной равновесным с температурой $T \propto \frac{1}{a^2}$.

Ясно, что если во Вселенной присутствуют несколько компонент, одна релятивистская, другая нерелятивистская, то расширение Вселенной может приводить к нарушению термодинамического равновесия, поскольку температуры двух компонент зависят от роста масштабного фактора поразному. В реальной Вселенной нарушение барионной (и, по-видимому, лептонной) симметрии очень мало, и количество частиц вещества в миллиард раз меньше количества фотонов реликтового фона, коим и определяется энтропия, а следовательно и изменение температуры.

* * *

В современной Вселенной на малых масштабах присутствуют очень большие неоднородности, и практически не имеет смысла ставить вопрос о сравнении температуры вещества в среднем с температурой реликтового излучения. Ясно, что излучение почти не взаимодействует с веществом, а дальнейшие вопросы сродни выяснению средней температуры по больнице.

Вместе с тем, в достаточно ранней Вселенной ($z \geqslant 1100$) температура была высока, вещество находилось в ионизованном состоянии, и в целом вся первичная плазма была в высокой степени однородной (те самые 10^{-5} для относительной флуктуации, о которых мы упоминали в связи с анизотропией реликтового фона). В такой ситуации будет вполне разумным задать вопрос о том, находилось ли вещество в равновесии с излучением.

К счастью, для выяснения этого вопроса применимы самые элементарные методы физической кинетики. Взаимодействие фотонов с заряженными массивными нерелятивистскими частицами описывается томсоновским сечением

$$\sigma_{\mathcal{T}} = \frac{e^4}{6\pi m_e^2} \approx 0.665 \cdot 10^{-24} \mathrm{cm}^2.$$

Доминирует, естественно, рассеяние на электронах как существенно более лёгких частицах. Время свободного пробега фотона даётся выражением

$$\frac{1}{\sigma_{\mathcal{T}} n_e c},$$

которое надо сравнить с характерным временем расширения Вселенной – Хаббловским временем $H^{-1} = \frac{a}{\dot{a}}$. Первое время убывает с ростом температуры как $\frac{1}{T^3}$ вместе с концентрацией нерелятивистских частиц. Второе же зависит от температуры медленнее, поскольку

$$H = H_0\sqrt{\Omega_R(1+z)^4} = H_0\sqrt{\Omega_R}\left(\frac{T}{T_0}\right)^2.$$

Подставив конкретные параметры, легко проверить, что вплоть до рекомбинации космического вещества время свободного пробега фотона много меньше характерного времени изменения внешних условий. Электронов намного меньше, чем фотонов, и для каждого конкретного электрона столкновения с фотонами происходят ещё чаще. Следовательно, расширение Вселенной оказывается квазистационарным процессом, и равновесие не нарушается.

Представляет однако интерес и более сложный вопрос о том, восстановится ли равновесие, если в какой-то момент времени будет нарушена равновесность распределения фотонов, например в результате распада каких-нибудь гипотетических тяжёлых нестабильных частиц. Оказывается, ответ на этот вопрос уже не столь очевиден, поскольку при каждом соударении с тяжёлой неподвижной частицей энергия фотона $E \sim kT$ меняется лишь на очень малую величину, $\frac{\Delta E}{E} = \frac{kT}{m_e c^2}$. Ясно, что характерное время, за которое может существенно измениться энергия фотона в результате столкновений с электронами, становится равным

$$\frac{m_e c}{\sigma_T n_e kT}$$

и зависит уже от четвертой степени температуры. Явные оценки показывают, что уже при $z \lesssim 10^5$ равновесие не достигается. Процессы с переменным числом фотонов выходят из равновесия ещё раньше.

Описанная здесь ситуация является довольно общей. Константа Хаббла на радиационно-доминированной стадии пропорциональна квадрату температуры, а характерные скорости микропроцессов определяются, как правило, степенями температуры не ниже третьей (которая отвечает изменению объема). Следовательно, в достаточно ранней Вселенной имеется равновесие, в то время как отсутствие явных проявлений неравновесности в более поздние эпохи позволяет накладывать ограничения на различные гипотетические модели [7].

Эпоха рекомбинации

Разумеется, есть определённые этапы в развитии Вселенной, когда равновесие было заведомо нарушено. Можно сказать, что они связаны с изменениями "агрегатного состояния" вещества во Вселенной. Одним из таких периодов было время рекомбинации, когда первичная плазма, состоявшая из ядер водорода и гелия, а также свободных электронов, перешла в состояние нейтрального вещества. Рекомбинация гелия происходит раньше, поскольку соответствующая энергия связи больше. Но окончательное отщепление излучения и нейтрализация вещества происходят, разумеется, при рекомбинации водорода.

Напомним, что в равновесной статистической физике такой процесс рекомбинации описывается в терминах равновесия по отношению к реакции $p^+ + e^- \longleftrightarrow H + \gamma$, которое выражается условием равенства химических потенциалов $\mu_p + \mu_e = \mu_H$. Мы не будем повторять всех деталей этих элементарных выкладок. Скажем лишь, что можно воспользоваться равновесными распределениями для всех участников реакции, максвелл-больцмановским для массивных частиц и планковским для фотонов, а также условием электронейтральности плазмы $n_e = n_p$, чтобы вывести уравнение Саха для степени ионизации первичной плазмы $X_p \equiv \frac{n_p}{n_p+n_H}$

$$X_p + (n_p + n_H)X_p^2 \left(\frac{2\pi\hbar^2}{m_e kT}\right)^{\frac{3}{2}} e^{\frac{U_H}{kT}} = 1,$$

где U_H - энергия связи атома водорода.

Уравнение Саха известно с начала XX века в применении к звёздным атмосферам, и во многих случаях оно прекрасно описывает ситуацию. В применении к первичной плазме во Вселенной при реалистичных параметрах оно относит процесс рекомбинации к температурам порядка 4000K, в то время как более аккуратный анализ показывает, что космологическая рекомбинация происходила сравнительно медленно при температурах (3000$\pm$300)K, а уравнение Саха достаточно хорошо описывает первичную плазму лишь при отличиях степени ионизации от единицы, не превышающих пары процентов.

Причина заключается в том, что как в атмосфере звезды, так и в лабораторной пробирке, фотоны лаймановской серии, излучаемые на последнем шаге рекомбинационного каскада, могут более или менее свободно покинуть объём вещества, что, разумеется, совсем не так для первичной плазмы в расширяющейся Вселенной. Даже фотон Ly_α способен возбудить соседа, или полностью ионизовать находящийся на неосновном уровне атом. Аккуратное рассмотрение естественной ширины линий и скорости расширения Вселенной показывает, что космологическое красное смещение не обладает достаточной эффективностью для нивелирования данного эффекта, что неудивительно, поскольку характерные времена расширения на данном этапе исчисляются сотнями тысяч лет.

В кинетику рекомбинации в этих условиях большой вклад вносит запрещённый процесс $2s \to 1s$, проходящий по двухфотонному каналу $H_{2s} \longrightarrow H_{1s} + 2\gamma$ (характерное время по порядку величины соответствует секунде). Получаемые фотоны уже не обладают достаточной энергией, чтобы противостоять дальнейшей рекомбинации.

Существуют стандартные программы для численного моделирования космологической рекомбинации [8]. Это очень важно для анализа наблюдательных данных в космологии. Мы не будем вдаваться в подробности методов физической кинетики, которые, впрочем, вполне стандартны. Пожалуй, единственным специфическим эффектом является (весьма медленное) уменьшение концентраций всех частиц за счёт роста масштабного фактора. Оно, кстати, приводит к тому, что по мере стремления температуры к нулю остаточная ионизация к нулю не стремится, а замораживается на некотором уровне, меньшем 10^{-3}.

Впрочем, остаточный уровень ионизации имеет скорее академический интерес. Вещество всё равно становится практически прозрачным для реликтового излучения (напомним, что характерные частоты фотонов много меньше резонансных частот оптической структуры атомов, поскольку, как и должно быть в разреженной плазме, даже без неравновесной задержки, рекомбинация происходит при температурах много меньших, чем потенциал ионизации), а в эпоху рождения первых звёзд происходит процесс повторной ионизации (reionisation). Он однако не представляет очень серьёзного препятствия для распространения фотонов реликтового фона в силу большой разреженности вещества к этому времени. Оптическая толщина слоя повторной ионизации не превышает десяти процентов.

О реликтовых нейтрино

Для равновесных газов любых других частиц при достаточно высоких температурах, когда они являются ультрарелятивистскими, все термодинамические характеристики могут быть вычислены аналогично случаю фотонов в предположении обращения в нуль химического потенциала или, что то же самое, симметрии между веществом и антивеществом (равновесие по отношению к реакции аннигиляции). Наблюдательные данные указывают на то, что такая симметрия выполняется с точностью примерно до миллиардной доли, по крайней мере для барионов. В случае с лептонами, это тоже скорее всего соответствует действительности, но ситуация осложняется ненаблюдаемостью фона реликтовых нейтрино.

Явные формулы выглядят достаточно просто. Для бозонов имеем равновесную плотность энергии

$$\rho_{BE} = g\frac{k^4T^4}{2\pi^2\hbar^3}\int\limits_0^\infty \frac{\mathrm{z}^3 d\mathrm{z}}{e^{\mathrm{z}}-1},$$

где g – статистический множитель (вырожденность уровней). Энтропия в расчёте на один барион равна

$$k\sigma_{BE} = \frac{2k^4T^3}{3\pi^2\hbar^3 n_B}\int\limits_0^\infty \frac{\mathrm{z}^3 d\mathrm{z}}{e^{\mathrm{z}}-1}.$$

Для фермионов аналогичные формулы выглядят как

$$\rho_{FD} = g\frac{k^4T^4}{2\pi^2\hbar^3}\int\limits_0^\infty \frac{\mathrm{z}^3 d\mathrm{z}}{e^{\mathrm{z}}+1}$$

и

$$k\sigma_{FD} = \frac{2k^4T^3}{3\pi^2\hbar^3 n_B}\int\limits_0^\infty \frac{\mathrm{z}^3 d\mathrm{z}}{e^{\mathrm{z}}+1}.$$

В частности, эти формулы можно применить для описания нейтринной компоненты (в предположении пренебрежимой малости химических потенциалов), которая дольше многих других остаётся ульрарелятивистской. Однако, в силу того, что нейтрино очень слабо взаимодействуют с остальным веществом, интересно задаться вопросом, остаются ли они в равновесии с ним, например, при температуре аннигиляции электронов и позитронов (эпоха, в которую электроны переходят от ультрарелятивистского режима к нерелятивистскому). Соответствующие оценки выполняются так же, как и для равновесия излучения. Для сечения взаимодействия здесь применима четырёхфермионная теория Ферми с константой взаимодействия $G_F \approx 1.17 \cdot 10^{-5}$ GeV^{-2}.

Из соображений размерности ясно, что сечение взаимодействия частиц пропорционально квадрату энергии в системе центра масс (по крайней мере, если их массами можно пренебречь): $\sigma \propto G_F^2 T^2$. Разумеется, рост сечения с энергией – это проявление неперенормируемости модели (отрицательная массовая размерность константы связи), или с более фундаментальной точки зрения – рост вероятности обмена массивным промежуточным векторным бозоном. Время свободного пробега равно

$$\tau = \frac{1}{n_e \sigma c} \propto \frac{1}{G_F^2 T^5}.$$

По порядку величины это около одной секунды при $T \sim 10^{10}$ K, или около одного МэВа. Характерное время расширения Вселенной при этой энергии тоже около одной секунды, и меняется как обратный квадрат температуры. Таким образом, нейтрино выходят из равновесия при энергиях около 1 МэВа и возрасте Вселенной порядка секунды. Причём, мюонные и тау-лептонные нейтрино делают это несколько раньше, поскольку мюоны и тау-лептоны давно уже к этому времени аннигилировали, а взаимодействие с электронами осуществимо для них только посредством нейтральных слабых токов, в то время как для электронных нейтрино доступны также заряженные слабые токи.

Итак, мы видим, что аннигиляция электронов и позитронов происходит уже после отщепления нейтрино. Поэтому, они не получают своей доли в выделяющейся энергии, и их температура оказывается несколько ниже фотонной. Пользуясь явными формулами для энтропии, легко найти количественный результат, преполагая равновесность процесса аннигиляции. Рассмотрим отношение энтропии фотонов и электронов к энтропии нейтрино, до аннигиляции

$$\frac{\sigma_\gamma + \sigma_{e^\pm}}{\sigma_\nu} = \frac{2 + \frac{7}{8} \cdot 4}{\frac{7}{8} \cdot 6} = \frac{22}{21}$$

и после аннигиляции

$$\frac{\sigma_\gamma + \sigma_{e^\pm}}{\sigma_\nu} = \frac{2T_\gamma^3 + 0}{\frac{7}{8} \cdot 6T_\nu^3} = \frac{8}{21} \left(\frac{T_\gamma}{T_\nu} \right)^3,$$

где мы учли существование трёх типов нейтрино, а также следующее почти очевидное соотношение:

$$\int_0^\infty \frac{\mathrm{z}^3 d\mathrm{z}}{e^{\mathrm{z}}+1} = \frac{7}{8}\int_0^\infty \frac{\mathrm{z}^3 d\mathrm{z}}{e^{\mathrm{z}}-1}.$$

Приравнивая два отношения друг к другу, имеем

$$\frac{T_\gamma}{T_\nu} = \sqrt[3]{\frac{11}{4}} \approx 1.4,$$

и окончательно для температуры реликтового газа нейтрино сегодня получаем $T_\nu \approx 1.95$ K.

Разумеется, нейтрино при таких температурах уже не должны быть релятивистскими, но в силу исчезающе малого взаимодействия их распределение по импульсам не термализуется и должно отвечать равновесному для газа безмассовых частиц такой температуры. К сожалению, экспериментальная регистрация газа нейтрино такой температуры не представляется на сегодняшний день реалистичной.

Первичный нуклеосинтез

Ещё один интересный период в жизни Вселенной приходится на энергии порядка ста кэВ. В это время протоны и нейтроны образуют первичные атомные ядра – происходит нуклеосинтез. Впрочем, история начинается с температур, когда слабые взаимодействия перестают поддерживать равновесное распределение нуклонов по двум типам, $kT \approx 0.8$ MeV. Получить эту оценку несколько сложнее, чем в случае нейтрино, поскольку нуклоны нельзя считать безмассовыми, однако никаких принципиальных трудностей здесь нет. А равновесное отношение концентраций на момент отщепления равно

$$\frac{n_n}{n_p} = e^{-\frac{m_n - m_p}{kT}} \approx e^{-\frac{1.3}{0.8}} \approx \frac{1}{5},$$

или в терминах доли нейтронов среди всех нуклонов $X_n \equiv \frac{n_n}{n_p+n_n} \approx \frac{1}{6}$. После этого нейтроны начинают распадаться,

$$X_n(t) \approx \frac{1}{6} \cdot e^{-\frac{t}{\tau_n}},$$

до тех пор, пока начавшийся нуклеосинтез не упакует их по атомным ядрам.

Поскольку концентрация нуклонов не очень велика, нуклеосинтез проходит через двухчастичные реакции, где первый шаг – образование дейтерия: $p^+ + n \to \mathcal{D} + \gamma$. Затем образуются гелий-3 и тритий, в основном по каналам $\mathcal{D}+\mathcal{D} \to He^3 + n$ и $\mathcal{D}+\mathcal{D} \to H^3 + p^+$, не требующим участия электромагнитного взаимодействия. Потом появляется основной изотоп гелия – альфа-частица: $He^3 + \mathcal{D} \to He^4 + p^+$, $H^3 + \mathcal{D} \to He^4 + n$.

Уже на данном этапе понятно, что это неравновесный процесс, поскольку энергетически наиболее выгодно образование гелия-4, но оно не может случиться без преодоления дейтериевого бутылочного горлышка – образования весьма слабо связанных ядер дейтерия. Впрочем, для первичного накопления дейтерия можно написать уравнение типа Саха, а потом методами кинетики установить, что последующие реакции эффективно включаются при энергиях порядка 70 кэВ, или времени порядка нескольких минут, когда $X_n = \frac{1}{8}$, что даёт четверть массовой доли гелия, и три четверти водорода.

Точные количественные результаты получают с помощью численных методов [9], которые показывают, что образование гелия-4 действительно происходит весьма эффективно – массовые доли непрогоревших дейтерия и гелия-3 (включая тритий, позднее распадающийся в гелий-3) имеют порядок 10^{-5}. Численные результаты определяются параметром η_{10}, и результаты измерения первичной концентрации дейтерия прекрасно согласуются с ограничениями, происходящими из изучения флуктуаций реликтового фона.

Отметим, что при этом также важно предположение о количестве лёгких типов частиц (фотоны и три поколения нейтрино), ибо от них зависит скорость расширения Вселенной, а следовательно и доля нейтронов, успевающих распасться к моменту начала нуклеосинтеза. Существование пяти поколений лёгких нейтрино вошло бы в решительное противоречие с измеряемыми первичными концентрациями ядер.

Более тяжелых ядер практически не образуется в силу отсутствия сколько-нибудь стабильных изотопов с массовыми номерами 5 и 8. Наиболее распространённый из них – Li^7, который может образоваться как напрямую, так и через нестабильное ядро Be^7, а также может распадаться на две альфа-частицы, захватив протон, что приводит к немонотонной зависимости от η_{10}. Его первичная концентрация имеет порядок 10^{-10}, и с ней связана определённая проблема, поскольку теория предсказывает $\approx 3 \cdot 10^{-10}$, а наблюдения дают $2.0 \div 2.4 \cdot 10^{-10}$.

Конечно, удивительно, что подобное расхождение вообще возникает как проблема в астрономических науках, но тем не менее она рассматривается весьма серьёзно с различными гипотетическими решениями, вплоть до модификаций низкоэнергетической ядерной физики. Нам представляется наиболее вероятным конвективное разрушение ядер лития в звёздных атмосферах [10].

1.2.4 О связях с проблемами физики элементарных частиц

При ещё более ранних временах температура вещества во Вселенной достигала таких значений, что для описания его состояния требуется физика элементарных частиц, в том числе при энергиях, недостижимых с помощью современных ускорителей. Поскольку эти вопросы имеют лишь косвенное отношения к исследованиям, представленным в данной Диссертации, мы ограничимся кратким упоминанием основных идей.

Фазовые переходы в ранней Вселенной

Уже в рамках стандартной модели физики элементарных частиц имеются интересные этапы, связанные с существенным изменением состояния вещества.

При температурах около 100 GeV происходит спонтанное нарушение электрослабой симметрии и приобретение частицами массы за счёт механизма Хиггса [11]. Если бы это был фазовый переход первого рода, он мог бы стать источником потенциально наблюдаемых эффектов. Однако, измеренное значение массы бозона Хиггса предполагает, что переход был скорее гладким кросс-овером, если только не вводить расширений стандартной модели, таких как суперсимметрия.

Около 200 MeV происходит переход от кварк-глюонной плазмы к адронной фазе, а также тесно с ним связанное нарушение (приближённой) киральной симметрии. От перехода первого рода можно было бы ожидать, например, рождения кварковых самородков. Поскольку важную роль в этом процессе играют лёгкие кварки, вопрос о типе перехода оказывается нетривиальным даже для решёточных вычислений, но экспертное мнение склоняется к тому, что здесь тоже был гладкий кроссовер [12].

Таким образом, в рамках стандартной модели физики элементарных частиц, фазовые переходы были, скорее всего, весьма скучными, не оставившими после себя заметных следов. Однако, разумеется, это заключение зависит от конкретной модели, и фазовый переход может оказаться переходом первого рода, например, в рамках суперсимметричного расширения. В свою очередь, достаточно бурный переход может породить дополнительные гравитационные волны и другие последствия, которые в принципе могут использоваться для нахождения ограничений на параметры теоретических моделей.

Возможно, ещё более интересные события могли происходить в очень ранней Вселенной, если реализовывались фазы Великого Объединения. В частности, в теории могут оставаться магнитные монополи и прочие топологические дефекты в количествах, намного превышающих наблюдательные ограничения (заметим, что это весьма мягкая формулировка).

Бариосинтез

От физики элементарных частиц космология вправе ожидать разрешения ряда загадок, таких как наличие барионной асимметрии. По всей видимости, во Вселенной в значимых количествах присутствует вещество, но не антивещество. Разумеется, отношение барионов к фотонам очень мало, и это означает, что при высоких температурах количество частиц превышало количество античастиц лишь на миллиардную долю. Но и это превышание следует так или иначе объяснить.

Конечно же, за счёт расширения Вселенной аннигиляция не может произойти полностью даже в абсолютно симметричной фазе, но явные оценки показывают, что остаточная концентрация частиц с массой нуклона была бы в нашей Вселенной ещё на много порядков ниже, чем наблюдаемая плотность вещества.

Необходимые для бариосинтеза условия уже сравнительно давно были сформулированы А.Д. Сахаровым. Это нарушение сохранения барионного заряда, нарушение C и CP четности, нарушение термодинамического равновесия. Все эти условия должны удовлетворяться одновременно.

В рамках стандартной модели физики элементарных частиц можно было надеяться на электрослабый фазовый переход. Однако, как мы уже обсуждали выше, по современным представлениям он не был связан с сильным нарушением термодинамического равновесия. К тому же нарушение CP симметрии в стандартной модели тоже очень слабое, происходящее только из фазы матрицы Каббибо-Кобаяши-Маскавы – смешивания кварков, которая появляется только начиная с трёх поколений, и при этом в реальности смешивание с третьим поколением существенно более слабое, чем внутри первых двух.

Что касается, нарушения барионного заряда, то в стандартной модели оно существует лишь на непертурбативном уровне в виде инстантонных эффектов, чрезвычайно подавленных при нулевой температуре. Впрочем, в горячей Вселенной может произойти перекатывание через потенциальный барьер посредством тепловой флуктуации (сфалерон). При этом для успешного бариосинтеза такие процессы должны быть эффективны во время бурного фазового перехода, но не после восстановления термодинамического равновесия. Эти условия в рамках стандартной модели не выполнены.

Возможно, решение надо искать в области моделей Великого Объединения. Уже в рамках простейшей $SU(5)$ модели нарушение барионного заряда происходит на древесном уровне, например в виде распада протона $p^+ \to e^+ + \pi^0$. Однако, к сожалению, это имеет место при очень высоких температурах, так что после восстановления термодинамического равновесия сфалеронные процессы стандартной модели вымоют нарушение барионного заряда. Надо отметить, что они не способны нарушать сохранение разности барионного и лептонного зарядов, но она не нарушена и в рамках $SU(5)$ объединения, а потому ненулевая разность не может быть порождена при спонтанном нарушении $SU(5)$ симметрии. Впрочем, данный вариант Великого Объединения не является жизнеспособным и с точки зрения физики элементарных частиц, так как приводит к слишком малому времени жизни протона.

Больший интерес представляют более широкие группы симметрии, в мультиплетах которых естественно появляется место для дополнительных фермионов, например, $SO(10)$. Они хороши также тем, что этими фермионами могут быть массивные майорановские праворукие нейтрино, которые могут быть использованы для объяснения ненулевых масс наблюдаемых нейтрино посредством механизма качелей (see-saw, первого типа). Майорановская масса нарушает и CP четность, и сохранение лептонного заряда (без непосредственного нарушения барионного, а значит нарушает и их разность). Тем самым становится возможным порождение лептонной асимметрии, которая в дальнейшем будет частично переведена в барионную сфалеронами стандартной модели. Это наиболее признанная и разработанная модель – бариосинтез через лептогенезис [13].

В суперсимметричных моделях возможно также нарушение барионного заряда в скалярном секторе (механизм Аффлека-Дайна [14]). Рассматриваются и ещё более экзотические модели – движущиеся доменные стенки, необычная физика в распадах Чёрных Дыр, и так далее...

Кандидаты на роль Тёмной Материи

Другая естественная задача для физики элементарных частиц – предоставить кандидатов на роль частиц Тёмной Материи. Они должны весьма слабо взаимодействовать с частицами стандартной модели, поскольку в противном случае изменился бы ход первичного нуклеосинтеза, а также претерпела бы изменения скорость звуковых волн в первичной плазме, которая проявляется во флуктуациях температуры реликтового фона по направлению на небесной сфере.

Одной из самых популярных версий долгое время были так называемые слабо взаимодействующие массивные частицы – WIMP. Это могут быть, например, легчайшие суперпартнеры (LSP). Если предположить равновесный тепловой механизм их генерации, а также сделать естественные предположения о сечениях аннигиляции, то можно обнаружить, что для объяснения наблюдаемого количества Тёмной Материи их масса должна быть порядка ТэВ. Удивительно, но примерно такие же массы для LSP назывались из соображений стабилизации радиационных поправок к массе бозона Хиггса (проблема иерархии). Это совпадение придало большую значимость подобным моделям, однако результаты работы Большого Адронного Коллайдера (а точнее их отсутствие в плане поиска суперсимметрии) выглядят обескураживающе.

В связи с этим сейчас всё большую популярность приобретает другой класс моделей – аксионы, (псевдо)скалярные частицы очень малой массы с нетепловым механизмом рождения в состоянии конденсата. Такие частицы исторически происходят из попыток решения проблемы сильного CP нарушения в квантовой хромодинамике, а также в огромных количествах могут возникать в низкоэнергетических решениях теории струн. Их существование также пока не подтверждается экспериментально, а многие феноменологически релевантные версии уверенно исключены (теории струн это, конечно, не касается [15]).

Вполне возможно, с точки зрения проблемы тёмных секторов в космологии, следует искать не расширения стандартной модели физики элементарных частиц, а более успешную теорию гравитации (тем более что корректировки в гравитационном секторе могут приводить к изменению космологических ограничений на параметры физики элементарных частиц). Мы ещё вернёмся к этому замечанию в разделе 1.3.

1.2.5 Инфляция

В описанной выше классической космологии есть определённые концептуальные проблемы. В самом деле, рассмотрим замедленно расширяющуюся Вселенную (доминирование Тёмной Энергии наступило в космологических масштабах недавно) $a(t) \propto t^p$, $p < 1$ (степенной закон, хоть и реалистичен, но выбран только для простоты). Фотоны, а также любые другие носители информации, удовлетворяющие обычным представлениям о релятивистской причинности, перемещаются не быстрее, чем по светоподобным геодезическим $a(t)dr = \pm dt$ (радиальное движение в сферических координатах). Поэтому с момента начальной сингулярности до момента времени t любое возмущение проходит путь не длиннее чем

$$d(t) = a(t) \int\limits_0^t \frac{d\tau}{a(\tau)} = \frac{t}{1-p}. \tag{1.75}$$

Как видим, при $t \to 0$ функция $d(t)$ стремится к нулю быстрее, чем линейные размеры содвижущихся областей, определяемые масштабным фактором.

Более того, несложная оценка показывает, что области, разделённые на небесной сфере углом примерно в полтора и более градуса, не имели возможности обменяться информацией от момента начальной сингулярности и до отщепления от вещества (рекомбинация) тех фотонов, которые мы видим сегодня в реликтовом фоне. Вместе с тем, из наблюдений известно, что относительная флуктуация температуры реликтового фона по направлению имеет порядок лишь 10^{-5}. Это порождает проблему неестественности начальных условий.

Есть и некоторые другие трудности. Как мы уже упоминали, пространственная кривизна нашего мира весьма близка к нулю, и это означает, что в прошлом пространственная плоскостность Вселенной выполнялась с совершенно невероятной точностью. Почему? Неясно. Далее, если в очень ранней Вселенной существовала симметрия Великого Объединения, то почему от тех времён не осталось магнитных монополей и/или других топологических дефектов?

Интересно однако, что все эти проблемы напрямую связаны с замедленным характером расширения Вселенной. Если до эпохи классической космологии Вселенная проходила через период ускоренного расширения, то далеко разнесённые области могли в прошлом находиться в каузальном контакте друг с другом, а пространственная кривизна после длительного расширения могла стать пренебрежимо малой (заметим, что плотность экзотической материи, способной вызывать ускоренное расширение Вселенной, убывает заведомо медленнее, чем вклад пространственной кривизны). Более того, если после окончания инфляции перенагрев Вселенной (reheating) ограничился температурами меньшими Великого Объединения, то и проблемы магнитных монополей не возникает.

Длины волн и хаббловский масштаб

Инфляция позволяет получить естественное объяснение происхождения первичных флуктуаций и предсказать их спектр, в хорошем согласии с наблюдениями. Это связано с тем, что для динамики возмущений в расширяющейся вселенной важную роль играет масштаб (четырёхмерной) кривизны, или Хаббловский масштаб $\frac{c}{H}$. Для более коротковолновых возмущений кривизна Вселенной не очень важна, и их поведение близко к обычным волновым процессам. А для более длинных волн пространственные производные в уравнениях движения оказываются малы, и справедливо приближение "отдельных вселенных". Типичное поведение при аккуратном выборе переменных – одна мода почти постоянная, а другая – быстро затухающая.

При этом надо иметь в виду, что длины волн меняются пропорционально масштабному фактору, а поэтому

$$\frac{\lambda}{H^{-1}} = \frac{a}{\left(\frac{\dot{a}}{a}\right)^{-1}} = \dot{a}.$$

Отсюда ясно, что в замедленно расширяющейся вселенной длины волн возмущений становятся всё более короткими по сравнению с Хаббловским масштабом, который в стандартной космологии с обычной материей по порядку величины совпадает с горизонтом видимой части вселенной. Образно говорят, что возмущения заходят под горизонт. При этом предсказание флуктуаций реликтового фона требует произвольного выбора начальных условий в каузально несвязанных областях. Спектр Харрисона-Зельдовича [16,17] получается из естественного, но строго говоря ничем не оправданного, предположения о масштабной инвариантности флуктуаций.

Совсем иное поведение присуще инфляционным моделям. Длины волн растут намного быстрее, чем масштаб кривизны (в квази-деситтеровской стадии первые растут почти экспоненциально, а второй – почти постоянен). Соответственно, первичные флуктуации могли рождаться на очень малых масштабах во время инфляции, и иметь квантовую природу. Их длины волн сильно росли во время инфляции (минимально необходимое инфляционное расширение Вселенной оценивается как в $e^{50\div60}$ раз, или как говорят, пятьдесят - шестьдесят e-фолдингов), так что во время выхода из инфляции и в очень ранней горячей Вселенной они были "за горизонтом" ($\lambda \gg cH^{-1}$), и на них практически не влияла плохо известная физика соответствующих состояний материи. Одна мода осталась почти неизменной, а другая затухла, обеспечив когерентность колебаний. В замедленной фазе расширения длины волн последовательно опять входят под горизонт и вызывают звуковые волны в первичной плазме, наблюдаемые посредством флуктуаций в реликтовом фоне.

Скалярный инфлатон

Возникает вопрос, можно ли предложить работающую модель инфляции, ведь, как мы уже видели ранее, для ускоренного расширения необходимо экзотическое состояние вещества с отрицательным давлением. Оказывается, достаточно использовать каноническое скалярное поле со стандартным действием

$$S = \int d^4x\sqrt{-g}\left(\frac{1}{2}(\nabla_\mu\varphi)(\nabla^\mu\varphi) - V(\varphi)\right). \qquad (1.76)$$

Для него нетрудно вычислить тензор энергии-импульса и убедиться, что плотность энергии и давление для пространственно однородного скалярного поля равны

$$\rho = \frac{1}{2}\dot\varphi^2 + V(\varphi), \qquad (1.77)$$

$$p = \frac{1}{2}\dot\varphi^2 - V(\varphi). \qquad (1.78)$$

Очевидно, требуется режим, в котором потенциальная энергия доминирует над кинетической. Ясно, что это будет режим медленного качения скалярного поля по достаточно плоскому плато потенциальной энергии. Более того, потенциал должен быть даже не столько именно плоским, сколько просто не крутым, поскольку уравнение движения

$$\ddot\varphi + 3H\dot\varphi + V' = 0, \qquad (1.79)$$

где $V' \equiv \frac{dV}{d\varphi}$, содержит слагаемое $3H\dot\varphi$, играющее роль вязкого трения (хаббловское трение).

По аналогии с движением в вязкой среде можно ожидать, что существует режим медленного качения, при котором

$$\dot\varphi \approx -\frac{V'}{3H}. \qquad (1.80)$$

Мы не будем вдаваться в подробности строгого установления условий, при которых существует этот режим. Обычно бывает достаточно контролировать малость двух параметров медленного качения:

$$\epsilon \equiv \frac{1}{16\pi G}\left(\frac{V'}{V}\right)^2 \ll 1, \qquad (1.81)$$

$$\eta \equiv \frac{1}{8\pi G}\frac{V''}{V} \ll 1. \qquad (1.82)$$

При рассмотрении негауссовости в первичных флуктуациях могут потребоваться аналогичные параметры со следующими производными потенциала.

Смысл первых двух параметров в том, что медленно меняется постоянная Хаббла

$$\frac{\dot{H}}{H^2} \approx -\epsilon,$$

а также можно пренебречь второй производной инфлатона по времени в его уравнении движения

$$\frac{\ddot{\varphi}}{H\dot{\varphi}} \approx \epsilon - \eta.$$

1.2.6 Основы космологической теории возмущений

Для теоретического изучения малых отклонений от однородности часто бывает удобно перейти от физического времени t к конформному τ посредством замены переменных $a^2(\tau)d\tau^2 = dt^2$. При этом метрика пространственно плоской фридмановской вселенной оказывается (четырёхмерно) конформно плоской $g_{\mu\nu} = a^2(\tau)\eta_{\mu\nu}$. Во избежание путаницы обычно договариваются производную по конформному времени обозначать штрихом вместо точки. Производные при этом пересчитываются очевидным образом, и в частности можно ввести аналог постоянной Хаббла $\mathcal{H} = \frac{a'}{a} = aH$.

Запишем линеаризованную флуктуацию метрики в виде

$$\begin{aligned} ds^2 = a^2(\tau)\big(&(1+2\phi)d\tau^2 + 2(\partial_i f + v_i)dx^i d\tau \\ &+((-1+2\psi)\delta_{ij} + 2\partial^2_{ij}s + \partial_i c_j + \partial_j c_i + h_{ij})dx^i dx^j\big)\,, \end{aligned} \tag{1.83}$$

где по определению принято $\partial_i v_i = \partial_i c_i = \partial_j h_{ij} = h_{ii} = 0$ – бездивергентность и бесследовость всех величин, для которых это имеет смысл. Очевидно, что такое разбиение на скаляры ϕ, ψ, f, s, векторы v, c и (симметричный) тензор h связано просто с неприводимыми представлениями группы трёхмерных вращений, так что в линейном приближении можно рассматривать три сектора независимо.

Например, если дано векторное уравнение, оно может содержать как бездивергентные вектора, так и градиенты скаляров. Возьмем от него дивергенцию. Тогда получается, что лапласиан от некоторого линейного выражения, составленного из скалярных величин, обращается в нуль. В теории возмущений мы это интерпретируем как уравнение, требующее обращения в нуль самого скалярного выражения. А бездивергентная часть исходного уравнения обращается в нуль самостоятельно, давая независимое уравнение для векторных величин. На языке квадратичного действия это расщепление проявляется в невозможности записать инвариантное квадратичное взаимодействие скаляра и бездивергентного вектора, не исчезающее при интегрировании по частям.

Калибровочная свобода

На данном этапе линейные флуктуации геометрии параметризованы десятью величинами - по числу компонент метрики. Однако необходимо ещё помнить о свободе координатных преобразований, ведь бесконечно малое координатное преобразование можно описать как изменение переменных, описывающих флуктуации метрики, без изменения самой фоновой метрики. Представив такое преобразование как $x^\mu \to x^\mu + \xi^\mu(x)$, легко видеть, что метрика в линейном приближении претерпевает изменение

$$\delta g_{\mu\nu} = -g_{\alpha\mu}\partial_\nu \xi^\alpha - g_{\alpha\nu}\partial_\mu \xi^\alpha - \xi^\alpha \partial_\alpha g_{\mu\nu}.$$

Тензорные флуктуации h при этом не меняются, отражая объективный характер гравитационных волн. Далее, разделив пространственные компоненты ξ на градиентную и бездивергентную части, а также отдельно выделив его временную часть, легко установить, что половина скалярных и половина векторных возмущений оказываются чистыми калибровками, а инвариантны следующие потенциалы:

$$\Phi = \phi - \frac{1}{a}\left(a(f - s')\right)',$$

$$\Psi = \psi + \mathcal{H}(f - s'),$$

$$V_i = v_i - c_i'.$$

Как всегда, возможны различные выборы калибровки. Для численных методов удобной бывает синхронная калибровка (все флуктуации – в пространственных компонентах метрики), которая очевидно получается при $\phi = f = v = 0$. Для теоретических исследований весьма хороша ньютоновская калибровка $f = s = c = 0$, оставшиеся потенциалы в которой, к тому же, численно совпадают с калибровочно-инвариантными переменными.

Если же рассматривается вселенная, наполненная веществом, то возможны и другие естественные выборы калибровки, например калибровка, в которой обращается в нуль флуктуация плотности энергии $\delta\rho = 0$. Можно считать, что при этом плотность энергии вещества в расширяющейся Вселенной использована для задания переменной времени.

Идеальная жидкость

Для многих, хотя и отнюдь не для всех, задач теоретической космологии достаточно приближения идеальной жидкости для вещества, заполняющего вселенную:

$$T^{\mu\nu} = -pg^{\mu\nu} + (p + \rho)u^\mu u^\nu \tag{1.84}$$

где u^μ – 4-скорость жидкости. Напомним, что в системе покоя $u^\mu = \frac{\delta_0^\mu}{a(\tau)}$, и это будет фоновым значением 4-скорости в космологической системе отсчёта.

Тензоры энергии-импульса скалярного поля и пылевидной материи всегда имеют такой вид, а для применимости этого приближения к электромагнитной плазме необходимо, чтобы диффузионная длина фотона была много меньше характерных линейных размеров изучаемых флуктуаций.

Получаем линеаризованные компоненты тензора энергии-импульса:

$$T_0^0 = \rho + \delta\rho,$$

$$T^0_i = \frac{1}{a}(\rho + p)u_i,$$
$$T^i_j = -(p + \delta p)\delta^i_j,$$

причём, в рамках раздельного анализа скалярных и векторных возмущений, необходимо разложить пространственные компоненты скорости на градиентную и бездивергентную части

$$u_i = \partial_i u + \tilde{u}_i,$$

где $\partial_i \tilde{u}_i = 0$, что соответствует разложению на потенциальное и вихревое течения.

Линеаризованные гидродинамические уравнения

Чтобы записать уравнения теории возмущений [18], нужно найти тензор Эйнштейна возмущённой метрики в линейном приближении. Это задача решается вполне прямолинейно, но весьма громоздко. Мы ограничимся выписыванием ответов в ньютоновской калибровке.

В тензоре энергии-импульса идеальной жидкости отсутствуют тензорные флуктуации, поэтому в линейном приближении гравитационные волны не взаимодействуют с идеальной жидкостью:

$$h''_{ij} + 2\mathcal{H}h'_{ij} - \Delta h_{ij} = 0.$$

В коротковолновом (высокочастотном) режиме можно пренебречь вторым слагаемым, и мы имеем волновое уравнение. В длинноволновом приближении малым оказывается последнее слагаемое, и уравнение описывает две моды: одну быстро затухающую, и одну почти постоянную.

Для векторных флуктуаций смешанные $(0i)$ компоненты уравнений Эйнштейна связывают скорость вихревого движения с векторными возмущениями метрики, в то время как пространственные компоненты уравнений показывают, что последние убывают. В терминах физических скоростей вихревых потоков материи получается затухание во всех интересных для космологии режимах, кроме точного доминирования излучения, когда физические вихревые скорости остаются постоянными. Это весьма общее свойство векторных возмущений, тесно связанное с сохранением момента импульса, и оно делает векторные возмущения нерелевантными для подавляющего большинства интересных задач.

Далее, идеальная жидкость отличается также тем свойством, что давление её изотропно. В скалярном секторе это проявляется в том, что внедиагональные компоненты пространственных уравнений Эйнштейна принимают вид $\partial^2_{ij}(\phi - \psi) = 0$, $i \neq j$, из которого в рамках философии космологических возмущений следует

$$\phi = \psi,$$

что интересно также сравнить с линейным порядком разложения решения Шварцшильда. В результате в скалярном секторе остаётся один независимый потенциал (вполне аналогичный ньютоновскому в нерелятивистской теории).

Смешанные компоненты уравнений Эйнштейна связывают его с потенциалом скоростей течения материи

$$\phi' + \mathcal{H}\phi = 4\pi G a(\rho + p)u,$$

а временную и пространственные компоненты уравнений следует собрать воедино, подставив $\delta p = c_s^2 \delta\rho$, где c_s – скорость звука. В результате получается уравнение

$$\phi'' + 3\mathcal{H}(1 + c_s^2)\phi' + \left(2\mathcal{H}' + (1 + 3c_s^2)\mathcal{H}^2\right)\phi - c_s^2 \Delta\phi = 0 \tag{1.85}$$

для распространения скалярных возмущений (флуктуаций плотности и температуры), или попросту звуковых волн в первичной плазме.

Фоновый режим расширения Вселенной проявляется здесь в явной зависимости $\mathcal{H}$ от времени, и этого уравнения оказывается вполне достаточно, чтобы убедиться, что на малых масштабах контраст плотности эффективно нарастает при доминировании материи ($c_s = 0$), но не при доминировании излучения ($c_s^2 = \frac{1}{3}$), в результате чего Тёмная Материя играет решающую роль в том, чтобы малые первичные флуктуации могли привести к образованию крупномасштабной структуры Вселенной. Мы не будем здесь заниматься этим вопросом.

Заметим однако, что в случае $c_s^2 = w = const$ уравнения Фридмана приводят к обращению в нуль коэффициента при ϕ без производных по времени. Соответственно, опять имеем одну затухающую и одну почти постоянную моду. Однако в общем случае это не так. Но, оказывается, существует другая переменная

$$\phi - \frac{\delta\rho}{3(\rho + p)} \equiv \xi \approx \mathcal{R} \equiv \phi + \frac{\mathcal{H}u}{a} = \phi + Hu,$$

которая обладает этим свойством в длинноволновом приближении (оно же использовано для приближенного равенства между ξ и $\mathcal{R}$). В частности, можно установить, что при переходе от доминирования излучения к доминированию материи ньютоновский потенциал длинноволновых мод умножается на $\frac{9}{10}$.

Отметим, что у новой переменной есть естественный геометрический смысл скалярной кривизны трехмерных поверхностей постоянной плотности энергии (или постоянного времени в калибровке $\delta\rho = 0$).

Флуктуации в реликтовом фоне

Надо заметить, что в случае реалистичной многокомпонентной жидкости следовало бы отдельно рассмотреть адиабатическую и энтропийные моды. Однако наблюдения показывают, что энтропийные моды не присутствуют в значимых количествах, и мы подробнее рассмотрим именно адиабатическую моду.

Поскольку доминирующий вклад в энтропию дают ульрарелятивистские степени свободы, сохранение энтропии (в расчете на один барион) можно записать как $\frac{T^3}{n_b} = const$, откуда нетрудно вывести, что адиабатическая мода характеризуется соотношениями

$$\frac{3}{4}\frac{\delta\rho_\nu}{\rho_\nu} = \frac{3}{4}\frac{\delta\rho_\gamma}{\rho_\gamma} = \frac{\delta\rho_b}{\rho_b} = \frac{\delta\rho_{CDM}}{\rho_{CDM}}$$

для нейтрино, фотонов, "барионов" и холодной Темной Материи.

В таком случае для ньютоновского потенциала, создаваемого заданной длинноволновой (пренебрежение пространственными производными) флуктуацией плотности, получаем, учитывая, что к моменту рекомбинации плотность энергии доминируется нерелятивистскими компонентами,

$$\phi = -\frac{4\pi G}{3H^2}\delta\rho = -\frac{\delta\rho}{2\rho} = -\frac{3\delta\rho_\gamma}{8\rho_\gamma} = -\frac{3}{2}\left(\frac{\delta T}{T}\right)_0,$$

и в результате (эффект Сакса-Вольфе [19]), с учетом гравитационного красного смещения, из области с флуктуацией температуры $\left(\frac{\delta T}{T}\right)_0$ приходят фотоны с

$$\frac{\delta T}{T} = \left(\frac{\delta T}{T}\right)_0 + \phi = -\frac{1}{2}\left(\frac{\delta T}{T}\right)_0.$$

Интересно заметить, что из более горячих областей приходят более холодные фотоны, и наоборот.

Отметим также, что выше описан лидирующий эффект. Для точных предсказаний нужно ещё учесть допплеровское смещение от движений первичной плазмы, а также интегрированный эффект Сакса-Вольфе – влияние переменности ньютоновского потенциала по ходу движения фотонов реликтового фона. Последний важен тогда, когда потенциал действительно меняется – переход к доминированию материи (ранний ISW) и переход к доминированию Тёмной Энергии (поздний ISW). Для точного анализа применяются численные методы [20].

Квантовое рождение флуктуаций

Как уже отмечалось выше, инфляционная теория позволяет получить первичные космологические неоднородности из квантовых флуктуаций во время инфляции. Опишем кратко, как это делается. Следует рассмотреть квадратичное действие для флуктуаций вокруг инфляционного решения.

В тензорном секторе имеем

$$\frac{1}{64\pi G}\int d^4xa^2(\tau)\left({h'_{ij}}^2-\left(\partial_k h'_{ij}\right)^2\right).$$

Переходя к двум канонически нормированным переменным для двух независимых поляризаций

$$h_{ij}=\frac{\sqrt{32\pi G}}{a}(v_1e_{ij}^{(1)}+v_2e_{ij}^{(2)}),$$

имеем для них действие

$$\frac{1}{2}\int\left(v_i'^2-(\partial_jv_i)^2+v_i^2\frac{a''}{a}\right)$$

с переменной эффективной массой, но в пространстве Минковского.

Приближённо считая пространство де-ситтеровским $a(\tau)=-\frac{1}{H\tau}$, где переменная конформного времени $-\infty<\tau<0$, получаем для растущей (или в терминах h асимптотически постоянной) монохроматической моды с волновым числом k зависимость от времени $v_k(\tau)=e^{-ik\tau}\left(1-\frac{i}{k\tau}\right)$. Проводя стандартным образом каноническое квантование, имеем для вакуумных флуктуаций при $\tau\to 0$ (бесконечно удалённое будущее): $\left\langle 0|v^2|0\right\rangle\to\int\frac{dk}{k}\left(\frac{Ha}{2\pi}\right)^2$ со спектральной плотностью $\delta^2=\left(\frac{Ha}{2\pi}\right)^2$. В терминах исходной переменной (и складывая вклады от двух поляризаций) это даёт

$$\delta_h^2=\frac{16H^2G}{\pi}.$$

Таким образом, мы получили масштабно-инвариантный спектр, что конечно же является проявлением симметрии мира де Ситтера. В реальности надо учесть, что константа Хаббла медленно меняется со временем, и для данной длины волны хорошим приближением будет подставить то значение константы Хаббла, при котором она выходила за "хаббловский горизонт" во время инфляции.

Более длинные волны рождались раньше, $d\ln k=Hdt$, и для них было большее значение константы Хаббла, что даёт наклон тензорного спектра в красную сторону, величина которого определяется скоростью изменения константы Хаббла:

$$n_T \equiv \frac{d \ln \delta_h^2}{d \ln k} = -2\epsilon.$$

* * *

В скалярном секторе квадратичное действие получается весьма громоздким, но можно убедиться, что канонической переменной с действием аналогичным случаю гравитационных волн является переменная Муханова-Сасаки

$$\mathfrak{u} = a\left(\delta\varphi + \frac{\phi\varphi'}{\mathcal{H}}\right),$$

для которой переменная эффективная масса $\frac{\mathfrak{z}''}{\mathfrak{z}}$, где $\mathfrak{z} \equiv a\frac{\varphi'}{\mathcal{H}} = a\frac{\dot\varphi}{H}$, в приближении медленного качения практически совпадает с таковой для гравитационных волн $\frac{a''}{a}$. Поэтому результат квантования получается тем же. И для естественной переменной скалярных флуктуаций (в пылевидную эпоху она равна $\frac{5}{3}\phi$)

$$\mathcal{R} = \phi + \frac{\mathcal{H}}{\varphi'}\delta\varphi$$

получаем (из второго слагаемого, которое при медленном качении доминирует):

$$\delta_{\mathcal{R}}^2 = \left(\frac{H^2}{2\pi\dot\varphi}\right)^2.$$

Наклон этого спектра зависит уже и от изменения скорости качения скалярного поля, и может быть вычислен как

$$n_s - 1 \equiv \frac{d \ln \delta_{\mathcal{R}}^2}{d \ln k} = 2\eta - 6\epsilon.$$

Интересно отметить, что для отношения амплитуд тензорных и скалярных возмущений получаем

$$r \equiv \frac{\delta_h^2}{\delta_{\mathcal{R}}^2} = 16\epsilon,$$

и поэтому для любых моделей инфляции с одним медленно катящимся скалярным полем имеем условие согласованности

$$r = -8n_T.$$

1.2.7 О наблюдательных данных

Существующие в космологии наблюдательные данные имеют различную природу. Разумеется, важнейшую роль играют классические методы наблюдательной астрономии. Отличительной особенностью космологичесих задач является необходимость изучения весьма удалённых объектов. И уже такая элементарная величина как расстояние до источника света представляет серьёзные трудности для измерения, даже без учета концептуальных сложностей, связанных с общерелятивистскими представлениями о пространстве-времени.

Как известно, для ближайших звезд удаётся использовать кинематические методы определения расстояний. Но при переходе ко внегалактической астрономии на первое место выходит метод стандартных свечей – если по каким-то причинам удаётся определить истинную светимость объекта, то измерение его наблюдаемой яркости позволяет определить расстояние. Мы опускаем здесь подробности, связанные с учетом поглощения света межзвёздной средой (соотношение между затуханием и покраснением).

На сравнительно небольших по внегалактическим меркам расстояниях используются периодически переменные звезды, в первую очередь – цефеиды. В них происходят радиальные пульсации атмосферы, связанные с переходами между однократной и двухкратной ионизацией гелия в некотором слое, которые меняют непрозрачность вещества и приводят к колебаниям радиуса, а с ним и блеска, причём период пульсаций оказывается связан простой зависимостью с максимумом светимости. Калибровав эту зависимость по цефеидам нашей галактики, можно использовать её для определения расстояний до соседних галактик (мы опускаем тонкости, связанные с принадлежностью цефеид к разным типам звёздного населения).

На этом этапе можно провести непосредственное измерение постоянной Хаббла, в предположении, что локальный регион адекватно отражает динамику всей видимой части Вселенной в целом: мы не живём, например, в центре большого войда (void) – области пониженной плотности. Мы намеренно упоминаем здесь экзотическую возможность того, что наша локальная группа чем то выделена, поскольку существуют определенные нестыковки между локальными измерениями (например $H_0 = (73.8 \pm 2.4)\frac{\mathrm{km/s}}{\mathrm{Mpc}}$ из работы [21] или более новое значение $H_0 = (73.24 \pm 1.74)\frac{\mathrm{km/s}}{\mathrm{Mpc}}$ [22]) и наилучшим определением константы Хаббла по флуктуациям реликтового фона и изучению крупномасштабной структуры Вселенной [23], $H_0 = (67.6 \pm 0.6)\frac{\mathrm{km/s}}{\mathrm{Mpc}}$.

Если заниматься более детальной космографией, и интересоваться историей расширения Вселенной, необходимо изучать ещё более далекие объекты, и здесь на первый план выходят сверхновые типа Ia. Взрыв такой сверхновой происходит в двойной системе при аккреции вещества со звезды-гиганта на белый карлик, когда оказывается превышен предел Чандрасекара. Поэтому светимость в максимуме оказывается очень точно определённой, а небольшой остаточный разброс (разная металличность и т.д.) скоррелирован с темпом спадания яркости.

Мы опустили многие другие важные методы, такие как определение расстояния до эллиптической галактики с помощью флуктуаций поверхностной яркости (зависящих от полного числа звёзд, а соответственно и полной светимости). Однако понятно, что шкала измерения расстояний строится по методу лестницы, когда последующая перекладина опирается на предыдущую.

Именно изучение яркости сверхновых типа Ia на космологических расстояниях позволило впервые обнаружить ускоренное расширение Вселенной. Когда красное смещение становится близким к единице (или даже больше), зависимость яркости стандартной свечи от красного смещения определяется уже не только текущим значением константы Хаббла, но и тем, какой она была в прошлом.

Заметим, что оптические методы позволяют определять не только яркость, но, разумеется, и спектр объектов, что даёт возможность не только установить тип объекта, но и детально изучить химический состав, вплоть до соотношения разных изотопов. Это обстоятельство позволяет, например, сравнивать предсказания теории первичного нуклеосинтеза с наблюдениями, подыскивая окружения с наиболее первозданным составом.

Но также изучение изотопного состава даёт возможность установить возраст звёзд (при некоторых достаточно обоснованных предположениях о первоначальном изотопном составе – тяжёлые элементы рождаются только в результате термоядерных взрывов сверновых), и для наиболее старых звёзд получается возраст порядка тринадцати-четырнадцати миллиардов лет. Для звёздных скоплений можно также определять возраст по типу звёзд, которые уже сошли с главной последовательности, комбинируя эту информацию с теоретическими представлениями об эволюции звёзд.

Эти измерения полезно сравнить с возрастом Вселенной в рамках принятой космологической модели. Вычислить возраст Вселенной нетрудно, заметив, что $dt = \frac{da}{\dot{a}} = -\frac{a_0 dz}{Ha(1+z^2)}$, и учитывая зависимость (1.73) константы Хаббла от красного смещения:

$$t_0 = \frac{1}{H_0} \int\limits_0^\infty \frac{dz}{(1+z)\sqrt{\sum\limits_i \Omega_i (1+z)^{n_i}}}. \tag{1.86}$$

Если считать, что наша Вселенная заполнена пылевидной материей, то получится около семи миллиардов лет, что явно противоречит возрасту наблюдаемых объектов. Если же учесть вклад Тёмной Энергии в виде космологической постоянной, то получается $13.8 \cdot 10^9$ лет, что с учётом погрешностей хорошо согласуется с данными.

* * *

XXI первый век ознаменовался тем, что в космологии появился принципиально новый источник информации о реальной Вселенной – изучение флуктуаций реликтового фона.

Открытый в 60-е годы Пензиасом и Вильсоном, он не мог дать детальной информации при наблюдениях с земли, поскольку максимум распределения приходится на длину волны порядка 0.3cm (частота 100GHz), на которой атмосфера непрозрачна. Наблюдения Пензиаса и Вильсона велись на 7.5cm, то есть в рэлей-джинсовской асимптотике $\rho(\nu) = \frac{8\pi h\nu^3 d\nu}{e^{\frac{h\nu}{kT}}-1} \approx 8\pi\nu^2 kT$, которая не позволяет даже уверенно установить чернотельный характер излучения (в смысле правильной общей нормировки для данной температуры). Отметим, правда, что было известно возбуждение колебательных уровней молекул циана в межзвёздной среде с населённостями, соответствующими нескольким градусам по Кельвину.

Затем было несколько экспериментов на воздушных шарах, и наконец в 90-е годы с помощью спутника COBE (COsmic Background Explorer) был доказан чернотельный характер спектра реликтового фона, а также обнаружены относительные флуктуации температуры порядка 10^{-5} (за вычетом большего дипольного вклада). В нулевые годы работал спутник WMAP, и он стал первым экспериментом, позволившим точно измерить флуктуации температуры до мультипольных моментов порядка тысячи (угловое разрешение). С лучшим разрешением, но зато с неполным покрытием неба, измерения делались на наземных телескопах ACT (Атакама) и SPT (Антарктида). А несколько лет назад работал спутник Planck, с лучшими чувствительностью и угловым разрешением, чем WMAP.

Для анализа данных флуктуации температуры раскладываются по сферическим гармоникам $\delta T = \sum a_{l,m} Y_{l,m}$, причём из условия вещественности следует $a^*_{l,m} = a_{l,-m}$, а из статистической изотропии –

$$< a_{l_1,m_1} a^*_{l_2.m_2} >= C_l \cdot \delta_{l_1,l_2} \delta_{m_1,m_2}.$$

Значения коэффициентов, а точнее $\frac{l(l+1)}{2\pi} C_l$, используются для сравнения теории с наблюдениями. При этом каждому мультипольному моменту l отвечает $2l+1$ независимых измерений, поэтому на малых мультипольных моментах велика неопределённость (cosmic variance). На больших же значениях мультипольного момента (малых угловых масштабах), данных вполне достаточно, и их объединяют в более крупные группы (bins).

Мультипольные моменты первой полусотни соответствуют длинам волн, которые на момент рекомбинации всё ещё превышали хаббловский масштаб. Они не вызывали звуковых колебаний в первичной плазме, и поэтому представляют собой первичный спектр флуктуаций от инфляции, почти масштабно инвариантный, но с небольшим красным уклоном. Далее в спектре реликтового фона следуют акустические пики, которые в большей мере отражают уже состав первичной плазмы, нежели начальные условия. После первой тысячи мультипольных моментов спектр первичных флуктуаций реликтового фона начинает быстро убывать в силу нескольких причин - диффузионное размытие фотонами, размытие свободным распространением нейтрино (зависит от числа их типов!), а также конечность времени, в течение которого происходила рекомбинация (грубо говоря, попытка сделать снимок быстрого процесса на объективе с большой выдержкой).

Помимо температуры в реликтовом фоне можно также измерять поляризацию. Поляризация лежит, разумеется, в касательном пространстве к небесной сфере (перпендикулярно направлению движения фотонов) и может быть задана симметричной матрицей, составленной из стандартных параметров Стокса $P_{ab} = \frac{1}{2}\begin{pmatrix} 1+Q & U-iV \\ U+iV & 1-Q \end{pmatrix}$. При этом первичная поляризация, как происходящая из томсоновского рассеяния, имеет чисто линейный характер, то есть $V = 0$. Для анализа космологических данных удобно вместо двух привычных для оптики параметров Стокса (U и Q) ввести два потенциала посредством

$$\mathcal{P}_{ab} = \frac{1}{2}(\nabla_a \nabla_b + \nabla_b \nabla_a - g_{ab}\Delta)\mathcal{P}_E - \frac{1}{2}(\epsilon_a^c \nabla_b \nabla_c + \epsilon_b^c \nabla_a \nabla_c)\mathcal{P}_B,$$

где $\mathcal{P}_{ab} \equiv P_{ab} - \frac{1}{2}g_{ab}$ – бесследовая часть (вычли полную интенсивность) матрицы поляризации, а ϵ - абсолютно антисимметричный тензор.

Очевидно, первый параметр – скалярный (E-мода), а второй – псевдоскалярный (B-мода). В силу сохранения чётности в электродинамике, в первичном фоне могут присутствовать корреляции ET (E-моды поляризации с температурой), но не BT или EB. Причём B-мода поляризации может иметься в первичном сигнале только за счёт первичных гравитационных волн (и на сравнительно больших мультипольных моментах – из-за нелинейных эффектов гравитационного линзирования). Современные методы позволяют наблюдать все типы корреляций, но B-мода зарегистрирована пока не была, кроме эффектов линзирования.

При сравнении теории с наблюдениями используется шестипараметрическая космологическая модель: состав современной Вселенной (грубо говоря, плотность барионов, плотность Тёмной Материи, космологическая постоянная), амплитуда и наклон первичных флуктуаций от инфляции, а также оптическая толщина слоя повторной ионизации. При этом фитирование шестью параметрами позволяет хорошо описать всю наблюдаемую кривую зависимости C_l от l, включая начальный почти плоский участок, три акустических пика и кривую спадания при $l > 1000$, а также определённые наборы внешних данных.

При этом в модель заложен целый ряд предположений – нулевая пространственная кривизна, отсутствие первичных гравитационных волн, нулевая масса нейтрино, космологическая постоянная в роли Тёмной Энергии... После этого рассматриваются одно- и/или двухпараметрические расширения, которые позволяют накладывать ограничения на такие параметры как первичное отношение тензоров к скалярам, сумма масс трёх поколений нейтрино, отличие уравнения состояния Тёмной Энергии от космологической постоянной, вклад пространственной кривизны и т.д..

Получающаяся из такого анализа плотность барионов прекрасно согласуется с результатами анализа процессов первичного нуклеосинтеза, а также с ещё одним очень важным источником данных – изучением крупномасштабной структуры Вселенной на основе составления и статистического анализа галактических каталогов. Дело в том, что все структуры развиваются из тех же флуктуаций плотности, что и наблюдаемые в реликтовом фоне. В частности, существует акустический пик в двухточечной корреляционной функции распределения галактик (барионные акустические осцилляции).

1.3 Проблемы стандартной ΛCDM модели и модифицированная гравитация как средство их решения

В целом согласие теории с наблюдениями очень хорошее. Ограничения на первичный спектр, в том числе – необнаружение первичных гравитационных волн, позволяют исключить ряд моделей инфляции. А акустическая часть спектра хорошо соответствует наблюдаемой Вселенной.

Существуют небольшие проблемы на малых мультипольных моментах [24]. Флуктуации в квадруполе и октуполе несколько подавлены (что, разумеется, сложнее объяснять, чем усиление, которое может быть вызвано независимым, нескоррелированным источником, скажем, в поздней Вселенной). Существует необъяснимая корелляция между разными малыми мультипольными моментами. Однако значимость этих аномалий отнюдь не очевидна в силу малого количества независимых данных (cosmic variance). Кроме того, есть подозрительная корреляция с плоскостью эклиптики.

Более серьёзная проблема приходит при сравнении константы Хаббла, получаемой из реликтового фона [23], и особенно через барионные акустические осцилляции, с локальным измерением методом стандартных свечей [21]. Расхождение около двух-трёх стандартных отклонений может быть следствием нашего экзотического положения (в центре большого войда), или результатом отличия истинной теории гравитации от эйнштейновской, или же может указывать на присутствие неучтённых систематических погрешностей в методах измерения расстояний в космологии.

Можно также отметить, что есть определённые указания на более сильное гравитационное линзирование, чем следует из общей терии относительности [23].

Следует также упомянуть и о некоторых расхождениях в определении космологических параметров спутниками Planck и WMAP. Они не столь велики, чтобы ставить под сомнение всю картину, но достаточно серьёзны для точного анализа. Нет единого мнения о происхождении этих различий. Возможно, у эксперимента WMAP были трудности с единой нормировкой двух частей спектра – основной части, измеренной самим спутником, и хвостом затухания, который учитывался по данным наземных телескопов (ACT, SPT) в силу недостаточного углового разрешения аппаратуры спутника [25]. В то же время у эксперимента Planck есть свои сложности [26, 27]. Долгое время не удавалось справиться с утечкой температурного сигнала в карты поляризации, было серьёзное загрязнение на линии $217GHz$ от электроники джоуль-томсоновского холодильника и т.д..

Помимо огромного прогресса, уже достигнутого в космологических наблюдениях, существуют многообещающие планы дальнейших исследований. Это и планируемые следующие микроволновые миссии для регистрации B-моды поляризации (CORE [28]), и спутники для изучения крупномасштабной структуры (EUCLID [29]). Уже сейчас активно исследуются леса Ly_α в спектрах поглощения квазаров как источник космологической информации, дополнительной к реликтовому фону [30]. Они позволяют накладывать ограничения на модели Тёмной Материи со сверхлёгкими аксионами [31]. А в перспективе считается, что этот анализ будет чувствителен не только к сумме масс трёх поколений нейтрино, но и к их иерархии [32].

Новая важная информация может быть получена при измерении таких величин, как негауссовость первичных флуктуаций [33], а также спектральные искажения реликтового фона [34].

Планируются инфракрасные миссии [35] для изучения эпохи повторной ионизации (образование первых звёзд), а также радионаблюдения [36], в том числе связанные с линией 21cm сверхтонкой структуры водорода [37]. Абсолютно новое окно возможностей открывает недавняя регистрация гравитационных волн от слияния чёрных дыр [38–41] и нейтронных звёзд [42].

Последовательное улучшение наблюдательных методов предъявляет серьёзные требования к точности теоретического описания. Космология вступает в эпоху точности измерений порядка процента. В соответствии с этим активно развиваются методы феноменологического описания. В первую очередь актуально аккуратное описание нелинейных фаз развития неоднородностей, поскольку эти трудности имеют совершенно объективный характер, не связанный ни с какими умозрительными кострукциями.

В этом ключе активно развиваются как аналитические методы эффективной теории (в теоретикополевом смысле) образования структур [43], так и численные методы, которые в последние годы имеют перед собой задачу моделирования расширения Вселенной с учетом неоднородностей в рамках полноценной общей теории относительности [44].

* * *

Описанная выше картина показывает почти идеальную ситауцию в науке. Есть сравнительно простая (в концептуальных своих основах) теоретическая модель. Существуют хорошо развитые методы наблюдений. И между теоретическими предсказаниями и наблюдениями имеется прекрасное согласие. При этом однако присутствуют некоторые "мелочи," которые могут указывать на необходимость дальнейшего развития. Более того, есть все основания полагать, что в наблюдательных методах и результатах будет достинут ещё больший прогресс в самое ближайшее время, включая и разработку совершенно новых подходов. А теоретики в то же время успешно развивают свои вычислительные (аналитические и численные) методы, чтобы соответствовать всё возрастающей точности работы экспериментаторов.

Тем не менее, многие физики не удовлетворены сложившейся ситуацией, и ищут радикально новых путей в описании действительности, например с помощью внесения модификаций в теорию гравитации. Почему?

Если ответить на этот вопрос кратко, и максимально удалённо от чисто теоретических размышлений, то цена описанного выше согласия – это отнесение 95% плотности энергии во Вселенной к совершенно неведомым формам, Тёмной Материи и Тёмной Энергии.

В случае с Тёмной Материей не только не понятна её природа, но есть также и явный кризис на малых масштабах [45]. Моделирование в рамках парадигмы холодной Тёмной Материи приводит к слишком большому числу спутников у гигантских галактик типа Млечного Пути, к профилям с сингулярным распределением плотности материи в центре галактики вместо наблюдаемого гладкого ядра (cusp vs. core) и некоторым другим проблемам.

Возможно, что здесь плохо учтены особенности "барионной" физики [46]. Возможно, требуется замена холодной Тёмной Материи на тёплую, или даже нечёткую (fuzzy; сверхлёгкие аксионы [15]). Но с другой стороны, динамика внутри галактик удивительно хорошо описывается феноменологической формулой типа MOND [47]. Это требует объяснения, либо с точки зрения парадигмы CDM, либо в рамках какой-либо теории модифицированной гравитации.

Другая непонятная сущность (70% в балансе энергии!) – Тёмная Энергия, вызывающая ускоренное расширение Вселенной. С одной стороны, тут прекрасно работает простейшая гипотеза – космологическая постоянная. Однако, с точки зрения квантовой теории поля – проблема скорее в том, почему эта постоянная столь мала, на 123 порядка величины меньше планковской плотности, или как минимум на 60 порядков меньше плотности, отвечающей восстановлению суперсимметрии, которая способна сократить вклады бозонов и фермионов в энергию вакуума. Эта проблема печально известна своим противодействием любым попыткам своего разрешения. Так, например, в унимодулярной гравитации [48] космологическая постоянная, хоть напрямую и не проявляется, но возвращается обратно в виде константы интегрирования [49]. Вайнбергу принадлежит весьма сильная "no go" теорема в этой области [50]. Возможно, требуется разработка нелокальных теорий гравитации.

С недавних пор существует концепция деградитации [51], предполагающая, что в теории эффективно присутствует некий фильтр, отключающий длинноволновые моды от взаимодействия с гравитацией. Одним из механизмов могла бы быть гравитация на бранах. Однако в случае коразмерности 1 демпфирование гравитационного взаимодействия недостаточно сильное [52], а старшие коразмерности, по видимому, неразрывно связаны с духовой модой [53] (есть и другое мнение [54], а также возможность каскадной модели [55]). Была надежда на массивную гравитацию, и связанное с ней юкавское ослабление. Однако оказалось, что хотя в ней и возможны режимы, "игнорирующие" затравочную космологическую постояную, но естественной деградитации тем не менее не возникает, не говоря уже о других проблемах этих теорий.

Из более умозрительных вещей, не ясна природа инфлатона. Физическая реальность, в отличие от фундаментальных теорий, бедна на скалярные поля; возможно – не без причины (радиационные поправки к массе скаляра). Но с другой стороны, минималистические модели, встраивающие инфлатон в Стандартную Модель физики элементарных частиц с помощью неминимального взаимодействия бозона Хиггса с гравитацией [56], выглядят несколько надуманными, как и сама философия отсутствия "новой физики" вплоть до энергий квантовой гравитации.

Многих теоретиков вообще в последнее время пугает инфляционная парадигма [57], поскольку в её рамках возможен режим вечной инфляции [58], при котором наша Вселенная оказывается лишь одним из бесчисленного множества пузырей в Мультиверсе, так что даже предсказание вероятности тех или иных свойств нашего мира оказывается проблематичным (проблема меры), даже после применения антропного принципа.

Возможно, многие проблемы космологии можно решить при помощи отскока (bounce) сжимающейся фазы в расширяющееся решение вместо инфляции, но для этого требуются гораздо более серьёзные нарушения энергетических условий, плохо совместимые даже с самыми базовыми идеями квантовой теории поля (присутствие духов). Лишь в последние годы стали рассматриваться модели, которые потенциально могут разрешить эти сложности – так называемые галилеоны [59], поля со старшими производными в действии, но при этом имеющие уравнения движения второго порядка.

Таким образом, несмотря на первое (и небезосновательное) впечатление грандиозного успеха стандартной космологической модели, с точки зрения фундаментальных принципов очевидно, что требуется серьёзный пересмотр нашего понимания гравитации (и при этом мы даже не обсуждали здесь проблем её квантования!). Более того, есть прямые наблюдательные указание на необходимость такой работы, пусть пока и не вполне убедительные, но многое станет ясным в ближайшем будущем.

Поле для теоретических фантазий здесь весьма велико, и может показаться, что экспериментальная ситуация ещё не готова к тому, чтобы оправдать массивного участия теоретиков в разработке альтернативных теорий гравитации. Однако, как хорошо известно и как мы отчасти увидим в дальнейшем, уже самые первые принципы физической самосогласованности теорий (отсутствие духов, корректность постановки задачи Коши, возможность воспроизведения твёрдо установленных свойств нашей Вселенной) накладывают столь серьёзные ограничения, что часто бывает очень сложно преобразовать простую идею в реально работающую модель.

Таким образом, задачи построения моделей модифицированной гравитации и выяснения их космологических последствий оказываются технически очень интересными для теоретика, и при этом они помогают нам лучше понять свойства гравитационного взаимодействия и особенности наших теорий, а также имеют все шансы получить намного более практические применения к реальной космологии в самом ближайшем будущем.

Глава 2

Ускоренное расширение с векторными полями

В данной главе рассматриваются возможности использования векторных полей в космологии, в первую очередь для описания ускоренного расширения Вселенной, хотя поведение векторных полей в расширяющейся вселенной, разумеется, связано также и с проблемой первичных магнитных полей в космологии [60–62]. Это популярная тематика исследований, интерес к которой был существенно подогрет нашей статьёй [1*] (номерами со звёздочкой мы даём ссылки на список своих публикаций, положенных в основу Диссертации, который приведён в Заключении). Кроме работы [1*], основные результаты этой главы опубликованы в статьях [2*,3*,4*,5*,6*].

Задача реализации инфляционного расширения в моделях с векторными полями привлекала внимание теоретиков с точки зрения потенциального расширения класса возможных теорий, тем более что на тот момент экспериментально не был обнаружен даже бозон Хиггса, то есть фундаментальных скалярных полей известно не было. Тем самым, использование векторных полей могло бы стать интересной альтернативой стандартной скалярной парадигме в инфляции, и более того, можно было ожидать нового взгляда на проблемы низших мультипольных моментов в реликтовом фоне.

При использовании векторных полей возникает, по сути, две принципиальные трудности: отсутствие медленного качения и анизотропия расширения. До нашей статьи [1*] были отдельные работы, предлагавшие модели с векторными полями. Возможно, одной из первых была статья Ларри Форда [63], в которой было предложено рассматривать экстремально плоские потенциалы для векторного поля $V(A^2)$, для которых быстрое убывание векторного поля не приводит к сильному изменению плотности энергии. В более поздней статье [64] была использована тахионная масса, чьё значение подстраивалось ровно так, чтобы обеспечить режим медленного качения. Анизотропия же компенсировалась выбором ортонормированной триады векторых полей. Были также предложены идеи реализации космической триады в рамках неабеливых теорий Янга-Миллса [65, 66].

Отметим, что есть ещё одна идея для реализации пространственно изотропной космологии – времениподобное векторное поле. И такие модели действительно рассматривались [67, 68]. Как мы увидим ниже, в простейших пространственно однородных сценариях решение с векторным полем, направленным вдоль оси времени, невозможно, поэтому, как правило, подобные работы используют спонтанное нарушение лоренц-инвариантности типа эйнштейновского эфира, когда на векторное поле наложена связь, фиксирующая его (времениподобную) норму. При этом тоже возникают проблемы неустойчивости, а также, с нашей точки зрения, в определённом смысле теряется векторный характер модели. Мы подобную возможность в дальнейшем не будем рассматривать.

В нашей работе [1*] было предложено использовать неминимальное взаимодействие векторного поля со скалярной кривизной вида $\frac{R}{6}A^2$, которое позволяет избавиться от необходимости тонкой подстройки тахионной массы. Проблему анизотропии можно решать либо посредством триады, либо большим количеством случайно ориентированных векторных полей. Эта работа породила широкий отклик и многие десятки новых статей о векторной инфляции, возможно, по причине того, что ситуация уже назрела, и многие космологи всерьёз интересовались векторными полями, см. например работу [69].

Разумеется, проблема заключается в том, что мы лишь предложили способ естественного введения эффективной добавки к массе векторного поля, не изменив её тахионного характера для ускоряющихся космологий. Как хорошо известно, тахионная масса векторного поля приводит к появлению духового возбуждения в продольной компоненте поля, что в применении к нашей модели векторной инфляции было впервые отмечено в заметке [70]. (В работе [64] есть аналогичное обсуждение в случае явной тахионной массы, но старые статьи по инфляции с векторными полями не пользовались широкой известностью.) Примерно в то же время мы [2*] обнаружили анизотропную неустойчивость в моделях с большими полями.

Наши модели векторной инфляции оказались фатально неустойчивыми. Однако они привели к активизации исследований по векторным полям в космологии, и в результате появились более жизнеспособные модели, о которых мы кратко расскажем в разделе 2.3, не содержащем наших результатов. В остальных же разделах данной главы представлены наши работы по векторной инфляции – в несколько сокращённом виде, делая упор на те аспекты, которые продолжают оставаться интересными даже после установления неустойчивости исходных моделей.

2.1 Векторная инфляция

Мы рассматриваем массивное векторное поле, неминимально взаимодействующее с гравитацией:

$$S = \int dx^4 \sqrt{-g} \left(-\frac{R}{16\pi} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} \left(m^2 + \frac{R}{6} \right) A_\mu A^\mu \right), \qquad (2.1)$$

где как обычно $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$. Заметим, что в этих работах мы принимали сигнатуру $(+,-,-,-)$. Легко получить уравнения движения:

$$\frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^\mu} \left(\sqrt{-g} F^{\mu\nu} \right) + \left(m^2 + \frac{R}{6} \right) A^\nu = 0.$$

В пространственно плоской фридмановской вселенной

$$ds^2 = dt^2 - a^2(t)\delta_{ik}dx^i dx^k$$

имеем для временной и пространственной компонент

$$-\frac{1}{a^2}\Delta A_0 + \left(m^2 + \frac{R}{6}\right) A_0 + \frac{1}{a^2}\partial_i \dot{A}_i = 0, \tag{2.2}$$

$$\ddot{A}_i + \frac{\dot{a}}{a}\dot{A}_i - \frac{1}{a^2}\Delta A_i + \left(m^2 + \frac{R}{6}\right) A_i - \partial_i \dot{A}_0 - \frac{\dot{a}}{a}\partial_i A_0 + \frac{1}{a^2}\partial_i\left(\partial_k A_k\right) = 0. \tag{2.3}$$

Инвариантным образом описать величину векторного поля можно посредством скаляра

$$I = A^\alpha A_\alpha = A_0^2 - \frac{1}{a^2}A_i A_i,$$

поэтому мы вводим новые переменные $B_i \equiv A_i/a$ вместо A_i. Для пространственно однородного векторного поля ($\partial_i A_\alpha = 0$) немедленно получаем из (2.2), что

$$A_0 = 0,$$

а уравнение (2.3) принимает вид

$$\ddot{B}_i + 3H\dot{B}_i + m^2 B_i = 0,$$

где $H \equiv \dot{a}/a$. Заметим, что эффективная масса для поля B оказалась малой (без поправок порядка H^2) только за счёт сокращения с тахионным вкладом от скалярной кривизны: в обычной ситуации векторные поля в расширяющейся вселенной убывают.

Поскольку полученное уравнение имеет точно такой же вид, что и уравнение движения пространственно однородного канонического скалярного поля, оно в случае достаточно малой массы описывает режим медленного качения. Требуется выяснить, будет ли оно совместно с квази-де-ситтеровским расширением вселенной.

Интерпретируя тензор энергии-импульса как величину, приравниваемую к тензору Эйнштейна, получаем из вариации действия (2.1) по метрике

$$
\begin{aligned}
T^\alpha_\beta &= \frac{1}{4}F^{\gamma\delta}F_{\gamma\delta}\delta^\alpha_\beta - F^{\alpha\gamma}F_{\beta\gamma} + \left(m^2+\frac{R}{6}\right)A^\alpha A_\beta - \frac{1}{2}m^2A^\gamma A_\gamma\delta^\alpha_\beta \\
&+\frac{1}{6}\left(R^\alpha_\beta - \frac{1}{2}\delta^\alpha_\beta R\right)A^\gamma A_\gamma + \frac{1}{6}\left(\delta^\alpha_\beta\Box - \nabla^\alpha\nabla_\beta\right)A^\gamma A_\gamma. \qquad (2.4)
\end{aligned}
$$

Во фридмановской метрике для пространственно однородного векторного поля имеем следующие ненулевые компоненты:

$$
T^0_0 = \frac{1}{2}\left(\dot{B}_k^2 + m^2B_k^2\right), \qquad (2.5)
$$

$$
\begin{aligned}
T^i_j &= \left[-\frac{5}{6}\left(\dot{B}_k^2 - m^2B_k^2\right) - \frac{2}{3}H\dot{B}_kB_k - \frac{1}{3}\left(\dot{H}+3H^2\right)B_k^2\right]\delta^i_j \\
&+\dot{B}_i\dot{B}_j + H\left(\dot{B}_iB_j + \dot{B}_jB_i\right) + \left(\dot{H}+3H^2-m^2\right)B_iB_j. \qquad (2.6)
\end{aligned}
$$

Присутствие внедиагональных компонент показывает, что фоновое векторное поле несовместимо с метрикой изотропной вселенной, что неудивительно.

Одним из возможных решений будет рассмотреть [64] тройку взаимно ортогональных векторных полей $B_i^{(a)}$ одной длины $|B|$, для которых $\sum\limits_i B_i^{(a)}B_i^{(b)} = |B|^2\,\delta^a_b$ и $\sum\limits_a B_i^{(a)}B_j^{(a)} = |B|^2\,\delta^i_j$. Соответственно, для тензора энергии-импульса получаем

$$
T^0_0 = \varepsilon = \frac{3}{2}\left(\dot{B}_k^2 + m^2B_k^2\right),
$$

$$
T^i_j = -p\delta^i_j = -\frac{3}{2}\left(\dot{B}_k^2 - m^2B_k^2\right)\delta^i_j,
$$

а компоненты всех векторных полей удовлетворяют уже знакомому уравнению

$$
\ddot{B}_i + 3H\dot{B}_i + m^2B_i = 0. \qquad (2.7)
$$

Константа Хаббла равна

$$
H^2 = 4\pi\left(\dot{B}_k^2 + m^2B_k^2\right).
$$

Таким образом, все уравнения имеют абсолютно тот же вид, что и в скалярном случае, а значит, возможен инфляционный режим медленного качения.

Другой возможный подход к разрешению проблемы изотропии – рассмотреть большое число N случайно ориентированных полей. Для простоты будем считать, что все поля имеют одинаковую массу и примерно равную амплитуду. Тогда их вклад в плотность энергии оценивается как

$$T_0^0 = \rho \simeq \frac{N}{2}\left(\dot{B}_k^2 + m^2 B_k^2\right).$$

В пространственных компонентах внедиагональные части оказываются подавлены в силу случайности направлений по механизму типа случайного блуждания. Условно запишем это в виде

$$\sum_{a=1}^{N} B_i^{(a)} B_j^{(a)} \simeq \frac{N}{3} B^2 \delta_j^i + O(1)\sqrt{N} B^2.$$

Из (2.6) видим, что в течение инфляции внедиагональные компоненты имеют порядок $H^2\sqrt{N}B^2$. Почти изотропное инфляционное решение получается самосогласованным только в том случае, если они много меньше диагональных компонент $T_i^i \sim T_0^0 \sim H^2$; соответственно должно выполняться $B < 1/N^{1/4}$. С другой стороны, условие медленного качения нарушается, и инфляция заканчивается, когда $H \simeq m$. Учитывая, что во время инфляции

$$H^2 = \frac{8\pi}{3}\rho \simeq \frac{4\pi}{3} N m^2 B^2, \tag{2.8}$$

находим, что при $B \simeq 1/N^{1/2}$ инфляция заканчивается, и векторные поля переходят в осциллирующий режим. Сравнивая два условия, получаем возможный интервал значений векторного поля

$$\frac{1}{\sqrt[4]{N}} > B > \frac{1}{\sqrt{N}}$$

для поддержания инфляционного расширения.

Число возможных e-фолдингов инфляции оценивается стандартным образом. А именно, учитывая, что в режиме медленного качения

$$\dot{B} \approx -\frac{m^2 B}{3H}, \tag{2.9}$$

можно в соотношении $a \propto \exp\left(\int H(t)dt\right)$, где константа Хаббла дается выражением (2.8), заменить переменную интегрирования с t на B, и тогда находим возрастание масштабного фактра за время инфляции

$$\frac{a_f}{a_i} \simeq \exp\left(2\pi N B_{in}^2\right),$$

где B_{in} – начальное значение полей (конечным значением пренебрегли). Принимая $B_{in} \simeq N^{-1/4}$, находим $2\pi\sqrt{N}$ e-фолдингов. Таким образом, для объяснения начальных условий для современной Вселенной требуется несколько сотен векторных полей. При этом будет предсказана анизотропия порядка $\simeq 1/\sqrt{N}$ (несколько процентов) в конце инфляции, которая может сказаться на низших мультиполях в реликтовом фоне.

Разумеется, данная модель может быть обобщена на случай более сложного распределения полей по массам. И в частности, самые лёгкие поля могли бы до сих пор пребывать в режиме медленного качения, представляя собой Тёмную Энергию.

Случай произвольного потенциала

Конечно же, массовый член в качестве потенциала для векторного поля был выбран только из эстетических соображений. Можно рассмотреть произвольный потенциал $V\left(A_\alpha A^\alpha\right)$ вместо $\frac{1}{2}m^2A^2$. В таком случае получаем уравнения движения

$$\ddot{B}_i + 3\frac{\dot{a}}{a}\dot{B}_i + V'(B^2)B_i = 0,$$

где штрихом обозначена производная потенциала по его аргументу (не путать с производной по конформному времени!). Тензор энергии-импульса во фридмановской вселенной принимает значение

$$T_0^0 = \frac{1}{2}\left(\dot{B}_k^2 + V\left(B^2\right)\right),$$

$$\begin{aligned} T_j^i &= \left[-\frac{5}{6}\dot{B}_k^2 + \frac{1}{2}V\left(B^2\right) - \frac{2}{3}H\dot{B}_k B_k - \frac{1}{3}\left(\dot{H} + 3H^2 - V'\left(B^2\right)\right)B_k^2\right]\delta_j^i \\ &+\dot{B}_i\dot{B}_j + H\left(\dot{B}_i B_j + \dot{B}_j B_i\right) + \left(\dot{H} + 3H^2 - V'\left(B^2\right)\right)B_i B_j, \end{aligned}$$

а после усреднения по N полям даёт

$$T_j^i = -p\delta_j^i \simeq \frac{N}{2}\left(-\dot{B}_k^2 + V\left(B^2\right)\right)\delta_j^i.$$

Выход из инфляции происходит при $V'B/V \sim \sqrt{N}$. Поскольку значение векторного поля при выходе из инфляции B_f зависит от формы потенциала, это позволяет управлять остаточной анизотропией, которая имеет порядок $\sqrt{N}B_f^2$.

2.2 Космологические возмущения в векторной инфляции

К сожалению, теории векторной инфляции для достижения режима медленного качения используют неминимальное взаимодействие с гравитацией, которое действует как тахионная масса. Для продольной поляризации векторного поля это очень плохая новость [70–72], поскольку она становится духовой степенью свободы (отрицательная кинетическая энергия), что легко понять с помощью трюка Штюкельберга.

Эта проблема проявляется в том, что несмотря на хорошие фоновые уравнения, уже линейные возмущения становятся неустойчивыми. Тем не менее, некоторые особенности возмущений оказываются сами по себе поучительными в результате вмешательства фоновых векторов, вообще говоря не позволяющих расщепить флуктуации на три невзаимодействующих сектора. Поскольку сейчас рассматриваются другие, потенциально жизнеспособные, модели с векторными полями, обсуждение общих свойств теории возмущений в присутствии фоновых векторных полей может быть полезно.

2.2.1 Гравитационные волны

Рассмотрим для начала простой случай, когда к фоновой метрике добавлены только поперечные бесследовые возмущения в пространственном секторе (гравитационные волны). Разумеется, это не вполне самосогласованная постановка задачи, но она позволяет увидеть некоторые интересные особенности модели. Итак, переходя к конформному времени, мы рассматриваем метрику

$$ds^2 = a^2(\tau)\left(d\tau^2 - (\delta_{ik} - h_{ik})\, dx_i dx_k\right),$$

где $h^i{}_i = 0$ и $h^i{}_{j,i} = 0$.

Других тензорных переменных в возмущениях нет. Однако слагаемые типа $A_i \delta A_j$, строго говоря, взаимодействуют с гравитационными волнами. Мы их откинем и посмотрим, что получится. Наивно (см. ниже) можно ожидать, что в случае большого числа случайно ориентированных векторов усреднение подавит подобные члены. А в случае триады отбрасывание этих слагаемых соответствует предположению, что по тем или иным причинам допустимы лишь возмущения, поворачивающие или растягивающие триаду как целое, но не нарушающие условие изотропии для полного тензора энергии-импульса. По крайней мере, так можно получить ту часть уравнений для тензорных мод, которая не зависит от взаимодействия со скалярным и векторным секторами.

Опуская громоздкие (но несложные) выкладки, получаем квадратичное действие

$$S_{gw} \approx \frac{1}{8}\int a^2 \left[\left(\frac{1}{8\pi} + \frac{NB^2}{6}\right)\left(h'^2_{ik} - h^2_{ik,j} - m_g^2 h^2_{ik}\right)\right] dx^3 d\tau \qquad (2.10)$$

и уравнение движения

$$h''_{ik} + 2\left(\mathcal{H} + \frac{4\pi N B B'}{3 + 4\pi N B^2}\right) h'_{ik} - \triangle h_{ik} = -m_g^2 h_{ik} \qquad (2.11)$$

в интересующем нас секторе, где эффективная масса равна

$$m_g^2 \equiv -16\pi N \frac{\left(\frac{a''}{a^3} - 2V_{,I} - \frac{4}{5}B^2 V_{,II}\right) a^2 B^2 + \left(B' + B\mathcal{H}\right)^2}{3 + 4\pi N B^2}, \qquad (2.12)$$

и мы поменяли обозначение производных потенциала на $V_{,I} \equiv \frac{dV}{dI}$, $V_{,II} \equiv \frac{d^2V}{dI^2}$, где $I \equiv A^2$, во избежание путаницы с производными по конформному времени. Массовое слагаемое происходит из членов типа $A_\mu A_\nu g^{\mu\nu}$ и $F_{\mu\nu} F_{\alpha\beta} g^{\mu\alpha} g^{\nu\beta}$. В режиме медленного качения m_g сама меняется медленно.

При медленном качении из $\ddot{B} \ll H\dot{B}$ следует $B'' \approx \mathcal{H}B'$, и можно получить

$$m_g^2 \approx 16\pi N \frac{\left(-8\pi N V + \frac{10}{3} V_{,I} + \frac{4}{5} B^2 V_{,II}\right) a^2 B^2}{3 + 4\pi N B^2}. \qquad (2.13)$$

Например, в случае простейшей хаотической ифляции на массовом слагаемом ($V\left(B^2\right) = \frac{1}{2} m^2 B^2$) получаем

$$m_g^2 \approx 16\pi m^2 a^2 N B^2 \left(\frac{5 - 12\pi N B^2}{9 + 12\pi N B^2}\right),$$

и вспоминая, что для инфляции требуется $B \gtrsim \frac{1}{\sqrt{N}}$, мы видим, что гравитационные волны могут начать расти за счёт эффективной тахионной массы.

Заметим, что, в отличие от стабильных случаев, это заключение вряд ли может измениться в полной картине от того, что у тензорных мод помимо тахионной массы появятся еще и "внешние силы" от других типов возмущений. В работе [73] было высказано предположение, что этот результат – не более чем артифакт приближённого анализа. Однако нетрудно убедиться, что даже в однородном режиме анизотропия нарастает [4*], рассмотрев пространство типа Бианки I.

Аналогично можно рассмотреть ряд других традиционно используемых потенциалов и убедиться, что в случае инфляции хаотического типа масса обычно оказывается большой тахионной (что и не удивительно, поскольку в этих моделях векторные поля имеют большую величину и действуют вполне аналогично случаю инфляции на массовом слагаемом), а для потенциалов типа новой инфляции гравитационные волны (в рамках нашего "приближения") оказываются стабильными. Конкретные вычисления и ответы вряд ли представляют существенный интерес в свете наличия духовой моды, но их можно найти в нашей работе [2*], при написании которой нам ещё не было очевидно, что у модели есть гораздо более серьёзные проблемы, чем тахионный рост гравитационных волн.

2.2.2 Общие возмущения – формализм

Перейдём к общему случаю произвольных линейных возмущений в векторной инфляции. Мы рассматриваем действие

$$S = \int \sqrt{-g}\left[-\frac{R}{2}\left(1+\sum_{n=1}^{N}\frac{1}{6}I_{(n)}\right)-\frac{1}{4}\sum_{n=1}^{N}F^{(n)}_{\mu\nu}F^{\mu\nu}_{(n)}-\sum_{n=1}^{N}V\left(I_{(n)}\right)\right]dx^4, \tag{2.14}$$

где $I_{(n)} \equiv -A^{(n)}_{\mu}A^{\mu}_{(n)}$. Фоновые уравнения движения имеют вид

$$\begin{aligned} A_0 &= 0 \\ B_i'' + 2\mathcal{H}B_i' + 2V_{,I}a^2B_i &= 0 \end{aligned} \tag{2.15}$$

где $I = B_iB_i \equiv B^2$, $B_i \equiv \frac{A_i}{a}$, а уравнения Эйнштейна

$$3\mathcal{H}^2 = 8\pi N\left(Va^2 + \frac{B'^2}{2}\right), \tag{2.16}$$

$$2\mathcal{H}' + \mathcal{H}^2 = 8\pi N\left(Va^2 - \frac{B'^2}{2}\right). \tag{2.17}$$

Флуктуации метрики в ньютоновской калибровке, как мы объясняли в первой главе, можно записать как

$$ds^2 = a^2(\tau) \cdot \big((1 + 2\phi)\, d\tau^2 + 2\mathcal{V}_i d\tau dx^i - ((1 - 2\psi)\, \delta_{ik} - h_{ik})\, dx_i dx_k \big)$$

с условиями $\mathcal{V}^i_{,i} = 0$, $h^i_i = 0$, $h^i_{j,i} = 0$. Для возмущений векторных полей напишем

$$\delta A_\alpha = \left(\delta A_0, \chi_{,i} + \delta A^T_i \right) \tag{2.18}$$

где по определению $\left(\delta A^T\right)^i_{,i} = 0$. Итого имеется $2 + 2N$ скалярных переменных (ϕ, ψ, δA_0 и χ), $2 + 2N$ компонент от $N + 1$ бездивергентного вектора (V_i и δA^T_i) и 2 тензорные поляризации (h_{ij}).

Линейная вариация тензора Эйнштейна стандартна, а вариация тензора энергии-импульса чрезвычайно громоздка. Явный вид всех этих выражений можно найти в нашей работе [3*], а здесь мы ограничимся выписыванием возмущённых уравнений движения для векторных полей:

$$(\phi + \psi)_{,i} A'_i + \triangle \left(\delta A_0 - \chi' \right) + \left(\frac{a''}{a} - 2a^2 V_{,I} \right) \left(\delta A_0 + \mathcal{V}_i A_i \right) = 0, \tag{2.19}$$

$$\begin{aligned} -2\phi A''_i - (\phi' + \psi')\, A'_i - \delta A'_{0,i} - \triangle \delta A^T_i - A'_k \mathcal{V}_{i,k} + h'_{ik} A'_k + \delta A''_i \\ + \left(2a^2 V_{,I} - \frac{a''}{a} \right) \delta A_i + 2\frac{d^2 V}{dI^2} A_i \left(2\psi A_l^2 + 2A_l \delta A_l + A_l A_j h_{lj} \right) \\ + \left(\psi'' + \frac{1}{3} \triangle (\phi - 2\psi) + \mathcal{H} (\phi' + 3\psi') + 2\phi \frac{a''}{a} \right) A_i = 0. \end{aligned} \tag{2.20}$$

Как видим, они тоже весьма громоздки. Однако легко усмотреть принципиальный момент, отличающий эту теорию возмущений от случая скалярной инфляции: разные типы возмущений не отщепляются, и в линеаризованные уравнения для векторного поля входит тензорная величина h_{ij}. В уравнениях Эйнштейна часть подобных слагаемых обратится в нуль, или по крайней мере будет подавлена. Так, можно убедиться, что в случае изотропного фона $\sum\limits_{n=1}^{N} h_{ik} A^{(n)}_i A^{(n)}_k = 0$. Но это верно отнюдь не для всех типов смешивающих эффектов. Напомним, хотя бы, что $\delta A^0 = \delta A_0 + \mathcal{V}_i B_i$ (смешивание скаляров и векторов).

Линейные возмущения при больших и при малых полях

Можно несколько продвинуться в анализе, предполагая, что инфляция происходит на малых полях $NB^2 \ll 1$ (модели типа новой инфляции), как того требует устойчивость по отношению к рождению гравитационных волн.

Легко видеть, что лидирующие члены в (2.4) имеют порядок $H^2 B^2$, но при усреднении по всем полям они сокращаются (точно для триады или приближенно для случайного набора полей). Но во флуктуациях подобные слагаемые, разумеется, надо учесть.

Поначалу может показаться, что флуктуации должны быть подавлены в силу усреднения по большому число независмо флуктуирующих полей даже в случае инфляции на больших полях. Например, в T_{00} есть слагаемое $H^2 \sum A_j \delta A^j$. Вклад каждого отдельного поля может быть как положительным, так и отрицательным, что наивно должно привести к подавлению флуктуаций при $N \to \infty$.

Однако это неправильное заключение. В самом деле, феноменологически инфляция характеризуется значением постоянной Хаббла, и мы должны удерживать её неизменной при увеличении числа полей. При инфляции на массовом члене это означает $A \propto 1/\sqrt{N}$, и несмотря на статистическое подавление типа $1/\sqrt{N}$ сумма N слагаемых $H^2 \sum A_j \delta A^j$ оказывается независящей от N. Отметим, что это вполне аналогично случаю скалярной N-фляции [74].

Итак, флуктуации оказываются значимыми, в том числе в тех слагаемых, которые выпадают из фоновых уравнений за счёт изотропной конфигурации фоновых полей. Так слагаемое $H^2 \sum B_i \delta B_j$ будет давать флуктуации порядка $\sqrt{N} H^2 B \langle \delta B \rangle$.

Если рассматривается инфляция на больших полях, то эффекты смешивания могут быть весьма значительны. Оценивая амплитуду флуктуаций отдельного поля при инфляции на массовом слагаемом как $\delta B \propto$

$H \propto mB\sqrt{N}$, получаем $\sqrt{N}\langle B\delta B\rangle \propto mB^2N$. Если инфляция начинается при $B \sim N^{-1/4}$, то эта флуктуация столь велика, что вообще делает всё рассмотрение ненадежным.

Таким образом, в моделях типа хаотической инфляции не только эффективная масса у гравитона велика и является тахионной, но ещё и присутствует большая вынуждающая сила со стороны флуктуаций векторного поля. Хоть и контринтуитивно, можно было бы предположить, что эти два эффекта способны компенсировать друг друга. Однако, как мы уже упоминали, это не так [4*]: анизотропия действительно нарастает.

Вместе с тем, при инфляции с малыми полями, величина $\sqrt{N}B\delta B \propto H\sqrt{N}B$ может быть сделана произвольно малой. Если бы не проблемы устойчивости продольной компоненты векторного поля, возмущения в такой модели могли бы быть космологически приемлемыми [75–77], имея при этом потенциально интересные эффекты. В частности, могли бы рождаться в некотором количестве векторные моды.

2.2.3 Проблема духа в продольных компонентах

Как уже упоминалось выше, продольная мода тахионного векторного поля – это духовое возбуждение [70,71]. С точки зрения трюка Штюкельберга после замены $A_\mu \longrightarrow A_\mu + \partial_\mu\chi$, которая не меняет кинетического слагаемого, видим, что отрицательность m^2 становится утверждением о неправильном знаке кинетической энергии поля χ. Обсудим это несколько подробнее.

Векторное поле в пространстве Минковского

Без обращения к языку Штюкельберга можно понять природу проблем с продольной модой путём рассмотрения уравнения связи (2.2). Для простоты будем предполагать пространство Минковского. Пространственные компоненты векторного поля удовлетворяют обычному уравнению Клейна-Гордона $\left(\Box + m^2\right)\overrightarrow{A} = 0$, которое при тахионной массе является неустойчивым, но обладает характерным временем развития неустойчивости, определяемым величиной $|m|^{-1}$. Настоящая проблема присутствует во временной компоненте, которая определяется уравнением связи (2.2), $\left(- \triangle + m^2\right) A_0 + \partial_i \dot{A}_i = 0$. В терминах Фурье компонент имеем

$$A_0 = -\frac{i k_i \dot{A}_i}{k^2 + m^2}, \tag{2.21}$$

что расходится при $k^2 = -m^2 > 0$. На этой длине волны продольная мода запрещена, а компонента A_0 произвольна. В проколотой окрестности этого значения волнового числа продольная мода существует, но сопряжена со сколь угодно большими значениями A_0.

Разумеется, столь просто эта картина выглядит только благодаря линейности рассматриваемого уравнения. При включении взаимодействий расходимость одной из компонент поля станет катастрофой. Конечно, наивно можно было бы надеяться на то, что нелинейность может, наоборот, помочь, устранив полюс резонансного характера в (2.21), например, сделав эффективную массу положительной при больших значениях A_0.

Ясно, однако, что это невозможно без нарушения лоренц-инвариантности, что хорошо согласуется с картиной Штюкельберга. Известно, что в лоренц-инвариантной теории поля дух представляет собой очень серьёзную проблему, если только речь идёт не об эффективной теории при духовом полюсе, лежащем выше масштаба обрезания теории, что в данном случае заведомо не так (эффективный массовый параметр m^2 должен быть мал, чтобы обеспечить режим медленного качения). Разумеется, несложно также построить гамильтонов формализм, чтобы убедиться, что гамильтониан неограничен снизу.

Заметим, что в работах [73,75–77] защищается точка зрения, что векторная инфляция является достаточно жизнеспособной теорией (и даже можно вычислить негауссовость). Однако, это достигается, по сути, игнорированием проблемы. В длинноволновом режиме применяется $\delta\mathcal{N}$ формализм [78], а проблемы на хаббловских и более мелких масштабах игнорируются на основе того понимания, что неминимальное взаимодействие может быть лишь эффективным слагаемым, а сама скалярная кривизна тоже не может уже считаться лишь числовым параметром на малых масштабах, поскольку испытывает флуктуации. Строго говоря, всё это верно, но ясно, что последовательной теории у нас при этом больше нет, тем более что дух появляется не слишком глубоко под гоизонтом. Интересно однако попытаться явно описать поведение продольной моды в опасном режиме.

Продольная мода при инфляции

Для продольной моды ($k_i B_i = kB$) можно выразить временную компоненту из уравнения связи (2.2), чтобы получить уравнение движения в замкнутом виде. Предполагая пространство де Ситтера ($\frac{R}{6} = -2H^2$), имеем

$$\ddot{B} + \left(3H + \frac{2H\frac{k^2}{a^2}}{\frac{k^2}{a^2} + m^2 - 2H^2}\right)\dot{B} + \left(\frac{k^2}{a^2} + m^2 + \frac{2H^2\frac{k^2}{a^2}}{\frac{k^2}{a^2} + m^2 - 2H^2}\right)B = 0. \tag{2.22}$$

Проблема существует при $\frac{k^2}{a^2} = 2H^2 - m^2$.

Для простоты будем пренебрегать массой векторного поля $m \to 0$ (её включение качественно результатов не меняет). Рассмотрим длину волны, которая в некоторый момент времени t_0 была глубоко под горизонтом, скажем $\frac{k^2}{a^2} = 4H^2$. Тогда имеем $\frac{k^2}{a^2} = 4H^2 e^{-2H(t-t_0)}$, а уравнение (2.22) принимает вид

$$\left(2 - e^{2H(t-t_0)}\right)\ddot{B} + \left(10H - 3He^{2H(t-t_0)}\right)\dot{B} + 8H^2 e^{-2H(t-t_0)}B = 0.$$

Вводя новую переменную времени $\tau = 2H(t-t_0) - \ln 2$ (не путать с конформным временем!), получаем

$$2(1 - e^{\tau})\ddot{B} + (5 - 3e^{\tau})\dot{B} + e^{-\tau}B = 0. \tag{2.23}$$

Критическое значение волнового числа пересекается при $\tau = 0$. В этот, и только в этот, момент коэффициент при $\ddot{B}$ обращается в нуль. Это означает, что двухпараметрическое семейство решений касается однопараметрического семейства кривых $\dot{B} = -\frac{B}{2}$ в этой точке. Поведение это устойчиво, поскольку если $\dot{B} \neq -\frac{B}{2}$ при некотором малом $\tau < 0$, то вторая производная $\ddot{B} \sim \frac{\dot{B}+\frac{B}{2}}{\tau}$ имеет подходящий знак и корректирует траекторию.

Мы хотим понять, как ведут себя решения уравнения (2.23) в окрестности точки $\tau = 0$. Будем считать, они могут быть представлены в виде ряда по степеням τ. Ясно, что разложение либо должно начинаться с члена τ^2 (продольная мода исчезает при подходе к сингулярной точке), либо следует взять $1 - \frac{1}{2}\tau$, чтобы удовлетворить уравнению касательного семейства. Итак, можно искать решение в виде $B = \alpha - \frac{\alpha}{2}\tau + \beta\tau^2 + \sum\limits_{n\geq 3} C_n\tau^n$. Легко проверить, что первые три члена решают уравнение (2.23) с точностью до $\mathcal{O}(\tau^2)$ при произвольных параметрах α и β, а рассмотрение следующих приближений позволяет последовательно определять все коэффициенты C_n через значения этих двух параметров. Для компоненты A_0 можно раскрыть неопределённость типа $\frac{0}{0}$ и определить конечное значение, через которое она проходит в особой точке.

Таким образом, эволюция отдельно взятой продольной моды проходит через особую точку гладко. Интересно, что тот же результат был получен в работе [72] для векторного поля во фридмановской вселенной с помошью численного моделирования. Однако это верно лишь в предположении пробного поля (пренебрежение реакцией метрики на присутствие векторного поля). При изучении полной системы уравнений решение расходилось в некоторый момент времени. С точки зрения нашего анализа этот результат может показаться несколько неожиданным, поскольку учёт флуктуаций метрики скорее размывает момент прохода через особое значение, а особенность определяется обращением в нуль коэффициента при $\ddot{B}$.

Отметим, что ещё одна особенность решений была отмечена в работе [72]. Она возникает при выходе из инфляции. Однако эта особенность совершенно естественна, поскольку происходит переход от тахионной массы к обычной, а в самый момент перехода оказывается восстановленной калибровочная инвариантность, что делает плохо определённым число степеней свободы.

2.2.4 Проблема дополнительной степени свободы

Оказывается, что у векторной инфляции есть ещё одна проблема – существование дополнительной степени свободы, которая при рассмотрении над пространственно однородным фоном находится в режиме сильной связи. Это связано с временными компонентами векторов.

Наивный подсчёт числа степеней свободы дает $3N + 2$, по 3 поляризации для каждого массивного векторного поля и 2 поляризации гравитона. Однако же, поле неминимально взаимодействует со скалярнной кривизной (RA^2), а скалярная кривизна содержит вторые производные метрики по времени. В результате временные компоненты (2.2) уравнений движения векторных полей перестают быть связями. С другой стороны, при интегрировании по частям в действии для избавления от вторых производных появится производная от A_0. Если быть точным, то появится производная от $\sum A_0^2$, где сумма берётся по всем векторным полям. Отсюда сразу можно сделать два заключения. Во-первых, дополнительная степень – одна для всех полей, и в этом смысле (а также по происхождению) носит по сути гравитационный характер. Во-вторых, в линейном приближении вокруг решений с $A_0 = 0$ (отсутствие продольных мод) она исчезает (сильная связь).

Полезно обсудить эту же проблему и в эйнштейновской картине. А именно, сделав конформное (или вейлевское) преобразование над метрикой $\tilde{g}_{\mu\nu} = e^{2\rho} g_{\mu\nu}$, можно убедиться, что скалярная кривизна меняется как

$$\tilde{R} = e^{-2\rho} \left(R - 6\Box\rho - 6(\nabla\rho)^2 \right) .$$

Соответственно, оставляя компоненты векторов с нижними индексами неизменными и выбирая конформный фактор в виде

$$\rho = \frac{1}{2} \ln \left(1 + \frac{1}{6} \sum_{n=1}^{N} I_{(n)} \right) ,$$

преобразовываем исходное действие к виду с эйнштейновской гравитацией с минимально связанной материей:

$$S = \int dx^4 \sqrt{-g} \left(-\frac{R}{2} + \frac{1}{2} \nabla_\mu \varphi \nabla^\mu \varphi - \mathbf{V} - \sum_n \frac{1}{4} F^{(n)}_{\mu\nu} F^{(n)}_{\alpha\beta} g^{\mu\alpha} g^{\nu\beta} \right). \quad (2.24)$$

Вся сложность модели перенесена в сектор (векторных) инфлатонов и проявляется в виде сложного кинетического слагаемого с переменной

$$\varphi = \sqrt{\frac{3}{2}} \cdot \ln \left(1 + \frac{1}{6} \sum_n I_{(n)} \right),$$

а также в форме потенциала

$$\mathbf{V} = \left(1 + \frac{1}{6} \sum_m I_{(m)} \right)^2 \cdot \sum_n V \left(\left(1 + \frac{1}{6} \sum_p I_{(p)} \right)^{-1} \cdot I_{(n)} \right).$$

Последний может быть существенно упрощён в частных случаях, таких как инфляция на массовом слагаемом, но для текущего обсуждения его форма вообще не важна.

Рассмотрим материальную часть действия (2.24) с одним ($N = 1$) векторным полем в пространстве Минковского. Канонические импульсы находятся как обычно:

$$\pi_0 = 4 A_0 \varphi_{,I}^2 \left(A_0 \dot{A}_0 - A_j \dot{A}_j \right),$$

$$\pi_i = -4 A_i \varphi_{,I}^2 \left(A_0 \dot{A}_0 - A_j \dot{A}_j \right) + \dot{A}_i - \partial_i A_0.$$

Рассматривая эти соотношения как уравнения на скорости, находим, что детерминант полученной системы равен

$$-\left(4\varphi_{,I}^2\right)^3 A_0^2 \left(A_1^2 A_2^2 + A_2^2 A_3^2 + A_3^2 A_1^2 \right),$$

что отлично от нуля при $A_0 \neq 0$, и следовательно (в присутствии продольных мод) динамическая система является системой без связей (4 степени свободы). Ни одна из компонент уравнений движения (или их комбинаций)

$$2A^{\nu}\frac{\partial\varphi}{\partial I}\left(\Box\varphi\right)+\nabla_{\mu}F^{\mu\nu}+2A^{\nu}\frac{\partial\mathbf{V}}{\partial I}=0 \tag{2.25}$$

не является связью из-за присутствия величины $\Box\varphi$.

В случае произвольного числа полей легко убедиться, что новая мода всё-равно только одна. В самом деле, взяв два произвольных векторных поля из имеющихся в действии (без нарушения общности назовём их первым и вторым), умножим уравнение движения первого из них на $2A^0_{(2)}\frac{\partial\varphi}{\partial I_{(2)}}$, а второго – на $2A^0_{(1)}\frac{\partial\varphi}{\partial I_{(1)}}$, и вычтем одно из другого. Получено уравнение связи. Таких независимых связей можно получить $N-1$.

Соответственно, если хоть одно из полей имеет $A_0\neq 0$, появляется дополнительная степень свободы. Поскольку фоновое космологическое решение соответствует обращающейся в нуль временной компоненте всех полей, оно отвечает режиму сильной связи для этой моды, делая теорию возмущений ненадёжной.

Заметим, однако, что если бы не было проблемы продольного духа, с этой трудностью можно было бы бороться. Дело в том, что степень свободы появляется из-за старших производных метрики (на самом деле, ньютоновского потенциала ψ). Если заставить эти производные "работать" вне зависимости от векторных полей (в том числе при всех $A_0=0$), то проблему сильной связи можно устранить. Сделать это несложно, хотя бы с помощью поправки типа R^2 в действии.

2.2.5 О возможностях модификации лагранжиана

Естественно возникает вопрос о том, сколь серьёзные модификации требуются для того, чтобы сделать векторную инфляцию более стабильной. Как мы видели, для векторных полей типа Прока треубуют решения проблема изотропии и отсутствие медленного качения. Если первую можно успешно решить, по крайней мере в моделях с малыми полями, то предпринятая попытка решения второй с помощью неминимального взаимодействия с гравитацией оказалась неудачной в силу духовой неустойчивости. В этом разделе мы, следуя нашей работе [5*], рассмотрим данный вопрос, не переходя ко многокомпонентным полям, таким как в теориях Янга-Миллса, и считая калибровочную симметрию нарушенной.

Если обратиться к старой работе [63], то там было рассмотрено два варианта решения проблемы медленного качения. Одна идея состояла в использовании большой тахионной массы. Однако она должна быть тонко подстроена к значению константы Хаббла. В этом смыле использование неминимального взаимодействия со скалярной кривизной позволяет обойти проблему неестественности, но к сожалению не духовую степень свободы.

Вторым примером из работы [63] было использование чрезвычайно, экспоненциально плоского потенциала для векторного поля, так чтобы весьма быстрое качение аргумента потенциальной энергии приводило бы лишь к очень медленному изменению последней. При всей возможной неестественности подобной идеи, даже она не всегда позволяет избавиться от неустойчивостей.

Например, можно попытаться использовать потенциал $V = -\frac{B^2}{|B^2|^{1-\epsilon}}$ с очень малым $\epsilon > 0$ (нас интересуют лишь большие отрицательные значения B^2). Однако в таком случае продольная мода, хоть и не является духовой, но развивает градиентную неустойчивость. В самом деле, рассмотрим фоновое решение в виде $B_\mu = \delta^1_\mu \mathcal{B}(t)$ и чисто продольное возмущение $\delta B_\mu = \partial_\mu \lambda$. Тогда разложение потенциала в ряд по степеням возмущения принимает вид

$$\begin{aligned}
-V &= \frac{-\mathcal{B}^2 - 2\mathcal{B}\partial_1\lambda + (\partial_\mu\lambda)(\partial^\mu\lambda)}{\left(\mathcal{B}^2 + 2\mathcal{B}\partial_1\lambda - (\partial_\mu\lambda)(\partial^\mu\lambda)\right)^{1-\epsilon}} \\
&= \mathcal{B}^{2\epsilon}\left(-1 - 2\epsilon\frac{\partial_1\lambda}{\mathcal{B}} + \epsilon\frac{(\partial_\mu\lambda)(\partial^\mu\lambda)}{\mathcal{B}^2} + (2\epsilon - 2\epsilon^2)\frac{(\partial_1\lambda)^2}{\mathcal{B}^2} + \mathcal{O}(\lambda^3)\right).
\end{aligned}$$

Разумеется, коэффициенты сильно зависят от времени, но при малых значениях ϵ продольная мода явно страдает градиентной неустойчивостью. Для стабильности требуется $\epsilon \geqslant \frac{1}{2}$, но тогда потенциальная энергия будет убывать уже быстрее $\frac{1}{a}$.

Работоспособный пример можно построить с потенциалом $V = C\left(e^{\kappa\sqrt{|B^2|}} - 1\right)$. Градиентная неустойчивость не возникает, поскольку

$$
\begin{aligned}
-V = C\left(1 - e^{\kappa\mathcal{B}}\right) \\
+ Ce^{\kappa\mathcal{B}}\left(-\kappa\partial_1\lambda + \frac{\kappa}{2\mathcal{B}}(\partial_\mu\lambda)(\partial^\mu\lambda) - \left(\frac{\kappa^2}{2} - \frac{\kappa}{2\mathcal{B}}\right)(\partial_1\lambda)^2\right) \\
+ \mathcal{O}(\lambda^3),
\end{aligned}
$$

и знак при $(\partial_1\lambda)^2$ всегда правильный.

В общем случае имеем

$$
-\delta V = -V'\left(2\mathcal{B}\partial_1\lambda + (\partial_\mu\lambda)(\partial^\mu\lambda)\right) - 2V''\mathcal{B}^2(\partial_1\lambda)^2 + \mathcal{O}(\lambda^3),
$$

а для стабильности надо $V' < 0$ и

$$
V'' \geqslant \frac{V'}{2\mathcal{B}^2}.
$$

В частности, легко проверить, что потенциал из [63], $V = C\left(1 - e^{\kappa B^2}\right)$ устойчив по отношению к градиентной нестабильности, если $\mathcal{B}^2 \geqslant \frac{1}{2\kappa^2}$.

* * *

Можно ли построить более элегантные решения, если перейти к более сложным лагранжианам? Путеводной нитью может служить задача построить теорию с векторным полем, которое не растворяется при расширении Вселенной. Причем следует иметь в виду, что конечно же речь дожна идти об энергии этого поля. Нельзя добиться физически интересных следствий, просто сделав замену переменных типа $A_\mu = f(\mathfrak{A}^2)\cdot\mathfrak{A}_\mu$ так, чтобы величина $\mathfrak{A}^2$ убывала бы медленно. В работе [5*] мы приводим аккуратный гамильтонов анализ так получающейся модели, чтобы показать, что она действительно эквивалентна исходной. За очевидностью результата (при удручающей громоздкости выкладок) в Диссертации мы эти рассмотрения опускаем.

Легко убедиться, что форма кинетического слагаемого, даже для поля с нарушенной калибровочной инвариантностью, выбрана неслучайно. Она определяется желанием иметь три степени свободы. Если рассмотреть все возможные квадратичные лагранжианы для вектор-потенциала, то в случае общего положения они приводят к четырём степеням свободы, одна из которых получается духовой в силу знаконеопределённости метрики на группе Лоренца. Исключением является лишь максвелловское слагаемое F^2, да неинтересный лагранжиан $(\partial_\mu A^\mu)^2$ с одной степенью свободы.

Таким образом, кирпичиками для конструирования теории являются обычные инварианты F^2 и A^2. Рассмотрим лагранжианы типа $\mathcal{L} = -f(F^2) - V(A^2)$ с двумя нелинейными функциями f и V. В работе [79] было построено ускоренно расширяющееся решение с функцией $f(F^2) = \frac{F^2}{4} - \frac{c}{F^2}$, у которой есть весьма проблематичная для динамики точка $f' = 0$. На самом деле, легко проверить [80], что для ограниченности гамильтониана снизу необходимые условия $f' > 0$ и $V' < 0$. К сожалению, в рамках этих условий невозможно построить инфляционной модели, если только не рассматривать чрезвычайно плоские функции как в работе Форда, поскольку в статье [80] было показано, что при этих условиях ни одна из величин $\frac{A_i}{a}$ и $\frac{F_{0i}}{a}$ не может находиться в режиме медленного качения. (Там же указано, что введение дуального тензора напряжённости в слагаемом типа $F\tilde{F}$ не приводит к успеху.) К разговору о "тонко подстроенных" моделях мы еще вернёмся в последнем разделе этой главы.

Остаётся ещё одна разумная возможность - кинетическое самодействие другой формы: $f(A^2)F^2$. Рассмотрим для простоты теорию с массивным членом в потенциале

$$\mathcal{L} = -\frac{1}{4}f(A^2)F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2A^2.$$

Канонические импульсы равны $p^i = fF_{0i}$ и $p^0 = 0$ (первичная связь). Гамильтонова плотность легко вычисляется

$$\mathcal{H} = \frac{(p^i)^2}{2f} - A_0 \partial_i p^i + \frac{1}{4} F_{ik} F_{ik} - \frac{1}{2} m^2 A^2$$

и порождает вторичную связь

$$-\partial_i p^i = m^2 A_0 + \frac{(p^i)^2 f'}{f^2} A_0 - \frac{1}{2} f' A_0 F_{ik} F_{ik},$$

которая позволяет записать гамильтониан в виде

$$\mathcal{H} = \frac{(p^i)^2}{2f} \left(1 + 2\frac{f'}{f} A_0^2\right) + \frac{1}{4} F_{ik} F_{ik} \left(1 - 2\frac{f'}{f} A_0^2\right) + \frac{1}{2} m^2 \left(A_0^2 + A_i^2\right).$$

Из вида первых двух слагаемых очевидно, что он никогда не бывает ограниченным снизу для непостоянных функций f.

Более того, рассмотрим уравнение движения

$$\nabla_\mu (f F^{\mu\nu}) - \frac{1}{2} f' F^2 A^\nu + m^2 A^\nu = 0.$$

Поле имеет эффективную массу $\tilde{m}^2 = \frac{m^2}{f} - \frac{f'F^2}{2f}$, причём $F^2 \approx H^2 B^2$, если динамика Вселенной доминируется векторным полем. Поскольку $F^2 < 0$ и $f > 0$, то для отрицательности квадрата массы (а тем более для его близости к $-2H^2$) требуется $f' < 0$ и $|f'F^2| > 2m^2$. Очевидно, что квадратичное действие для продольной моды

$$\mathcal{L} = \left(-\frac{1}{4} f' F^2 + \frac{1}{2} m^2\right) (\partial_\mu \lambda)(\partial^\mu \lambda)$$

в этом случае описывает дух.

В нашей работе [5*] показано, что и более сложные нелинейные функции от нескольких аргументов не приводят к желаемому результату. Конкретные рассуждения нагоняют тоску, не вознаграждая при этом приятными сюрпризами, поэтому мы их опускаем.

Замечания о нарушенной лоренц-инвариантности

Разумеется, все сложности, порождаемые духовыми степенями свободы, так или иначе связаны с лоренц-инвариантностью. Если от неё отказаться, то можно вообще рассмотреть модель только с пространственными компонентами $\mathcal{L} = \frac{1}{2}\left(\partial_\mu A_i\right)\left(\partial^\mu A_i\right) - \frac{1}{2}m^2 A_i^2$ (в космологической системе отсчета). Конечно, этого не хотелось бы делать, поскольку лоренц-инвариантность на фундаментальном уровне проверена очень хорошо.

Однако нарушение инвариантности не обязано быть фундаментальным, оно может быть спонтанным, за счёт взаимодействия с каким-нибудь нетривиальным фоном (например, эйнштейновский эфир). Эти возможности остаются за рамками нашего рассмотрения, но хотелось бы заметить, что даже при отказе от лоренц-инвариантности, если этот отказ не столь радикален как полное исключение временной компоненты, избавиться от неустойчивостей не так то просто.

В самом деле, простейшей идеей может стать введение разных масс для разных компонент поля:

$$V(A^2) = -(m_1^2 A_0^2 - m_2^2 A_i^2)$$

без изменения максвелловского кинетического слагаемого. Такое отличие непросто ввести ввести во время инфляции взаимодействием с тензором Риччи, ибо он почти пропорционален метрике – потребуются коэффициенты порядка $\mathcal{O}\left(\frac{1}{\epsilon}\right)$, но эйнштейновский эфир мог бы работать.

В этом случае временная компонента определяется из уравнения

$$\left(-\triangle + m_1^2\right) A_0 + \partial_i \dot{A}_i = 0$$

без проблем с нулевыми модами, если её масса имеет нормальный знак. Однако в пространственной части уравнения движения

$$\ddot{A}_i + m_2^2 A_i - \triangle A_i + \left(1 - \frac{m_2^2}{m_1^2}\right)\partial_i \partial_k A_k = 0$$

обнаруживается градиентная неустойчивость продольной моды (что, конечно, сразу очевидно в картине Штюкельберга) в случае тахионного характера второй массы. В принципе, эту проблему можно решать вмешательством в кинетическое слагаемое, но мы оставляем такие возможности за кадром.

2.3 Краткий обзор современного состояния инфляции с векторными полями

Наша статья [1*] породила очень большой отклик в сообществе. Возник интерес к построению моделей инфляции с необычными для этой цели полями, вплоть до спиноров [81, 82]. Спинорные модели, конечно, весьма плохо мотивированы, поскольку трудно объяснить физический смысл макроскопически (и даже астрономически) большого классического спинорного поля. Однако другие варианты могут быть интересны, как с позиций построения новых моделей инфляции, так и с фундаментальных точек зрения, таких как организация устойчивого анизотропного расширения, что, как известно, нетривиально [83].

Одно из интересных обобщений – инфляция на дифференциальных p-формах при $p > 1$ (антисимметричных тензорах) [84, 85]. Так, например, для 2-форм имеем антисимметричный тензорный потенциал с двумя значками $A_{\mu\nu}$, а напряжённость поля имеет три значка $F_{\mu\nu\alpha}$. При этом неминимальное взаимодействие с гравитацией может содержать как слагаемые вида $RA_{\mu\nu}A^{\mu\nu}$, так и $R_{\mu\nu}A^{\mu}{}_{\alpha}A^{\alpha\nu}$.

Оказывается, что (при аккуратном подборе коэффициентов во взаимодействии с кривизной) существует дуальность [84,86]: 2-формы дуальны векторам, а 3-формы – скалярам. Надо кстати отметить, что и для векторной инфляции можно было бы ввести взаимодействие не только со скаляром кривизны, но и с тензором Риччи ($R_{\mu\nu}A^{\mu}A^{\nu}$) – это по очевидным причинам мало меняет динамику инфляционного режима.

Правда, эти преобразования дуальности содержат кривизну, а с ней и вторые производные от метрики, которые приводят к третьим производным в тензоре энергии-импульса в дуальной картине, которые исчезают вокруг однородных фоновых решений. Это является любопытным отражением найденной нами дополнительной степени свободы.

Инфляция на 2-формах подвержена всем сложностям, что и векторная: духи [84] и катастрофический рост анизотропии [87]. Интересно однако, что, если не пытаться абсолютно точно воспроизвести фоновую динамику скаляра (позволить поправки к эффективной массе порядка $\dot{H}$), то инфляцию на 3-формах можно иметь без неминимального взаимодействия с гравитацией и духовых степеней свободы [88, 89].

Ограничиваясь только векторными полями, существует другая интересная возможность, свободная от типичных проблем векторной инфляции, но она скорее содержит векторную примесь к скалярной инфляции (что само по себе интересно как успешное построение расширения со стабильной анизотропией) [90–92]. Плотность энергии доминируется скалярным инфлатоном, но векторное поле не убывает за счёт кинетического взаимодействия вида $f(\phi)F_{\mu\nu}F^{\mu\nu}$. Такие модели продолжают активно исследоваться [93]. Важной особенностью является возможность сохранения калибровочной инвариантности, что позволяет избежать появления любых проблем с продольными модами.

Другое интересное направление [94–96] – модели инфляции с неабелевыми полями, которые используют добавку вида $\left(F\tilde{F}\right)^2$ к действию Янга-Миллса. Калибровочная инвариантность сохранена, поэтому такие модели также не подвержены проблемам с продольными компонентами векторных полей. Изотропную триаду обычно записывают, преполагая калибровочную группу $SU(2)$, – один генератор из базиса алгебры Ли на одно пространственное направление. Естественность такого подхода может (справедливо) вызывать вопросы. Однако, это даёт возможность получать интересные результаты, и модель продолжает активно исследоваться [97].

В последнее время также интерес представляют "векторные галилеоны" [98] – векторные поля, обладающие уравнениями движения второго порядка, несмотря на старшие производные в действии, не сводящиеся к поверхностным слагаемым. Этот вопрос остаётся пока недостаточно разработанным.

2.4 О гиперболичности уравнений движения

Поскольку рассмотрение обобщённых вариантов векторной инфляции и родственных ей моделей предполагает использование нестандартных лагранжианов векторных полей, интересно выяснить более детально их динамические свойства. В работе [80] было указано, что для лагранжианов вида

$$\mathcal{L} = -f(F^2) - V(A^2) \tag{2.26}$$

с нетривиальной функцией f всегда наступает нарушение гиперболичности (чуть ниже мы укажем в каком смысле) уравнения движения

$$\nabla_\mu \left(f'(F^2) \cdot F^{\mu\nu} \right) = \frac{1}{2} V' \cdot A^\nu \tag{2.27}$$

хотя бы где-нибудь в фазовом пространстве теории.

Обсудим этот вопрос подробнее. Легко найти главную (второго порядка) часть дифференциального оператора в уравнении (2.27):

$$\mathfrak{D}^\nu_\mu A^\mu \equiv \left[f' \cdot \left(\delta^\nu_\mu \Box - \partial_\mu \partial^\nu \right) + 4 f'' \cdot F^{\alpha\nu} F_{\beta\mu} \partial_\alpha \partial^\beta \right] A^\mu . \tag{2.28}$$

Гиперболичность системы уравнений можно обсуждать на языке спектральных свойств главного символа ее дифференциального оператора, в нашем случае

$$D^\nu_\mu(p) \equiv (M^\nu_\mu)^{\alpha\beta} p_\alpha p_\beta \equiv \left[f' \cdot \left(\delta^\nu_\mu g^{\alpha\beta} - \delta^\alpha_\mu g^{\nu\beta} \right) + 4 f'' \cdot F^{\alpha\nu} F^\beta{}_\mu \right] p_\alpha p_\beta . \tag{2.29}$$

Если удаётся диагонализовать $(M^\nu_\mu)^{\alpha\beta}$ по отношению к индексам μ и ν, $(M^\nu_\mu)^{\alpha\beta} \to G^{\alpha\beta}_{(\mu=\nu)} \cdot \delta^\nu_\mu$, то каждая из матриц $G^{\alpha\beta}$ должна иметь одно отрицательное и три положительных собственных значений.

Строго говоря, в нашем случае $\mathfrak{D}_\mu^\nu$ заведомо не гиперболичен, поскольку уравнения движения содержат хорошо известную связь

$$\nabla_\mu \left(V' \cdot A^\mu\right) = 0, \tag{2.30}$$

которая уменьшает количество независимых компонент до трёх. На самом деле мы требуем, чтобы матрицы $G^{\alpha\beta}$ для этих независимых компонент были гиперболической природы.

Кроме того, следуя работе [80], мы позволим себе ещё одну вольность – мы потребуем, чтобы матрица $g_{\nu\alpha}G^{\mu\alpha}$ была положительно определённой. Это означает, что отрицательное собственное значение матрицы $G^{\alpha\beta}$ имеет времениподобный собственный вектор (по отношению к метрике g). Это совершенно разумное требование, поскольку в конце концов нам надо эволюционировать полную систему уравнений, описывающих вселенную.

Отметим, что в работе [80] было наложено условие $\partial_\mu A^\mu = 0$, которое справедливо только для квадратичного потенциала, или как калибровочное условие в безмассовом случае. Тогда легко видеть, что

$$(M_\mu^\nu)^{\alpha\beta} = f'\delta_\mu^\nu g^{\alpha\beta} + 4f'' F^{\alpha\nu} F^\beta{}_\mu,$$

а главный символ принимает вид

$$D(p) = f' p^2 I + 4f'' \left|w\right\rangle \left\langle w\right|,$$

где I – единичная матрица, $p^2 \equiv p_\mu p^\mu$ и $w^\mu \equiv F^{\alpha\mu} p_\alpha$.

Диагонализация даёт три матрицы $G^{\alpha\beta} = f' \cdot g^{\alpha\beta}$, которые отвечают очевидно гиперболическому оператору $f'\Box$ с естественным условием $f' > 0$. Однако, четвёртая $f'p^2 + 4f'' \left\langle w|w\right\rangle$ – более интересна

$$G^{\alpha\beta} = f' \cdot g^{\alpha\beta} + 4f'' \cdot F^{\alpha\mu} F^\beta{}_\mu. \tag{2.31}$$

Соответствующая матрица $g_{\nu\alpha}G^{\mu\alpha} = f' \cdot \delta_\nu^\mu + 4f'' \cdot F^{\alpha\mu} F_{\alpha\nu}$ была изучена в работе [80], и мы не будем повторять здесь несложных выкладок. В качестве окончательного результата получается, что эта матрица диагонализуема с двумя собственными значениями кратности два:

$$\lambda = f' + f''F^2 \pm f''\sqrt{(F^2)^2 + (F\tilde{F})^2}, \tag{2.32}$$

где $\tilde{F}$ – дуальный тензор напряжённости поля. Можно убедиться, что эти собственные значения всегда можно сделать отрицательными, вопреки нашим требованиям, с помощью подходящего выбора значений полевых переменных для любой нелинейной функции $f(F^2)$.

В случае общего потенциала условием $\partial_\mu A^\mu = 0$ пользоваться нельзя, ибо настоящая связь устроена сложнее. Однако, результат остаётся справедливым. В самом деле, главный символ (2.29) имеет вид

$$D(p) = f'p^2 I - f' \left| v \right\rangle \left\langle v \right| + 4f'' \left| w \right\rangle \left\langle w \right|,$$

где $w^\mu \equiv F^{\alpha\mu}p_\alpha$, $v^\mu \equiv g^{\alpha\mu}p_\alpha$ и $\langle v|w\rangle = 0$ в силу антисимметрии $F_{\mu\nu}$. После диагонализации два оператора сохраняют тривиальный вид $f'\Box$, один оператор снова отвечает эффективной метрике (2.31) с тем же результатом для гиперболичности, а еще один равен $f'\left(p^2 - \langle v|v\rangle\right) = 0$. Последний результат, естественно, просто соответствует существованию связи (2.30) в модели.

Можно было бы надеяться на улучшение поведения уравнений после разрешения связи (2.30), которую мы не приняли во внимание, сразу рассматривая только главный символ. Но, к сожалению, это не так. В самом деле, подставляя связь (2.30) в дифференциальный оператор (2.28), получаем для главного символа

$$\begin{aligned} D^\nu_\mu(p) &= (M^\nu_\mu)^{\alpha\beta} p_\alpha p_\beta \\ &= \left[f' \cdot \left(\delta^\nu_\mu g^{\alpha\beta} + 2g^{\nu\beta}\frac{V''}{V'}A^\alpha A_\mu \right) + 4f'' \cdot F^{\alpha\nu}F^\beta{}_\mu \right] p_\alpha p_\beta. \end{aligned}$$

Теперь диагонализацию провести нетрудно. Главный символ оператора приобретает вид

$$D(p) = f'p^2 I + 2f'\frac{V''}{V'}\left(A^\mu p_\mu\right)|v\rangle\langle A| + 4f''|w\rangle\langle w|\,,$$

где w^μ и v^μ такие же, как раньше. Поскольку $\langle v|w\rangle = 0$, сопряжённая матрица имеет собственный вектор $\langle w|$ с собственным значением, отвечающим эффективной метрике (2.31). Следовательно, характер нарушения гиперболичности остаётся тем же.

2.4.1 Виды нарушений

Перейдём теперь к обсуждению конкретных видов нарушения гиперболичности в моделях с лагранжианом (2.26). Рассмотрим случай, похожий на векторную инфляцию: пространственно однородное векторное поле $A^\mu(t)$ во фридмановской вселенной $ds^2 = -dt^2 + a^2(t)d\overrightarrow{x}^2$. Напряженность поля имеет только электрическую часть

$$F_{0i} \equiv E_i(t) = \dot{A}_i(t),$$

а временная компонента векторного поля A_0, как мы видели ранее, обращается в нуль. Для эффективной метрики (2.31) получаем

$$G^{\mu\nu}\partial_\mu\partial_\nu = -\left(f' - 4f''\frac{E^2}{a^2}\right)\partial_t^2 + \frac{1}{a^2}\left(f'\delta_{ij} - 4f''\frac{E_iE_j}{a^2}\right)\partial_i\partial_j.$$

В моделях, в которых $f' > 0$ и $f'' < 0$, гиперболичность вокруг таких космологических решений никогда не нарушается. Более того, скорость распространения (для фронта ударной волны – по виду главной части волнового оператора) равна скорости света вдоль электрического поля, и субсветовая – поперёк. В самом деле, без ограничения общности можно считать, что электрическое поле направлено вдоль первой оси $E_i = E\delta_i^1$. Тогда получаем

$$G^{\mu\nu}\partial_\mu\partial_\nu = -\left(f' - 4f''\frac{E^2}{a^2}\right)\partial_t^2 + \frac{1}{a^2}\left(f' - 4f''\frac{E^2}{a^2}\right)\partial_1^2 + \frac{1}{a^2}f'\partial_2^2 + \frac{1}{a^2}f'\partial_3^2.$$

Напротив, если $f'' > 0$ (при $f' > 0$), то скорости вдоль осей x^2 и x^3 – сверхсветовые, равные $\sqrt{\frac{1}{1-4\frac{f''}{f'}\frac{E^2}{a^2}}}$ в единицах скорости света. Отметим, что собственные значения (2.32) к этому нечувствительны. Но как только поля становятся слишком большими, $\frac{E^2}{a^2} > \frac{f'}{4f''}$, временная координата и пространственная координата вдоль электрического поля меняются местами, приводя к двум отрицательным собственным значениям матрицы $g_{\nu\alpha}G^{\mu\alpha}$. Разумеется, к тому же выводу можно было придти, просто рассматривая формулу (2.32) при $F\tilde{F} = 0$ и $F^2 = -2\frac{E^2}{a^2}$.

Рассмотрим теперь чисто магнитное поле вдоль оси x^2: $F_{31} = B$, соответственно $A_1 = \frac{B}{2}x^3$ и $A_3 = -\frac{B}{2}x^1$ с некоторой константой B. В таком случае легко видеть, что

$$G^{\mu\nu}\partial_\mu\partial_\nu = -f'\partial_t^2 + \frac{1}{a^2}\left(f' + 4f''\frac{B^2}{a^2}\right)\partial_1^2 + \frac{1}{a^2}f'\partial_2^2 + \frac{1}{a^2}\left(f' + 4f''\frac{B^2}{a^2}\right)\partial_3^2.$$

При $f' > 0$ и $f'' > 0$ каузальных патологий в смысле работы [80] не обнаруживается, хотя скорости сверхсветовые в направлениях, поперечных к магнитному полю ($\sqrt{1 + 4\frac{f''}{f'}\frac{B^2}{a^2}}$ в единицах скорости света).

Если же положить $f'' < 0$, то патологий не возникает до тех пор, пока $\frac{B^2}{a^2} < -\frac{f'}{4f''}$, а скорости распространения при этом субсветовые. При больших магнитных полях гиперболичность нарушается, или в более привычных терминах: система оказывается гиперболичной по отношению к координате x^2 в роли переменной времени.

Наконец, можно исследовать скрещенные конфигурации электрического и магнитного полей. Положим $F_{01} = E$ и $F_{31} = B$, а в качестве метрики возьмём для простоты пространство Минковского, тогда в волновом операторе получаем (2.31):

$$\begin{aligned} G^{\mu\nu}\partial_\mu\partial_\nu = &-\left(f' - 4f''E^2\right)\partial_t^2 - 8f''EB\partial_t\partial_3 + \left(f' + 4f''B^2\right)\partial_3^2 \\ &+ \left(f' + 4f''(B^2 - E^2)\right)\partial_1^2 + f'\partial_2^2. \end{aligned}$$

При любом знаке f'' вдоль направления x^1 могут возникать проблемы с каузальностью, что подтверждает общий вывод работы [80].

2.4.2 Космологические векторные поля

Рассмотрим теперь выводы, которые можно сделать для космологии на основе картины нарушения гиперболичности для векторных полей. Будем предполагать, что потенциальная энергия подавлена по сравнению с кинетической (можно даже думать о чистой $\mathcal{L} = -f(F^2)$ модели, но нам не хочется обсуждать проблемы калибровочной инвариантности). Будем рассматривать динамику пробного пространственно однородного векторного поля в расширяющейся Вселенной.

Следуя логике работы [80], будем для начала искать режимы с медленно меняющимся электрическим полем $\frac{E^2}{a^2}$. Для этого требуется $E \sim a$ или $\dot{E} \approx HE$, в то время как уравнение движения (2.27) в пренебрежении массой дает

$$\left(1 - 4\frac{E^2 f''}{a^2 f'}\right)\dot{E} + H\left(1 + 4\frac{E^2 f''}{a^2 f'}\right)E \approx 0. \tag{2.33}$$

Ясно, что нам необходимо, чтобы отношение $\frac{4\frac{E^2 f''}{a^2 f'}+1}{4\frac{E^2 f''}{a^2 f'}-1}$ было близко к единице, а следовательно $4\frac{E^2|f''|}{a^2 f'} \gg 1$.

Мы уже видели, что при $f'' > 0$ в случае таких больших полей возникает нарушение гиперболичности (а ещё раньше появляются сверхсветовые скорости распространения).

Случай $f'' < 0$ выглядит более обнадёживающе, хотя и надо заботиться от том, чтобы не перейти в область $f' < 0$ (мы считаем поля уменьшающимися, но $F^2 = -2\frac{E^2}{a^2}$ растёт у уменьшением $\frac{E}{a}$). Рассмотрим, например, функцию $f(F^2) = 1 - e^{-\alpha F^2}$ с некоторой постоянной $\alpha > 0$. В таком случае всегда $f' = \alpha e^{-\alpha F^2} > 0$ и $f'' = -\alpha^2 e^{-\alpha F^2} < 0$. Из уравнения (2.33) получаем $\dot{E} = \frac{4\alpha\frac{E^2}{a^2}-1}{4\alpha\frac{E^2}{a^2}+1}\cdot HE$, откуда следует $\dot{E} \approx HE$ если $4\alpha\frac{E^2}{a^2} \gg 1$.

Казалось бы, возможен режим медленного качения. Но, именно поскольку функция выбрана чрезвычайно крутой, это не совсем то, что нам нужно. Можно убедиться, что решение имеет вид $\frac{E^2}{a^2}e^{4\alpha\frac{E^2}{a^2}} = \frac{const}{a^4}$ и отвечает медленному качению электрического поля только за счет экспоненициально сильной зависимости в левой части. Тензор энергии-импульса доминируется слагаемым $\alpha F^{\mu\beta}F_{\nu\beta}e^{-\alpha F^2}$, которое быстро убывает (хотя и с сохранением относительной анизотропии). Соответственно, характерная плотность энергии убывает, и такое поле не может быть источником инфляции.

Соответственно, этот пример учит нас, что может быть лучше искать функции f с большой второй производной f'', но малой f', для которых нет такого усиления качения f в сравнении с аргументом $\frac{E}{a}$. Ясно однако, что малость первой производной при большой второй требует еще большей подстройки параметров. К тому же, заведомо опасная область $f' < 0$ оказывается совсем рядом.

Разумеется, остаётся ещё возможность вообще не гнаться за медленностью изменения электрического поля, а просто рассмотреть предельно (экспоненциально) плоскую функцию f, которая будет давать практически постоянную плостность энергии несмотря на быстрый бег аргумента. Однако это мало отличается от простой космологической постоянной, когда в тензоре энергии-импульса доминирует член $fg_{\mu\nu}$.

Можно подтвердить выводы работы [80] о том, что жизнеспособные инфляционные режимы возможны только в результате тонкой подстройки параметров или выбора экспоненциально плоского потенциала, хотя последний вариант, с нашей точки зрения, не обязательно должен считаться неестественным. Однако рассмотрение в терминах невозможности медленного качения $\frac{A_i}{a}$ и $\frac{F_{0i}}{a}$ не вполне точно для моделей с очень сильной зависимостью функции f от аргумента, поскольку медленное качение данных величин оказывается не тем условием, которое обеспечивает правильное уравнение состояния.

* * *

Снова отметим, что в последнее время также стало интересно рассматривать поля типа векторных галилеонов [98]. Примечательно, что наши условия гиперболичности очень похожи на условия устойчивости в работе [98] с заменой фоновых конфигураций векторного поля на геометрические величины.

Глава 3

Массивная гравитация

В этой главе исследуются особенности теорий массивной гравитации, результаты опубликованы в работах [7*,8*,9*,10*]. В литературе существуют прекрасные обзоры по данной теме: написанный на заре дРГТ теорий общий обзор Курта Хинтербихлера [99], подробный обзор по дРГТ теориям и их приложениям, написанный с точки зрения эффективной теории поля, от Клаудии де Рам [100], а также обзор [101] по биметрическому варианту теории с упором на гамильтонов формализм. Поэтому мы ограничимся лишь теми аспектами, которые важны для наших собственных работ [7*,8*,9*,10*]. В частности, мы полностью игнорируем тетрадную формулировку [102], которая является очень важной составной частью теории.

История массивной гравитации восходит к классической статье Фирца и Паули 1939 года [103], в которой, в рамках линеаризованной теории, была найдена форма потенциала, обеспечивающая наличие только пяти степеней свободы (за счёт пары связей второго рода). Через некоторое время (в 1970 году) было замечено [104,105], что в теории нет непрерывности предела при стремлении массы гравитона к нулю: скалярная мода не отщепляется при $m \to 0$, увеличивая притяжение массивных тел, но не меняя отклонения лучей света, что противоречит экспериментальным данным.

Однако же, очень скоро было выяснено [106], что чем меньше масса, тем больше область, в которой скалярная мода находится в режиме сильной связи (при учёте нелинейных поправок в действии Эйнштейна-

Гильберта), и можно рассчитывать на восстановление непрерывности предела $m \to 0$ в полной нелинейной теории – это механизм Вайнштейна. С современной точки зрения можно рассматривать его как один из возможных способов "спрятать" скалярную моду от детектирования в модифицированной теории гравитации, наряду с хамелеоном [107] и симметроном [108]. Отличие механизма Вайнштейна в том, что он использует нелинейность в кинетической части действия. Разумеется, это весьма нетривиальный механизм, полностью доказать наличие которого непросто, см. современный обзор [109]. Это одна из очень серьёзных проблем, связанных с применением (лоренц-инвариантной) массивной гравитации к реальным ситуациям, особенно с точки зрения квантования.

По иронии судьбы, в том же самом году (1972), когда был предложен механизм Вайнштейна, Боулваром и Дезером было показано, что при учёте нелинейных поправок духовая степень свободы неизбежно возвращается обратно в теорию [110], и исследования по массивной гравитации на долгие годы утратили свою привлекательность.

Прорыв произошёл в 2010 году, когда Клаудия де Рам и Грегори Габададзе обнаружили [111] пробел в рассуждениях [110, 112], устанавливающих неизбежность духа. Нелинейные поправки строились по теории возмущений, но авторам удалось, совместно с Эндрю Толли, обнаружить [113], что они складываются в ряд Тейлора для квадратного корня из некоторой матрицы, см. ниже раздел 3.2. Разумеется, дать общее доказательство отсутствия духа во всех порядках теории возмущений было весьма непросто, см. статьи [113, 114].

В серии работ [115–119] Фавад Хассан, Рейчел Розен и другие дали общее непертурбативное доказательство отсутствия духа в дРГТ теории, а также обобщили этот результат на случай произвольной опорной метрики (вместо метрики Минковского) и на биметрический случай (динамическая опорная метрика). Это вызвало мощный всплеск исследований по теории массивной гравитации, включая космологические решения, Чёрные Дыры и т.д., см. обзоры [100, 101].

Наша работа [7*] была написана на заре исследований по гамильтонову анализу массивной гравитации, и в ней был предложен новый метод проведения анализа, со вспомогательными полями, но без явного извлечения квадратного корня из матрицы, который оказался полезен [120] для рассмотрения проблемы на языке полей Штюкельберга для нарушенных диффеоморфизмов.

Работы [8*,9*] посвящены важному аспекту теории – проблеме существования и единственности квадратного корня из матрицы. Эти вопросы тесно переплетены со взаимоотношениями тетрадной и метрической формулировок [121,122], с которыми пока ещё связано много неясностей и которые выходят за рамки данной Диссертации.

Возможности использования необычных корней упоминались, в весьма частных случаях приложений к космологии, в статьях [123–125], но сколько-нибудь общего рассмотрения до наших работ предложено не было. Нами была полностью описана имеющаяся здесь неоднозначность, а также предложен вариант теории возмущений (в терминах собственных значений вместо матриц), который позволяет рассматривать такие ситуации, когда стандартные методы терпят полное фиаско в силу плохой поставновки задачи об извлечении матричного квадратного корня, см. раздел 3.6.

Наконец, в статье [10*] мы обнаруживаем наличие духа в одном из расширений массивной гравитации – в расширенном квазидилатоне. Про мотивы для рассмотрения этих теорий и актуальность данного исследования см. разделы 3.7 и 3.8.

Отметим, что разделы 3.1, 3.2 и 3.7 являются обзорными и не содержат результатов, выносимых на защиту. В них приведены необходимые сведения для понимания смысла наших работ и их места в текущей литературе. Раздел 3.4 также не является новым, ибо содержит математическое рассмотрение проблемы извлечения квадратных корней из матриц. Наши результаты приведены в разделах 3.3, 3.5, 3.6 и 3.8.

3.1 Теория Фирца-Паули и её проблемы

Особенность массивной гравитации заключается в том, что есть несоответствие ожидаемого числа степеней свободы после разрушения калибровочной инвариантности общей теории относительности $(2+4=6)$ и размерности представления группы Лоренца, отвечающего массивному спину 2 (5). Печальная новость заключается в том, что лишняя скалярная мода оказывается духом.

Опасность придания гравитону массы довольно таки ясна уже на уровне линейной теории. В самом деле, рассмотрим линеаризованное действие Эйнштейна-Гильберта (сигнатура $(-,+,+,+)$)

$$S=\int d^4x\left(-\frac{1}{4}(\partial_\alpha h_{\mu\nu})(\partial^\alpha h^{\mu\nu})+\frac{1}{2}(\partial^\alpha h_{\mu\nu})(\partial^\nu h^\mu_\alpha)\right.$$
$$\left.-\frac{1}{2}(\partial_\alpha h^{\alpha\mu})(\partial_\mu h^\beta_\beta)+\frac{1}{4}(\partial_\mu h^\alpha_\alpha)(\partial^\mu h^\beta_\beta)+\mathcal{O}(h^3)\right), \quad (3.1)$$

в стандартных переменных космологической теории возмущений (см. первую главу) оно принимает вид

$$S=\int d^4x\left(-\frac{1}{4}(\partial_\alpha h^{(TT)}_{ij})(\partial^\alpha h^{(TT)}_{ij})+\frac{1}{2}\left(\partial_j\left(\dot{v}_i-s_i\right)\right)^2\right.$$
$$\left.-6\dot{\psi}^2+2(\partial_i\psi)^2+4\psi\,\triangle\left(\phi-\dot{b}+\ddot{\sigma}\right)+\mathcal{O}(h^3)\right), \quad (3.2)$$

где, как обычно,

$$h_{00}=2\phi,$$

$$h_{0i}=\partial_i b+s_i,$$

$$h_{ij}=2\psi\delta_{ij}+2\partial^2_{ij}\sigma+\partial_i v_j+\partial_j v_i+h^{(TT)}_{ij}$$

при $\partial_i s_i\equiv 0,\quad \partial_i v_i\equiv 0,\quad \partial_i h^{(TT)}_{ij}\equiv 0,\quad h^{(TT)}_{ii}\equiv 0.$

Как видим, в секторе гравитационных волн всё в порядке, они подчиняются безмассовому волновому уравнению.

Два поперечных вектора отвечают четырём независимым параметрам, но только два из них, дающиеся вектором $\dot{v}_i-s_i$, калибровочно инвариантны.

Наконец, у скаляров тоже есть две калибровочно инвариантные комбинации, ψ и $\phi-\dot{b}+\ddot{\sigma}$. Разумеется, все эти переменные получаются из космологических инвариантных потенциалов при стремлении константы Хаббла к нулю, см. раздел 1.2.6.

У величин s_i и ϕ нет производных по времени в действии, а производная от b может быть устранена интегрированием по частям. Таким образом, эти четыре переменные не являются динамическими. И, более того, они служат множителями Лагранжа, накладывающими четыре связи (первого рода, отвечающие калибровочной инвариантности).

Легко видеть, что тем самым скалярный и векторный сектора лишаются динамических гравитационных степеней свободы и полностью определяются связями. Очевидно, что это обстоятельство спасительно для теории, поскольку в скалярном секторе действия (3.2) мы видим неправильный знак кинетической энергии. Следовательно, вводя массу гравитона, которая гарантированно разрушит калибровочную симметрию, следует убедиться, что мы не возвращаем дух в теорию.

Вообщя говоря, можно было бы думать о двух способах добавить массовое слагаемое к действию (3.1), $h_{\mu\nu}h^{\mu\nu}$ или $h^\mu_\mu h^\nu_\nu$. Заметим однако, что избавиться от неправильной кинетической энергии ψ удавалось благодаря тому, что поле ϕ (по сути – h_{00}) входило в действие (3.2) линейно. Попытка сохранить это свойство приводит нас к слагаемому Фирца-Паули:

$$V = \frac{m^2}{4}\left(h_{\mu\nu}h^{\mu\nu} - h^\mu_\mu h^\nu_\nu\right). \tag{3.3}$$

При этом ϕ остается множителем Лагранжа квадратичного действия и порождает связь, а точнее – пару связей второго рода.

* * *

Другая поучительная точка зрения на массовый член Фирца-Паули заключается в использовании трюка Штюкельберга. В линейном приближении калибровочное преобразование метрики при замене координат $x^\mu \longrightarrow x^\mu + \xi^\mu(x)$ выглядит следующим образом:

$$h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu.$$

При подстановке этой замены в массовое слагаемое (3.3) полученное выражение обладает тем замечательным свойством, что лагранжиан штюкельбергова поля совпадает с максвелловским

$$(\partial_\mu \xi_\nu - \partial_\nu \xi_\mu)(\partial^\mu \xi^\nu - \partial^\nu \xi^\mu),$$

единственно возможным, свободным от духа. Если бы мы имели другое массовое слагаемое, то при выделении уже из векторного поля продольной компоненты с помощью нового трюка Штюкельберга, $\xi_\mu \to \xi_\mu + \partial_\mu \zeta$, последняя получила бы старшие производные в действии – дух Остроградского.

3.2 Нелинейная бездуховая массивная гравитация

Как мы уже обсуждали выше, необходимо выйти за рамки линейного приближения, хотя бы для того, чтобы не противоречить экспериментам по проверке общей теории относительности. В принципе, можно пойти по прямолинейному пути и искать поправки следующего порядка. А именно, если мы учли, например, третий порядок в действии Эйнштейна-Гильберта, какие поправки надо добавить к слагаемому Фирца-Паули? Так можно поступить в каждом порядке теории возмущений и получить целый ряд для потенциала в полностью нелинейной теории массивной гравитации [111]. Отметим, во избежание недоразумений, что при работе на языке полей Штюкельберга придётся выйти за рамки линейного приближения при записи преобразования метрики.

Однако, довольно рано было замечено, что ряд подозрительно похож на ряд Тейлора для квадратного корня. И действительно, оказалось, что бездуховые потенциалы представимы с помощью квадратного корня из матрицы $g^{\alpha\mu}\eta_{\mu\beta}$ [115]:

$$V_1 = \mathrm{Tr}\sqrt{g^{-1}\eta},$$

$$V_2 = \left(\mathrm{Tr}\sqrt{g^{-1}\eta}\right)^2 - \mathrm{Tr}\left(\sqrt{g^{-1}\eta}\right)^2,$$

$$V_3 = \left(\mathrm{Tr}\sqrt{g^{-1}\eta}\right)^3 - 3\left(\mathrm{Tr}\sqrt{g^{-1}\eta}\right)\left(\mathrm{Tr}\left(\sqrt{g^{-1}\eta}\right)^2\right) + 2\mathrm{Tr}\left(\sqrt{g^{-1}\eta}\right)^3.$$

Следует заметить, что у этих выражений есть замечательный математический смысл: это элементарные симметрические полиномы собственных значений матрицы $\sqrt{g^{-1}\eta}$. (На самом деле, в качестве фоновой метрики можно взять любую другую метрику $f_{\mu\nu}$ вместо метрики Минковского и иметь дело с матрицей $\sqrt{g^{-1}f}$.) В частности, можно было бы взять и четвёртый полином, но он равен детерминанту, и поэтому, будучи умножен на $\sqrt{-g}$, просто сдвигает действие на константу (пропорциональную $\sqrt{-f}$ в случае произвольной вспомогательной метрики – добавка к её космологической постоянной в биметрической теории).

Отметим, что при использовании потенциала Фирца-Паули (с почти любыми нелинейными поправками) теория возмущений перестаёт работать на масштабе $\Lambda_5 = \left(m^4 M_{Pl}\right)^{1/5}$, который соответствует классическому радиусу Вайнштейна, внутри которого скалярная мода становится сильно связанной. Это масштаб обрезания для эффективной теории массивного гравитона [112]. Потенциал массивной гравитации дРГТ получается последовательным сокращением лидирующих взаимодействий, самодействий скаляра. Можно убедиться, что он позволяет отодвинуть обрезание до масштаба $\Lambda_3 = \left(m^2 M_{Pl}\right)^{1/3}$, на котором уже важна и векторная мода [126], и выше которого, по-видимому, масштаб обрезания поднять невозможно при сохранении лоренц-инвариантности (см. однако [127]).

Интересно отметить, что хотя старшие производные и не исчезают из действия скалярной моды, они принимают вид галилеонов, обладающих уравнениями движения второго порядка. В этом можно убедиться в пределе отщепления $m \to 0$, $M_{Pl} \to \infty$, $\Lambda_3 = const$, в котором остаётся только скалярная мода [126].

Мы не будем повторять всех подробностей, хотя это и очень красивая тема, поскольку главной задачей данной главы является непертурбативный гамильтонов анализ. Для нас важно, что действие массивной гравитации можно записать в виде [115]

$$S = \int d^N x \sqrt{-g} \left(R + m^2 \sum_{n=0}^{N} \beta_n e_n \left(\sqrt{g^{-1}\eta} \right) \right), \tag{3.4}$$

где N – размерность пространства-времени с метрикой $g_{\mu\nu}$, R – скалярная кривизна данной метрики, а e_n – элементарные симметрические полиномы собственных значений той матрицы, что стоит у них в аргументе (в исходных пертурбативных работах [113] использовалась матрица $\mathbb{I} - \sqrt{g^{-1}\eta}$ вместо $\sqrt{g^{-1}\eta}$, но это неважно, поскольку симметрические полиномы можно, конечно же, легко пересчитать).

Опишем более подробно понятие симметрических полиномов. Пусть матрица $\mathcal{M}^\mu_\nu$ обладает собственными значениями λ_i. Положим

$$e_n \equiv \sum_{i_1 < i_2 < \ldots < i_n} \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_n}. \tag{3.5}$$

По определению будем считать $e_0 \equiv 1$. Тогда симметрические полиномы от собственных значений могут быть представлены (теорема Виета) как коэффициенты характеристического многочлена матрицы $\mathcal{M}$:

$$\det\left(\mathcal{M} - \lambda \mathbb{I}\right) = \prod_{n=1}^{N} (\lambda_i - \lambda) = \sum_{n=0}^{N} (-\lambda)^{N-n} \cdot e_n(\mathcal{M}). \tag{3.6}$$

Ясно, что e_0 описывает лишь космологическую постоянную, а e_N – вообще константу (или космологическую постоянную для второй метрики, если она динамическая). Оставшиеся $N-1$ параметров описывают возможные потенциалы для гравитона. В частности в четырёхмерном пространстве-времени есть трёхпараметрическое семейство бездуховых массовых слагаемых.

В дальнейшем мы будем часто пользоваться упрощённой формой записи, в которой $[\mathcal{M}]$ обозначает след матрицы $\mathcal{M}$. В соответствии с этим имеем $[\mathcal{M}] \equiv \mathcal{M}^\mu_\mu$, $[\mathcal{M}]^2 \equiv (\mathcal{M}^\mu_\mu)^2$, $[\mathcal{M}^2] \equiv \mathcal{M}^\mu_\nu \mathcal{M}^\nu_\mu$, и так далее. В этих обозначениях можно написать

$$e_1(\mathcal{M}) = \sum_i \lambda_i = [\mathcal{M}], \tag{3.7}$$

$$e_2(\mathcal{M}) = \sum_{i<j} \lambda_i\lambda_j = \frac{1}{2}\left(\left(\sum_i \lambda_i\right)^2 - \sum_i \lambda_i^2\right) = \frac{1}{2}\left([\mathcal{M}]^2 - [\mathcal{M}^2]\right), \tag{3.8}$$

$$e_3(\mathcal{M}) = \sum_{i<j<k} \lambda_i\lambda_j\lambda_k = \frac{1}{6}\left([\mathcal{M}]^3 - 3[\mathcal{M}][\mathcal{M}^2] + 2[\mathcal{M}^3]\right), \tag{3.9}$$

$$e_4(\mathcal{M}) = \frac{1}{24}\left([\mathcal{M}]^4 - 6[\mathcal{M}]^2[\mathcal{M}^2] + 3[\mathcal{M}^2]^2 + 8[\mathcal{M}][\mathcal{M}^3] - 6[\mathcal{M}^4]\right). \tag{3.10}$$

В четырёх измерениях при этом получаем

$$e_4(\mathcal{M}) = \det(\mathcal{M}),$$

а в общем случае можно доказать полезное рекуррентное соотношение

$$e_n(\mathcal{M}) = \frac{1}{n}\sum_{i=1}^{n}(-1)^{i-1}[\mathcal{M}^i]\cdot e_{n-i}(\mathcal{M}), \tag{3.11}$$

которое автоматически обеспечивает $e_n = 0$ при $n > N$ (и, разумеется, детерминант при $n = N$).

3.2.1 Предел Фирца-Паули

Убедимся, что теория дРГТ действительно даёт массовое слагаемое Фирца-Паули для малых возмущений вокруг пространства Минковского. Для этого рассмотрим флуктуации метрики $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. Договоримся опускать и поднимать индексы у h с помощью фоновой метрики η. Напомним, что при этом

$$g^{\mu\alpha}\eta_{\alpha\nu} = \delta^\mu_\nu - h^\mu_\nu + h^{\mu\alpha}h_{\alpha\nu} + \mathcal{O}(h^3). \tag{3.12}$$

Явно вычислим матричный квадратный корень $\sqrt{g^{-1}\eta}$, раскладывая в ряд Тейлора вокруг $\sqrt{\mathbb{I}} = \mathbb{I}$,

$$\sqrt{\mathbb{I} - H} = \mathbb{I} - \frac{1}{2}H - \frac{1}{8}H^2 + \mathcal{O}(H^3),$$

где $H = h - h^2 + \mathcal{O}(h^3)$. Подставляя в формулы (3.7), (3.8), (3.9), (3.10), получаем

$$\begin{aligned}
e_1(\sqrt{g^{-1}\eta}) &= 4 - \frac{1}{2}h^\mu_\mu + \frac{3}{8}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3), && (3.13)\\
e_2(\sqrt{g^{-1}\eta}) &= 6 - \frac{3}{2}h^\mu_\mu + \frac{1}{8}(h^\mu_\mu)^2 + h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3), && (3.14)\\
e_3(\sqrt{g^{-1}\eta}) &= 4 - \frac{3}{2}h^\mu_\mu + \frac{1}{4}(h^\mu_\mu)^2 + \frac{7}{8}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3), && (3.15)\\
e_4(\sqrt{g^{-1}\eta}) &= 1 - \frac{1}{2}h^\mu_\mu + \frac{1}{8}(h^\mu_\mu)^2 + \frac{1}{4}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3). && (3.16)
\end{aligned}$$

Разумеется, последнее выражение (3.16) может быть получено и другим способом: $e_4(\sqrt{g^{-1}\eta}) = \frac{1}{\sqrt{-g}}$, где

$$\sqrt{-g} = 1 + \frac{1}{2}h^\mu_\mu + \frac{1}{8}(h^\mu_\mu)^2 - \frac{1}{4}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3). \qquad (3.17)$$

В квадратичном приближении β_i-слагаемые в действии (3.4) могут быть получены умножением формул (3.13) – (3.15) на (3.17):

$$\begin{aligned}
\sqrt{-g}\cdot e_1(\sqrt{g^{-1}\eta}) &= 4 + \frac{3}{2}h^\mu_\mu + \frac{1}{4}(h^\mu_\mu)^2 - \frac{5}{8}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3), && (3.18)\\
\sqrt{-g}\cdot e_2(\sqrt{g^{-1}\eta}) &= 6 + \frac{3}{2}h^\mu_\mu + \frac{1}{8}(h^\mu_\mu)^2 - \frac{1}{2}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3), && (3.19)\\
\sqrt{-g}\cdot e_3(\sqrt{g^{-1}\eta}) &= 4 + \frac{1}{2}h^\mu_\mu - \frac{1}{8}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3), && (3.20)
\end{aligned}$$

при этом $\sqrt{-g}\cdot e_4(\sqrt{g^{-1}\eta}) = 1$ точно, и разумеется $\sqrt{-g}\cdot e_0 = \sqrt{-g}$ раскладывается в соответствии с формулой (3.17).

На данный момент структура Фирца-Паули (3.3) совсем не видна. Однако надо ещё потребовать, чтобы параметры модели были так подобраны, чтобы пространство Минковского действительно являлось решением. Линейная вариация дРГТ потенциала дает

$$V(h) \equiv m^2 \sum_{n=0}^{N} \sqrt{-g} \cdot \beta_n e_n(\sqrt{g^{-1}\eta})$$
$$= V(0) + m^2 \left(\frac{1}{2}\beta_0 + \frac{3}{2}\beta_1 + \frac{3}{2}\beta_2 + \frac{1}{2}\beta_3\right) h^\mu_\mu + \mathcal{O}(h^2),$$

Соответственно, необходимо наложить условие

$$\beta_0 = -3\beta_1 - 3\beta_2 - \beta_3.$$

После подстановки этого требования в квадратичное действие, получаем знакомый результат:

$$V(h) - V(0) = \frac{m^2}{8}(\beta_1 + 2\beta_2 + \beta_3) \cdot \left(h^{\mu\nu}h_{\mu\nu} - (h^\mu_\mu)^2\right) + \mathcal{O}(h^3).$$

Полезное замечание

Выше мы проделали вычисления в стандартном виде. Однако полезно иметь в виду, что при работе с массивной (не полностью биметрической) гравитацией бывает удобно воспользоваться следующим трюком. Заметим, что симметрические полиномы обратной матрици $e_n(\mathcal{M}^{-1})$ – это полиномы по переменным $\frac{1}{\lambda_i}$, которые очевидно можно получить делением $e_{N-n}(\mathcal{M})$ на $\det\mathcal{M}$. Поэтому

$$\sqrt{-g} \cdot e_n(\sqrt{g^{-1}f}) = \sqrt{-f} \cdot e_{N-n}(\sqrt{f^{-1}g}),$$

что, в частности, наводит на правильную мысль о том, что опорную метрику можно сделать динамической, записав для неё тоже слагаемое Эйнштейна-Гильберта и не возраждая при этом духа Боулвара-Дезера в теории [118]. Также из этого следует хорошо известная симметрия биметрической дРГТ теории

$$g_{\mu\nu} \leftrightarrow f_{\mu\nu}, \quad \beta_n \leftrightarrow \beta_{N-n}.$$

Соответственно, беря метрику Минковского в качестве f, можно вычислять напрямую $\sqrt{-g}\beta_n e_n(\sqrt{g^{-1}\eta})$ слагаемые действия просто как $\beta_n e_{N-n}(\sqrt{\eta^{-1}g})$. В частности,

$$\sqrt{-g}\cdot e_3(\sqrt{g^{-1}\eta}) = e_1(\sqrt{\eta^{-1}g}) = e_1(\sqrt{\mathbb{I}+h}) = 4 + \frac{1}{2}[h] - \frac{1}{8}[h^2] + \mathcal{O}(h^3),$$

что, между прочим, объясняет загадочное выпадение слагаемого с $(h^\mu_\mu)^2$ из выражения для $\sqrt{-g}\cdot e_3(\sqrt{g^{-1}\eta})$.

3.2.2 Доказательство Хассана и Розен

Отсутствие духа в полученной теории (конечно же, на классическом уровне) можно также доказать и за рамками теории возмущений. Лучше всего для этого подходит гамильтонов анализ.

Для действия теории относительности в форме АДМ с добавлением потенциала V можно записать гамильтониан

$$\begin{aligned} H &= -\int d^3x\sqrt{\gamma}\left(N\left(\overset{(3)}{R} + \frac{1}{\gamma}\left(\frac{1}{2}\left(\pi^j_j\right)^2 - \pi_{ik}\pi^{ik}\right) - V\right) + 2N^i \overset{(3)}{\nabla}{}^k \pi_{ik}\right) \\ &= \int d^3x\left(\sqrt{\gamma}N\cdot V(N, N_i, \gamma_{ij}) - NR_0 - N^i R_i\right), \end{aligned} \tag{3.21}$$

где импульсы введены стандартным образом, $\pi^{ij} \equiv \frac{\partial\mathcal{L}}{\partial\gamma_{ij}}$. Как видим, в отсутствие потенциала переменные шага и сдвига входят линейно, играя роль множителей Лагранжа при стандартных связях общей теории относительности (R_0 и R_i).

Однако, в случае потенциальной функции V общего положения, шаг и сдвиг входят в лагранжиан и гамильтониан нелинейно, так что вариация по ним даёт уравнения для определения их самих вместо связей на пространственную метрику γ_{ij}. Лишённая ограничений, последняя соответствует шести поляризациям массивного гравитона, включающим также и дух Боулвара-Дезера.

Непертурбативное доказательство было дано Хассаном и Розен в серии работ, начиная со статьи [116]. Они показали, что для определяющей теорию матрицы

$$g^{-1}\eta = \frac{1}{N^2}\begin{pmatrix} 1 & N^l\delta_{lj} \\ -N^j & (N^2\gamma^{il} - N^iN^l)\delta_{lj} \end{pmatrix}$$

можно, после подходящей замены переменных сдвига, представить квадратный корень в виде

$$\sqrt{g^{-1}f} = \frac{1}{N\sqrt{1-n^kn^k}}\begin{pmatrix} 1 & n^i \\ -n^j & -n^in^j \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & X^{ij}(\gamma, n) \end{pmatrix}, \qquad (3.22)$$

где n_i – новые переменные сдвига. В таком случае, умножая β_1 потенциал на $\sqrt{-g} = N\sqrt{\gamma}$, мы видим, что переменная шага N входит в действие линейно, и значит снова накладывает связь на физический сектор (после этого для следующих потенциалов, β_2 и β_3, легко проверить, что старшие степени $\frac{1}{N}$ в них сокращаются).

Приведём некоторые подробности доказательства Хассана и Розен. Они ищут преобразование сдвигов в виде

$$N^i = (\delta^i_j + ND^i{}_j(n,\gamma)) \cdot n^j, \qquad (3.23)$$

требуя выполнения свойства

$$N\sqrt{g^{-1}\eta} = \mathbb{A} + N\mathbb{B}$$

с некоторыми матрицами $\mathbb{A}$ и $\mathbb{B}$, не зависящими от N. Довольно прямолинейные выкладки показывают, что эти матрицы можно получить в виде

$$\mathbb{A}^i{}_k = \frac{1}{\sqrt{1-n^r\delta_{rs}n^s}}\begin{pmatrix} 1 & n^r\delta_{rk} \\ -n^i & -n^in^r\delta_{rk} \end{pmatrix},$$

$$\mathbb{B} = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{(\gamma^{is} - D^i{}_rn^rn^lD_l{}^s)\delta_{sk}} \end{pmatrix}$$

при условии, что матрица преобразования D удовлетворяет уравнению

$$\left(\sqrt{1-n^r\delta_{rs}n^s}\right)D^i{}_k = \sqrt{(\gamma^{is} - D^i{}_rn^rn^lD_l{}^s)\,\delta_{sk}},$$

которое заведомо разрешимо.

Уравнение движения для переменных сдвига n^i принимает вид

$$\left(R_i + \frac{2m^2\sqrt{\gamma}n^l\delta_{li}}{\sqrt{1-n^r\delta_{rs}n^s}}\right)\cdot\left(\delta^i_{\ k} + N\frac{\partial}{\partial n^k}(D^i_{\ j}n^j)\right) = 0,$$

где второй множитель является якобианом преобразования (3.23) и не может обращаться в нуль, а обращение в нуль первого множителя оказывается уравнением, которое позволяет однозначно найти вектор сдвига:

$$n^i = -R_j\delta^{ji}\left[4m^4\mathrm{det}\gamma + R_k\delta^{kl}R_l\right]^{-1/2}.$$

После этого вариация по отношению к переменной шага N дает

$$R^0 + R_iD^i_{\ j}n^j + 2m^2\sqrt{\gamma}\left[3 - \sqrt{1-n^r\delta_{rs}n^s}D^k_{\ k}\right] = 0.$$

После подстановки в полученное уравнение найденных значений компонент сдвига, получаем связь на компоненты пространственной метрики. Поскольку калибровочная свобода полностью нарушена, это не может быть связью первого рода. И действительно, можно показать [119], что коммутация этой связи с гамильтонианом порождает вторичную связь, и вместе эти связи образуют пару связей второго рода, а дальнейшая проверка самосогласованности определяет значение переменной шага. Тем самым гамильтонов анализ завершается, и получается система с пятью степенями свободы. Доказательство может быть обобщено на случай произвольной опорной метрики [117], а также для биметрических теорий [118].

3.2.3 Биметрическая теория и уравнения движения

Как мы уже упоминали, есть замечательная симметрия между симметрическими полиномами для заданной матрицы и для её обратной, поэтому бездуховый потенциал, записанный для метрики g с произвольной опорной метрикой f, выглядит точно также, как бездуховый потенциал, записанный для метрики f с произвольной опорной метрикой g. В результате оказывается, что можно записать полностью биметрическую теорию, в которой обе метрики будут динамическими [118]:

$$
\begin{aligned}
S = &-\frac{M_g^2}{2}\int d^4x\sqrt{-g}R_g - \frac{M_f^2}{2}\int d^4x\sqrt{-f}R_f \\
&+ m^2M_g^2\int d^4x\sqrt{-g}\sum_{n=0}^{4}\beta_n e_n\left(\sqrt{g^{-1}f}\right) + \int d^4x\sqrt{-g}\mathcal{L}_m(g,\Phi),
\end{aligned}
\quad (3.24)
$$

где мы ввели лагранжиан (абстрактных) полей материи Φ, взаимодействующих с одной из метрик.

Оказывается, если некоторое поле взаимодействует сразу с двумя метриками, это неизбежно возвращает дух Боулвара-Дезера. Впрочем, известна особая комбинация метрик, для которой этот дух появляется выше обрезания эффективной теории [128, 129].

Отметим, что мы ввели две независимые массы Планка (в коэффициентах при слагаемых Эйнштейна-Гильберта), и биметрическая симметрия имеет теперь вид

$$
g \leftrightarrow f; \qquad \beta_n \leftrightarrow \beta_{4-n}; \qquad M_g \leftrightarrow M_f; \qquad m^2 \leftrightarrow m^2\frac{M_g^2}{M_f^2}. \qquad (3.25)
$$

Для вывода уравнений движения удобно заметить, что из

$$
\delta\mathcal{M} = \delta\left(\sqrt{\mathcal{M}}\cdot\sqrt{\mathcal{M}}\right) = \left(\delta\sqrt{\mathcal{M}}\right)\cdot\sqrt{\mathcal{M}} + \sqrt{\mathcal{M}}\cdot\left(\delta\sqrt{\mathcal{M}}\right)
$$

следует

$$
\sqrt{\mathcal{M}}^{-1}\cdot\delta\mathcal{M} = \sqrt{\mathcal{M}}^{-1}\cdot\left(\delta\sqrt{\mathcal{M}}\right)\cdot\sqrt{\mathcal{M}} + \delta\sqrt{\mathcal{M}}.
$$

Беря след от этого выражения, получаем $\left[\delta\sqrt{\mathcal{M}}\right] = \frac{1}{2}\left[\sqrt{\mathcal{M}}^{-1}\cdot\delta\mathcal{M}\right]$, и в результате

$$
\delta\left[\left(\sqrt{g^{-1}f}\right)^n\right] = \frac{n}{2}\cdot\left[g\left(\sqrt{g^{-1}f}\right)^n\delta g^{-1}\right],
$$

откуда легко выходит

$$\frac{2}{\sqrt{-g}}\delta\left(\sqrt{-g}e_n\left(\sqrt{g^{-1}f}\right)\right)$$
$$=\sum_{m=0}^{n}(-1)^{m+1}\left[g\left(\sqrt{g^{-1}f}\right)^m\delta g^{-1}\right]\cdot e_{n-m}\left(\sqrt{g^{-1}f}\right). \quad (3.26)$$

Вариация действия (3.24) по отношению к метрике $g_{\mu\nu}$ даёт с использованием (3.26):

$$\overset{g}{R}_{\mu\nu}-\frac{1}{2}\overset{g}{R}\,g_{\mu\nu}=\frac{1}{M_g^2}T^{\mu}_{\ \nu}$$
$$-\frac{m^2}{2}\sum_{n=0}^{3}(-1)^n\beta_n\left(g_{\mu\lambda}\mathbb{Y}_{(n)}{}^{\lambda}_{\nu}\left(\sqrt{g^{-1}f}\right)+g_{\nu\lambda}\mathbb{Y}_{(n)}{}^{\lambda}_{\mu}\left(\sqrt{g^{-1}f}\right)\right), \quad (3.27)$$

где $\mathbb{Y}_{(n)}$ определяются как

$$\mathbb{Y}_{(0)}(\mathcal{M})=\mathbb{I},$$

$$\mathbb{Y}_{(1)}(\mathcal{M})=\mathcal{M}-\mathbb{I}\cdot[\mathcal{M}],$$
$$\mathbb{Y}_{(2)}(\mathcal{M})=\mathcal{M}^2-\mathcal{M}\cdot[\mathcal{M}]+\frac{1}{2}\mathbb{I}\cdot\left([\mathcal{M}]^2-[\mathcal{M}^2]\right),$$

$$\mathbb{Y}_{(3)}(\mathcal{M})=\mathcal{M}^3-\mathcal{M}^2\cdot[\mathcal{M}]+\frac{1}{2}\mathcal{M}\cdot\left([\mathcal{M}]^2-[\mathcal{M}^2]\right)$$
$$-\frac{1}{6}\mathbb{I}\cdot\left([\mathcal{M}]^3-3[\mathcal{M}]\cdot[\mathcal{M}^2]+2[\mathcal{M}^3]\right),$$

или в общем виде

$$\mathbb{Y}_{(n)}(\mathcal{M})=\sum_{i=0}^{n}(-1)^i e_i(\mathcal{M})\cdot(\mathcal{M})^{n-i}. \quad (3.28)$$

Разумеется, это уравнение будет точно таким же и при фиксированной (нединамической) метрике f. Мы откладывали его выписывание до обсуждения биметрической гравитации только во избежание излишних повторений.

Аналогично, варьируя действие (3.24) по отношению к метрике $f_{\mu\nu}$, получаем

$$\overset{f}{R}_{\mu\nu} - \frac{1}{2}\overset{f}{R}\, f_{\mu\nu} = -\frac{m^2}{2M_*^2}\sum_{n=0}^{3}(-1)^n\beta_{4-n}\left(f_{\mu\lambda}\mathbb{Y}_{(n)}{}^{\lambda}_{\nu}\left(\sqrt{f^{-1}g}\right) + f_{\mu\lambda}\mathbb{Y}_{(n)}{}^{\lambda}_{\mu}\left(\sqrt{f^{-1}g}\right)\right), \tag{3.29}$$

где введено безразмерное отношение планковских масс $M_* = \frac{M_f}{M_g}$. Буквы g или f над геометрическими величинами означают, что оные были вычислены для метрики $g_{\mu\nu}$ или $f_{\mu\nu}$ соответственно.

Для полноты отметим, что вследствие тождеств Бьянки (ковариантного сохранения тензора Эйнштейна) из уравнений движения следуют соотношения

$$\overset{f}{\nabla}_\mu \sum_{n=0}^{3}(-1)^n\beta_{4-n}\left(\mathbb{Y}_{(n)}{}^{\mu}_{\nu}\left(\sqrt{f^{-1}g}\right) + f^{\mu\alpha}f_{\nu\lambda}\mathbb{Y}_{(n)}{}^{\lambda}_{\alpha}\left(\sqrt{f^{-1}g}\right)\right) = 0$$

и, с учетом ковариантного сохранения тензора энергии-импульса материи,

$$\overset{g}{\nabla}_\mu \sum_{n=0}^{3}(-1)^n\beta_{n}\left(\mathbb{Y}_{(n)}{}^{\mu}_{\nu}\left(\sqrt{g^{-1}f}\right) + g^{\mu\alpha}g_{\nu\lambda}\mathbb{Y}_{(n)}{}^{\lambda}_{\alpha}\left(\sqrt{g^{-1}f}\right)\right) = 0.$$

3.2.4 Проблема квадратных корней

Во всём предыдущем обсуждении мы игнорировали проблемы существования и единственности квадратного корня. Они не возникают при пертурбативном определении теории, поскольку квадратный корень возникает в виде ряда Тейлора. Однако это очень важный вопрос для понимания того, как рассматриваемая модель устроена на фундаментальном уровне. Для вещественнозначных невырожденных матриц вещественнозначный квадратный корень существует при очень широких предположениях: не должно быть отрицательных собственных значений, отвечающих нечётному числу идентичных жордановых клеток. Можно убедиться, что несуществование вещественного матричного корня связано с конфигурациями, в которых две метрики теории в некотором смысле каузально несовместимы друг с другом [130].

Вместе с тем существует также проблема неединственности квадратного корня. В ней присутствуют два аспекта. Есть дискретная неопределённость в выборе знаков собственных значений, аналогичная двум ветвям обычной функции квадратного корня. Но в некоторых случаях возникает и континуальная неопределённость. Причём связана она в первую очередь с наиболее интересными симметричными ситуациями. В частности, имеется бесконечное множество квадратных корней из единичной матрицы, например $\begin{pmatrix} 3/5 & -4/5 \\ -4/5 & -3/5 \end{pmatrix} \cdot \begin{pmatrix} 3/5 & -4/5 \\ -4/5 & -3/5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Вместе с тем, для представленного выше вывода уравнений движения принципиально важно предположение о том, что существует матрица $\sqrt{\mathcal{M}}$ как гладкая функция матрицы $\mathcal{M}$, так что можно корректно понимать все промежуточные действия. Оказывается, что это предположение может нарушаться в случаях континуальной свободы.

Касательно доказательств отсутствия духов, очевидно, что доказательство Хассана и Розен допускает, как минимум, некоторую свободу выбора квадратного корня, а в следующем разделе мы приведём подход, вообще избегающий явного выбора этой матрицы. Поэтому вопрос о значении нетривиальных квадратных корней приобретает особую актуальность. Мы ещё вернемся к этому важному аспекту теории в дальнейшем.

3.3 Анализ со вспомогательными полями

В этом разделе мы изложим альтернативный подход к доказательству отсутствия духа Боулвара-Дезера, предложенный в нашей работе [7*]. Введём вспомогательные тензорные поля Φ^μ_ν в β_1 модель дРГТ, и запишем потенциал в виде

$$V = \frac{m^2}{N} \left(\Phi^\mu_\mu + \left(\Phi^{-1}\right)^\mu_\nu N^2 g^{\nu\alpha} f_{\alpha\mu} \right). \tag{3.30}$$

Переменные Φ^μ_ν – нединамические. Их уравнения движения чисто алгебраические: $\Phi^2 = N^2 g^{-1} f$. Подставляя это соотношение в потенциал (3.30), получаем знакомое слагаемое $2m^2 \mathrm{Tr}\sqrt{g^{-1}f}$ в действии (3.4).

Мы будем в дальнейшем предполагать опорную метрику Минковского $f_{\mu\nu} = \eta_{\mu\nu}$. При это можно потребовать от вспомогательных полей следующих свойств симметрии: $\Phi^k_i = \Phi^i_k$ и $\Phi^0_i = -\Phi^i_0$.

Если абсолютно последовательно выполнять все шаги гамильтонова анализа, то в модели имеются следующие первичные связи (обращение в нуль импульсов переменных без временных производных):

$$\mathcal{C}_1 = \pi_N,$$

$$\mathcal{C}_{2i} = \pi_{N^i},$$

$$\mathcal{C}_{3\,\nu}^{\ \mu} = \pi_{\Phi^\nu_\mu}.$$

Их надо коммутировать (брать скобки Пуассона) с гамильтонианом. Очевидно, что связи $\mathcal{C}_3$ приводят к тому результату, который мы уже обсудили, то есть вторичным связям вида

$$\mathcal{C}_4 = \Phi^2 - N^2 g^{-1} f,$$

а нас теперь интересует, что даёт коммутация с гамильтонианом связей $\mathcal{C}_1$ и $\mathcal{C}_{2i}$.

Для $\mathcal{C}_{2i}$ несложным вычислением получаем

$$\mathcal{C}_{6i} = \sqrt{\gamma}\left(-2 \overset{(3)}{\bigtriangledown}{}^k \pi_{ik} + m^2 N^2 \left(\Phi^{-1}\right)^\mu_\nu \frac{\partial}{\partial N^i} g^{\nu\alpha} f_{\alpha\mu}\right),$$

причём производную матрицы в последнем слагаемом легко представить в виде

$$N^2 \frac{\partial}{\partial N^i} g^{\nu\alpha} f_{\alpha\mu} = \begin{pmatrix} 0 & \delta^j_i \\ -\delta^k_i & -\delta^j_i N^k - \delta^k_i N^j \end{pmatrix}.$$

Для связи же $\mathcal{C}_1$ имеем

$$\mathcal{C}_5 = \sqrt{\gamma}\left(-\overset{(3)}{R} - \frac{1}{\gamma}\left(\frac{1}{2}\left(\pi^j_j\right)^2 - \pi_{ik}\pi^{ik}\right) + 2m^2 N \left(\Phi^{-1}\right)^i_j \gamma^{ij}\right).$$

На данном этапе получаем полную гамильтонову плотность в виде

$$\begin{aligned}\mathcal{H} = -\sqrt{\gamma}N\left(\overset{(3)}{R} + \frac{1}{\gamma}\left(\frac{1}{2}\left(\pi^j_j\right)^2 - \pi_{ik}\pi^{ik}\right)\right) - 2\sqrt{\gamma}N^i \overset{(3)}{\nabla}{}^k \pi_{ik} \\ + \sqrt{\gamma}m^2\left(\Phi^\mu_\mu + \left(\Phi^{-1}\right)^\mu_\nu N^2 g^{\nu\alpha} f_{\alpha\mu}\right) + \\ + \lambda_1\pi_N + \lambda_2{}^i\pi_{N^i} + \lambda_3{}^\mu_\nu\pi_{\Phi^\mu_\nu} + \sqrt{\gamma}\lambda_4{}^\mu_\nu\left(\Phi^\nu_\alpha\Phi^\alpha_\mu - N^2 g^{\nu\alpha} f_{\alpha\mu}\right) + \\ + \sqrt{\gamma}\lambda_5\left(-\overset{(3)}{R} - \frac{1}{\gamma}\left(\frac{1}{2}\left(\pi^j_j\right)^2 - \pi_{ik}\pi^{ik}\right) + 2m^2 N\left(\Phi^{-1}\right)^i_j\gamma^{ij}\right) + \\ + \sqrt{\gamma}\lambda_6{}^i\left(-2\overset{(3)}{\nabla}{}^k\pi_{ik} + 2m^2\left(\left(\Phi^{-1}\right)^i_0 + \left(\Phi^{-1}\right)^i_j N^j\right)\right).\end{aligned} \tag{3.31}$$

Каждому нулевому импульсу соответствует некоторое (нелинейное) уравнение на переменные теории. В случае общего положения мы получили бы систему связей второго рода, вторая половина которых (вторичные) позволяет алгебраически определить значения нединамических переменных в терминах компонент пространственной метрики и их канонических импульсов.

Однако же можно убедиться, что при выбранной нами форме потенциала существует определённая линейная комбинация (нефизических) импульсов, которая коммутирует с гамильтонианом. Это означает, что полученные уравнения $\mathcal{C}_4$, $\mathcal{C}_5$, $\mathcal{C}_6$ не позволяют однозначно найти значения нединамических переменных. И, следовательно, среди этих уравнений есть комбинация, которая порождает нетривиальную связь в физическом (пространственном) секторе вместо определения заведомо нединамических переменных. Тем самым, число степеней свободы оказывается строго меньшим шести. Ниже мы приводим некоторые технические детали.

3.3.1 Технические подробности

Нам надо явно вычислить скобки Пуассона полного гамильтониана с нефизическими импульсами. Причём, поскольку требуется обращение в нуль только в слабом по Дираку смысле, то есть на поверхности связей, то коммутирование с первыми слагаемыми гамильтониана новой информации не даёт, ибо оно уже породило вторичные связи. Соответственно, мы находим

$$\frac{1}{\sqrt{\gamma}}\{\mathcal{C}_1\ ,\ H\} = -2{\lambda_4}^i_k N\gamma^{ik} + 2\lambda_5 m^2 \left(\Phi^{-1}\right)^i_j \gamma^{ij},$$

$$\frac{1}{\sqrt{\gamma}}\{\mathcal{C}_{2i}\ ,\ H\} = -2{\lambda_4}^i_0 + 2{\lambda_4}^i_k N^k + 2m^2 {\lambda_6}^k \left(\Phi^{-1}\right)^i_k,$$

$$\frac{1}{\sqrt{\gamma}}\{{\mathcal{C}_3}^\mu_\nu\ ,\ H\} = 2{\lambda_4}^\mu_\alpha \Phi^\alpha_\nu + 2m^2\lambda_5\gamma^{ij}N\frac{\partial}{\partial\Phi^\nu_\mu}\left(\Phi^{-1}\right)^i_j$$
$$+ 2m^2 {\lambda_6}^i \frac{\partial}{\partial\Phi^\nu_\mu}\left(\left(\Phi^{-1}\right)^i_0 + \left(\Phi^{-1}\right)^i_j N^j\right),$$

где была использована симметрия вспомогательных полей и отвечающих им множителей Лагранжа: ${\lambda_4}^k_i = {\lambda_4}^i_k$ и ${\lambda_4}^0_i = -{\lambda_4}^i_0$. Число уравнений в этой системе совпадает с числом неизвестных (множителей Лагранжа), и в случае общего положения можно ожидать, что все множители Лагранжа будут определены (а коммутирование гамильтониана со вторичными связями точно так же определит множители Лагранжа при нефизических импульсах), и на этом анализ закончится, оставив шесть степеней свободы (по числу компонент γ_{ij}) и четырнадцать нефизических переменных, ограниченных четырнадцатью связями второго рода.

Однако, используя связь $\mathcal{C}_4$ и элементарное соотношение

$$\left(\Phi^{-1}\right)^\alpha_\mu \frac{\partial}{\partial\Phi^\alpha_\nu}\left(\Phi^{-1}\right)^\beta_\kappa = -\left(\Phi^{-2}\right)^\beta_\mu\left(\Phi^{-1}\right)^\nu_\kappa,$$

из последнего коммутатора легко получить

$$\lambda_{4}{}^{i}_{j} = \frac{1}{2\sqrt{\gamma}} \left(\Phi^{-1}\right)^{i}_{\alpha} \{\mathcal{C}_{3}{}^{\alpha}_{j} \ , \ H\} \\ + m^2 \left(\frac{\lambda_5}{N} \left(\Phi^{-1}\right)^{i}_{j} + \frac{\lambda_6{}^{k} \gamma_{kj}}{N^2} \left(\left(\Phi^{-1}\right)^{i}_{0} + \left(\Phi^{-1}\right)^{i}_{l} N^l \right) \right),$$

$$\lambda_{4}{}^{k}_{0} = \frac{1}{2\sqrt{\gamma}} \left(\Phi^{-1}\right)^{k}_{\alpha} \{\mathcal{C}_{3}{}^{\alpha}_{0} \ , \ H\} \\ + m^2 \left(\frac{\lambda_5 N^j}{N} \left(\Phi^{-1}\right)^{k}_{j} + \frac{\lambda_6{}^{i} N_i}{N^2} \left(\left(\Phi^{-1}\right)^{k}_{0} + \left(\Phi^{-1}\right)^{k}_{l} N^l \right) \right).$$

Мы видим, что комбинация множителей Лагранжа $\lambda_{4}{}^{k}_{0} - \lambda_{4}{}^{k}_{j} N^j$ определяется при этом только коммутатором с $\mathcal{C}_3$, в то время как λ_5 и λ_6 выпадают. Используя теперь коммутатор с $\mathcal{C}_2$, получаем (в слабом смысле, разумеется)

$$2m^2 \lambda_6{}^{k} \left(\Phi^{-1}\right)^{i}_{k} = \frac{1}{\sqrt{\gamma}} \left\{ \mathcal{C}_{2i} + \left(\Phi^{-1}\right)^{i}_{\alpha} \mathcal{C}_{3}{}^{\alpha}_{0} - \left(\Phi^{-1}\right)^{i}_{\alpha} N^j \mathcal{C}_{3}{}^{\alpha}_{j} \ , \ H \right\}.$$

С другой стороны, можно вычислить величину $\lambda_{4}{}^{i}_{j} \gamma^{ij}$ и сравнить с коммутатором $\{\mathcal{C}_1 \ , \ H\}$. Это тоже позволяет (в слабом по Дираку смысле) определить множители λ_6:

$$2m^2 \lambda_6{}^{k} \left(\left(\Phi^{-1}\right)^{k}_{0} + \left(\Phi^{-1}\right)^{k}_{l} N^l \right) = -\frac{1}{\sqrt{\gamma}} \left\{ N^2 \left(\Phi^{-1}\right)^{i}_{\alpha} \gamma^{ij} \mathcal{C}_{3}{}^{\alpha}_{j} + N \mathcal{C}_1 \ , \ H \right\}.$$

Полученные два результата должны быть согласованы друг с другом, и исходя из этого, мы получаем комбинацию импульсов

$$\pi_N + N \left(\Phi^{-1}\right)^{i}_{\alpha} \gamma^{ik} \pi_{\Phi^k_\alpha} \\ + \frac{\left(\Phi^{-1}\right)^{i}_{0} + \left(\Phi^{-1}\right)^{i}_{l} N^l}{N} \left(\left(\overset{(3)}{\Phi^{-1}} \right)^{-1} \right)^{k}_{i} \times \\ \left(\pi_{N^k} + \left(\Phi^{-1}\right)^{k}_{\alpha} \pi_{\Phi^0_\alpha} - \left(\Phi^{-1}\right)^{k}_{\alpha} N^j \pi_{\Phi^j_\alpha} \right),$$

которая слабо коммутирует с гамильтонианом вне зависимости от значений множителей Лагранжа. Это и есть та особенность дРГТ потенциала, которую мы хотели выявить нашим методом. Для этого нам не пришлось явно извлекать квадратный корень из матрицы.

В нашем методе это определяет направление в пространстве нединамических переменных, вдоль которого пока не получено никаких ограничений. В стандартном подходе это соответствует тому, что четыре связи, обобщающие обычные эйнштейновские связи, не зависят от переменной шага, если их выразить в терминах нового вектора сдвига. Далее следует прокоммутировать гамильтониан со связями $\mathcal{C}_4$, $\mathcal{C}_5$, $\mathcal{C}_6$ для определения множителей Лагранжа λ_1, λ_2, λ_3. Однако мы знаем, что есть комбинация первичных связей, которая без всяких дополнительных условий прокоммутирует со всеми вторичными связями, а следовательно система уравнений не позволит определить всех искомых множителей Лагранжа, и вместо этого возникнет новая (для нас уже третичная) связь. Поскольку в системе нет калибровочной свободы, то следует ожидать, что, в паре с физической комбинацией вторичных связей, они будут связями второго рода, уменьшая число степеней свободы до пяти.

3.3.2 Дополнительные замечания

Полнота рассмотрения

Разумеется, хотелось бы, чтобы гамильтонов анализ был явно проведён до конца. И действительно, в результате утомительных вычислений можно убедиться, что третичная связь возникает, но выражения оказываются чрезвычайно громоздкими, и довести эти вычисления до полного логического завершения с определением всех связей (второго рода) и множителей Лагранжа оказывается очень сложно. Насколько мы знаем, никто этого не сделал. В стандартном подходе это было тоже проделано отнюдь не сразу, наиболее полное рассмотрение представлено в работе Хассана и Розен [119].

Нашим методом оказывается сравнительно просто установить факт появления одной связи в чисто пространственном секторе, поскольку не приходится явно извлекать квадратный корень из матрицы, но при этом полный последовательный анализ усложняется в силу появления большого числа вспомогательных переменных. Тем не менее, наш метод анализа оказался полезным и удобным для обсуждения отсутствия духа на языке полей Штюкельберга [120]. С теоретической же точки зрения важно, что наше рассуждение не привязано к конкретному способу извлечения квадратного корня.

Произвольная опорная метрика

Мы использовали опорную метрику Минковского. Известно, что отсутствие духа можно доказать и для произвольной опорной метрики [117]. Вычисления при этом становятся более сложными, изменяется вид анзатца (3.22) для квадратного корня.

Если для опорной метрики тоже использовать АДМ разложение в виде

$$f_{\mu\nu} = \begin{pmatrix} -\left(M^2 - M_k M^k\right) & M_i \\ M_j & s_{ij} \end{pmatrix},$$

то требуется извлекать квадратный корень из матрицы

$$g^{\mu\alpha} f_{\alpha\nu} = \begin{pmatrix} \frac{M^2 - M_k\left(M^k - N^k\right)}{N^2} & \frac{s_{ij}N^j - M_j}{N^2} \\ -\frac{N^j\left(M^2 - M_k\left(M^k - N^k\right)\right)}{N^2} + \gamma^{ij} M_j & s_{ik}\gamma^{kj} - \frac{s_{ik}N^k N^j - N_i M^j}{N^2} \end{pmatrix}.$$

В нашем подходе возникающие усложнения относительно небольшие: в связи $\mathcal{C}_5$ возникает более сложное слагаемое $2m^2 N\sqrt{\gamma}\left(\left(\Phi^{-1}\right)^0_i M_j + \left(\Phi^{-1}\right)^k_i s_{kj}\right)\gamma^{ij}$ вместо простого $2m^2 N\sqrt{\gamma}\left(\Phi^{-1}\right)^i_j \gamma^{ij}$, и в связи $\mathcal{C}_6$ происходят аналогичные изменения. Однако, все изменения соответствуют скорее некоторому повороту в пространстве параметров, чем усложнению по существу.

В принципе, так оно и должно было быть. Ключевое свойство – линейность потенциала по одной комбинации нединамических параметров. Она устанавливается в каждой точке пространства-времени без участия производных. Вместе с тем, в заданной точке опорная метрика может быть приведена к виду Минковского подходящей заменой координат (разумеется, физическая метрика при этом тоже преобразуется).

За пределами β_1

Как мы уже ранеее упоминали, для произвольного дРГТ потенциала отсутствие духа легко доказывается, если уже доказано, что β_1 потенциал может быть представлен в виде (3.22). Дело в том, что первое слагаемое в правой части (3.22) сохраняет свою форму при возведении в квадрат, и можно убедиться, что из β_2 и β_3 членов выпадают старшие степени $\frac{1}{N}$. Однако, в принципе, нашим методом можно рассмотреть произвольный потенциал и напрямую, по той же самой схеме.

Интересная особенность возникает при этом для случая чистой β_2-модели. Легко видеть, что вариирование слагаемого

$$\Phi^\alpha_\alpha \left(\Phi^{-1}\right)^\mu_\nu \left(g^{-1}f\right)^\nu_\mu - \left(g^{-1}f\right)^\alpha_\alpha,$$

воспроизводящего β_2-потенциал, по отношению ко вспомогательным полям Φ определяет последние только с точностью до общего (несущественного) множителя. Это очень напоминает конформную инвариантность в действии для струны по Полякову, и, по-видимому, должно говорить о наличии дополнительных структур, связанных с β_2-моделью. И действительно, именно с чистой β_2-моделью связывались надежды на нелинейную реализацию частично безмассовой гравитации [131], хотя они и не оправдались [132].

3.4 Квадратные корни из матриц

В этом разделе мы кратко обсудим избранные разделы теории матриц (см., например, книгу [133]), которые важны для понимания особенностей непертурбативной теории массивной гравитации. Разумеется, новых результатов здесь не представлено. Однако эти факты были существенно использованы в нашей работе [9*]. И поскольку они часто оказываются за пределами математического кругозора физика-теоретика, мы считаем необходимым обсудить их здесь. Всюду в этом разделе мы имеем в виду квадратные матрицы над полем комплексных чисел.

Спектральная теорема

В конечномерном случае можно дать полную спектральную классификацию линейных операторов, описав все классы эквивалентности по отношению к преобразованиям подобия с помощью жордановой нормальной формы. А именно, любая квадратная матрица $\mathbb{X}$ может быть записана как блочно-диагональная с блоками из жордановых клеток,

$$\mathcal{S}\mathbb{X}\mathcal{S}^{-1} = J_{k_1}(\lambda_1) \bigoplus J_{k_2}(\lambda_2) \bigoplus \ldots,$$

где $\mathcal{S}$ – невырожденная матрица преобразования подобия (описывающая замену базиса в векторном пространстве), $J_{k_i}(\lambda_i)$ – жорданова клетка с собственным значением λ_i и размером k_i:

$$J(\lambda) \equiv \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}.$$

Каждая такая клетка прибавляет для своего собственного значения 1 к геометрической и k_i – к алгебраической кратности. В компонентах жорданов блок можно определить так: $J_{i,i} = \lambda$ и $J_{i,i+1} = 1$, остальные матричные элементы равны нулю. Если все жордановы блоки одномерны, то матрица, конечно же, диагонализуема.

Приведём набросок одного из возможных вариантов доказательства.

1) У любой $n \times n$ матрицы всегда есть собственный вектор. В самом деле, характеристическое уравнение $P(\lambda) = \det(\mathbb{X} - \lambda\mathbb{I}) = 0$ является условием разрешимости системы уравнений $\mathbb{X}e = \lambda e$, где e – вектор из того пространства, где действует линейный оператор $\mathbb{X}$. По основной теореме алгебры, оно всегда имеет решение. Выберем любой собственный вектор в качестве последнего элемента базиса (иными словами, выпол-

ним соответствующее преобразование подобия), тогда получаем $\mathbb{X}_{n,k} = 0$ при $k < n$.

2) Теперь проделаем то же самое с $(n-1) \times (n-1)$-мерным минором в левом верхнем углу, и так далее. Получаем верхнетреугольную матрицу. Очевидно, что на диагонали стоят корни характеристического многочлена, то есть собственные значения. Без ограничения общности можно считать, что одинаковые собственные значения стоят рядом друг с другом.

3) Выберем теперь одно из собственных значений, пусть это λ_1. Тогда можно записать $\mathbb{X} - \lambda_1 \mathbb{I} = \begin{pmatrix} \tilde{X} & M \\ \mathbb{O} & X_0 \end{pmatrix}$, где $\tilde{X}$ невырожденна, а X_0 нильпотентна. Матрицу $\mathbb{X} - \lambda_1 \mathbb{I}$, а с ней и $\mathbb{X}$, можно привести к блочно-диагональному виду преобразованием подобия $\begin{pmatrix} \mathbb{I} & C \\ \mathbb{O} & \mathbb{I} \end{pmatrix}$, если оно удовлетворяет уравнению $\tilde{X}C - CX_0 = M$. Последнее имеет решение $C = \sum\limits_{i \geqslant 0} \tilde{X}^{-i-1} M X_0^i$, являющееся хорошо определённой конечной суммой, поскольку $\tilde{X}$ невырожденна, а X_0 нильпотентна. Таким образом можно привести матрицу к блочно-диагональному виду, причём у каждого конкретного блока все собственные значения равны друг другу, но у любых двух различных блоков они отличаются.

4) Теперь нам надо установить структуру матрицы X_{λ_1}, все собственные значения которой равны некоторому фиксированному числу λ_1. Очевидно, достаточно изучить нильпотентную матрицу $X_0 = X_{\lambda_1} - \lambda_1 \mathbb{I}$ заданного размера $k \times k$. На самом деле, то, что мы хотим сейчас доказать, не зависит даже от приятных свойств комплексных чисел. А именно, по отношению к любому нильпотентному оператору всякое конечномерное векторное пространство может быть разложено в прямую сумму циклических подпространств.

5) Возьмём такое минимальное число $k_1 \leqslant k$, что $X_0^{k_1} = 0$. Тогда существует вектор, для которого $X_0^{k_1 - 1} e_0 \neq 0$. Этот вектор вместе со всеми векторами $X_0^i e_0$ порождает инвариантное подпространство $\mathcal{V}_0$, в котором матрица X_0 принимает в так построенном базисе вид жорданова блока

с $\lambda = 0$, а X_{λ_1} – аналогичного жорданова блока со своим собственным числом.

6) Если $k_1 = k$, то всё доказано. В противном случае возьмём подходящий вектор e_1 из линейного дополнения к подпространству $\mathcal{V}_0$. Под подходящим мы понимаем такой, который требует максимально возможной степени X_0 для перехода внутрь $\mathcal{V}_0$, включая нуль. (Альтернативно можно также рассмотреть факторпространство по $\mathcal{V}_0$. Поскольку последнее является инвариантным подпространством, то на факторпространстве корректно определено действие X_0, и можно заново провести построение предыдущего шага.) Построим новое подпространство, натянутое на вектора $X_0^i e_1$.

7) У нас не получится блочно-диагональной формы, если окажется, что $0 \neq X_0^{k_2} e_1 \in \mathcal{V}_0$. Но в таком случае вектор $X_0^{k_2} e_1$ является линейной комбинацией базисных векторов $X_0^j e_0$ подпространства $\mathcal{V}_0$, причём только с $j \geqslant k_2$, поскольку в противном случае оказалось бы, что $X_0^{k_1} e_1 \neq 0$. Следовательно, существует такой вектор $\tilde{e} \in \mathcal{V}_0$, что $X_0^{k_2} \tilde{e} = X_0^{k_2} e_1$. Тогда при построении нового циклического подпространства мы можем взять в роли старшего вектора $e_1' = e_1 - \tilde{e}$ вместо e_1, и проблема будет разрешена.

8) Описанную процедуру можно при необходимости повторить, разбив в конечном итоге каждый блок из совпадающих собственных чисел на отдельные жордановы блоки. Спектральная теорема доказана.

Аналитические функции матриц

Поскольку матрицы можно складывать и умножать, то с самого начала корректно и однозначно определены полиномиальные функции от матриц. Заметим, что полиномиальные функции уважают отношение эквивалентности, задаваемое преобразованиями подобия, в том смысле, что

$$P(\mathcal{S}\mathbb{X}\mathcal{S}^{-1}) = \mathcal{S}P(\mathbb{X})\mathcal{S}^{-1},$$

или преобразования подобия сохраняют полиномиальные соотношения. Разумеется, определение функции от матрицы можно непосредственно обобщить на любые функции, задающиеся сходящимися степенными рядами. При этом удобно пользоваться жордановой нормальной формой:

$$f\left(\mathcal{S}^{-1}\left(\bigoplus_i J_{k_i}(\lambda_i)\right)\mathcal{S}\right) = \mathcal{S}^{-1}\left(\bigoplus_i f\left(J_{k_i}(\lambda_i)\right)\right)\mathcal{S},$$

поскольку степени жордановой клетки легко вычисляются, приводя к простому ответу:

$$f(J_k(\lambda)) \equiv \begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2}f''(\lambda) & \cdots & \frac{1}{(k-2)!}f^{(k-2)}(\lambda) & \frac{1}{(k-1)!}f^{(k-1)}(\lambda) \\ 0 & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(k-3)!}f^{(k-3)}(\lambda) & \frac{1}{(k-2)!}f^{(k-2)}(\lambda) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & f(\lambda) & f'(\lambda) \\ 0 & 0 & 0 & \cdots & 0 & f(\lambda) \end{pmatrix}.$$

Пользуясь полученным соотношением, можно распространить определение функции от матрицы на все функции класса $\mathcal{C}^{k-1}(\mathfrak{U})$, где k – максимальный размер имеющихся жордановых клеток, а область $\mathfrak{U} \in \mathbb{R}$ должна содержать в себе весь спектр матрицы.

В частности, поскольку функция $f(x) = \sqrt{x}$ является гладкой при $x \neq 0$, мы получаем корректное определение функции квадратного корня для любой невырожденной матрицы. Поскольку квадратный корень является многозначной функцией, появляется дискретная свобода в выборе ветви квадратного корня из матрицы – выбор знака числа $\sqrt{\lambda}$ для каждого из имеющихся жордановых блоков. (Ниже мы увидим, что есть ещё дополнительные тонкости в случае, если квадратный корень определяется не как гладкая функция, а как решение уравнения $\mathbb{A} \cdot \mathbb{A} = \mathbb{X}$.) Легко убедиться, что матрица $\sqrt{J_k(\lambda)}$ относится к классу эквивалентности $J_k(\pm\sqrt{\lambda})$ при $\lambda \neq 0$.

Замечание о вырожденных матрицах

Случай вырожденных матриц нас интересовать не будет (хотя вырождения и встречались в литературе при исследовании массивных космологий [123,124]), но для полноты отметим, что здесь появляются новые интересные явления.

Возведём, например, вырожденную трёхмерную жорданову клетку в квадрат:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Простой перенумерацией элементов базиса находим, что $(J_3(0))^2$ эквивалентно $J_2(0) \bigoplus J_1(0)$ в смысле преобразований подобия. Жорданова клетка расщепилась. Легко понять, что это явление связано с тем, что для функции возведения в квадрат $f' = 0$ в точке спектра данной матрицы. Можно доказать, что расщепление жордановых блоков может произойти только при этом условии [133].

Отметим теперь также, что $(J_2(0))^2 = \mathbb{O}$. Поскольку корень из $J_3(0)$ может состоять только из вырожденных жордановых блоков, то очевидно, что его просто не существует. Иными словами, не существует такой 3×3 матрицы $\mathbb{A}$, что $\mathbb{A}^2 = J_3(0)$. В случае вырожденных матриц квадратные корни существуют не всегда, даже над полем комплексных чисел.

Теорема Гамильтона-Кэли

Пусть $P(\lambda)$ – характеристический полином матрицы $\mathbb{X}$. Теорема Гамильтона-Кэли утверждает, что $P(\mathbb{X}) = \mathbb{O}$. Она легко следует из того, что мы уже обсудили. В самом деле, если в жордановой форме матрицы есть клетка $J_{k_i}(\lambda_i)$, то число λ_i является корнем многочлена P порядка не меньшего, чем k_i. Следовательно, $P(J_{k_i}(\lambda_i)) = \mathbb{O}$ для любой жордановой клетки в спектральном представлении матрицы $\mathbb{X}$.

Коммутирующие матрицы

Для решения матричных квадратных уравнений оказывается чрезвычайно полезным сперва научиться решать уравнение коммутации

$$\mathbb{A}\mathbb{X} = \mathbb{X}\mathbb{A},$$

где требуется найти все возможные матрицы $\mathbb{A}$, если дана матрица $\mathbb{X}$.

Выберем базис, в котором $\mathbb{X}$ принимает жорданову нормальную форму

$$\mathbb{X} = \bigoplus_i J_{k_i}(\lambda_i).$$

Это блочно-диагональная форма, поэтому рассмотрим матрицу $\mathbb{A}$ в блочном виде: состоящую из $k_i \times k_j$ матриц $A_{i,j}$, где k_i – размеры жордановых клеток матрицы $\mathbb{X}$. Тогда уравнение принимает вид (нет суммирования)

$$A_{i,j} J_{k_j}(\lambda_j) = J_{k_i}(\lambda_i) A_{i,j}.$$

При $\lambda_i \neq \lambda_j$ получаем

$$(\lambda_i - \lambda_j) A_{i,j} = J_{k_i}(0) A_{i,j} - A_{i,j} J_{k_j}(0).$$

Матрицы $J(0)$ в правой части нильпотентны. Можно умножить это уравнение на $\lambda_i - \lambda_j$ столько раз, сколько потребуется, пока не получим $A_{i,j} = 0$.

В случае $\lambda_i = \lambda_j$ уравнение принимает простой вид

$$A_{i,j} J_{k_j}(\lambda) = J_{k_i}(\lambda) A_{i,j}.$$

Учитывая, что жордановы блоки тоже очень простые, нетрудно найти при каких условиях на компоненты $A_{i,j}$ это уравнение будет удовлетворено. При этом получается континуальное семейство решений.

При $k_i = k_j$ (например, в диагональном блоке) получаем решение в виде $(A_{i,j})_{a,a+b} = C_b$, где C_b – произвольные числа, $b = 0, 1, \dots k-1$.

Если же $k_i \neq k_j$, то имеем либо $A_{i,j} = \begin{pmatrix} \mathbb{O} & \tilde{A}_{i,j} \end{pmatrix}$, либо $A_{i,j} = \begin{pmatrix} \tilde{A}_{i,j} \\ \mathbb{O} \end{pmatrix}$, где $\tilde{A}_{i,j}$ – верхнетреугольная матрица описанного выше вида и размера $\min(k_i, k_j) \times \min(k_i, k_j)$.

* * *

Заметим, что для диагонализуемых матриц мы просто получаем, что они должны быть диагонализуемы одновременно. При $\lambda_i = \lambda_j$ матричный элемент $A_{i,j}$ может быть произволен: в этом подпространстве $\mathbb{X}$ кратна единичной и коммутирует с чем угодно, а матрицу $\mathbb{A}$ можно при желании диагонализовать, не меняя матрицы $\mathbb{X}$. В случае же $\lambda_i \neq \lambda_j$, у матрицы $\mathbb{A}$ не может быть внедиагонального элемента. Это результат, хорошо известный, например, из квантовой механики – условие одновременной измеримости двух наблюдаемых.

Квадратные корни из матриц

Теперь мы можем вернуться к вопросу об извлечении квадратного корня из матрицы. Предположим, нам дана произвольная невырожденная матрица

$$\mathbb{X} = \mathcal{S}^{-1} \left(\bigoplus_i J_{k_i}(\lambda_i) \right) \mathcal{S},$$

и мы хотим решить квадратное уравнение

$$\mathbb{A}^2 = \mathbb{X}$$

относительно неизвестной матрицы $\mathbb{A}$.

Искомую матрицу тоже можно представить в виде прямой суммы жордановых клеток, причём без явлений расщепления, поскольку $f' \neq 0$ для функций квадратного корня и возведения в квадрат на спектрах невырожденных матриц. Есть взаимно-однозначное соответствие между жордановыми клетками $\mathbb{X}$ и $\mathbb{A}$, причём, очевидно, мы можем использовать $\sqrt{J_k(\lambda)}$ вместо $J_k(\sqrt{\lambda})$ в спектральном разложении матрицы $\mathbb{A}$:

$$\mathbb{A} = \mathcal{T}^{-1} \left(\bigoplus_i \sqrt{J_{k_i}(\lambda_i)} \right) \mathcal{T},$$

где $\mathcal{T}$ – неопределённое пока преобразование подобия, с простейшим возможным решением $\mathcal{T} = \mathcal{S}$.

Наше уравнение $\mathbb{A}^2 = \mathbb{X}$ можно записать теперь в виде

$$\mathcal{T}^{-1} \left(\bigoplus_i J_{k_i}(\lambda_i) \right) \mathcal{T} = \mathcal{S}^{-1} \left(\bigoplus_i J_{k_i}(\lambda_i) \right) \mathcal{S},$$

который легко преобразовать к форме уравнения коммутации

$$\mathcal{S}\mathcal{T}^{-1}\left(\bigoplus_i J_{k_i}(\lambda_i)\right) = \left(\bigoplus_i J_{k_i}(\lambda_i)\right)\mathcal{S}\mathcal{T}^{-1}.$$

Общее решение получается в виде утверждения о том, что $\mathcal{T}\mathcal{S}^{-1}$ коммутирует с $\bigoplus\limits_i J_{k_i}(\lambda_i)$. Иными словами, $\mathcal{T}$ является произведением $\mathcal{S}$ и произвольной матрицы, коммутирующей с $\bigoplus\limits_i J_{k_i}(\lambda_i)$. Структуру последней мы уже описали в предыдущем пункте.

Разумеется, новый (отличный от случая $\mathcal{T} = \mathcal{S}$) квадратный корень получается только тогда матрица $\mathcal{T}\mathcal{S}^{-1}$ хоть и коммутирует с $\bigoplus\limits_i J_{k_i}(\lambda_i)$, но не коммутирует с $\bigoplus\limits_i \sqrt{J_{k_i}(\lambda_i)}$.

Теперь мы видим, что при извлечении квадратного корня в смысле решения квадратного уравнения может появляться ещё и континуальная симметрия. Это происходит в тех случаях, когда у матрицы $\mathbb{X}$ есть совпадающие собственные числа $\lambda_i = \lambda_j$ в разных жордановых блоках. Для такой матрицы существует континуальное семейство коммутирующих матриц, вид $A_{i,j}$ блоков которых мы выше определили, или, что то же самое, существует континуальное семейство преобразований подобия, не меняющих данную матрицу. Если же мы для двух таких собственных чисел выберем разные ветви квадратного корня, $\sqrt{\lambda_i} = -\sqrt{\lambda_j}$, то полученная матрица $\sqrt{\mathbb{X}}$ уже не будет оставаться инвариантной при этих преобразованиях. Тем самым получается континуальное множество корней. Отметим, однако, что хорошие аналитические свойства, связанные с определением в виде аналитической функции, при этом не гарантированы. Более того, как мы ниже увидим, их просто нет.

3.5 Новый метод в теории возмущений

Выяснив структуру квадратных корней, мы попробуем теперь обойтись без их явного нахождения. Как будет показано в следующем разделе, такой метод позволяет работать со случаями континуальной симметрии, при которых теория в её стандартной формулировке просто не работает. Этот способ был предложен в нашей работе [8*].

3.5.1 Соотношение между симметрическими полиномами

Заметим, что для записи действия массивной гравитации нам достаточно знания лишь собственных значений матрицы $\sqrt{g^{-1}\eta}$, а точнее даже их симметрических полиномов. Можно надеяться тогда сформулировать теорию без явного нахождения матрицы квадратного корня, тем самым нивелируя проблему континуальной симметрии, описанной в предыдущем разделе, которая, очевидно, не затрагивает собственных чисел.

Попробуем связать симметрические полиномы исходной матрицы и её квадратного корня. Вот главное наблюдение:

$$\begin{aligned}\sum_{n=0}^{N}(-\lambda^2)^{N-n}\cdot e_n(\mathcal{M}^2) &= \det\left(\mathcal{M}^2-\lambda^2\mathbb{I}\right)\\ &= \det\left((\mathcal{M}-\lambda\mathbb{I})\cdot(\mathcal{M}+\lambda\mathbb{I})\right) = \det\left(\mathcal{M}-\lambda\mathbb{I}\right)\cdot\det\left(\mathcal{M}+\lambda\mathbb{I}\right)\\ &= \left(\sum_{k=0}^{N}(-\lambda)^{N-k}\cdot e_k(\mathcal{M})\right)\cdot\left(\sum_{m=0}^{N}\lambda^{N-m}\cdot e_m(\mathcal{M})\right). \end{aligned} \tag{3.32}$$

Сравнивая степени λ с обеих сторон, получаем как тривиальное соотношение

$$\sum_{k+m=2n+1}(-1)^k e_k(\mathcal{M})e_m(\mathcal{M}) = 0,$$

так и то, что нам надо:

$$\sum_{k+m=2n}(-1)^k e_k(\mathcal{M})e_m(\mathcal{M}) = (-1)^n e_n(\mathcal{M}^2). \tag{3.33}$$

Мы связали элементарные симметрические полиномы произвольной матрицы $\mathcal{M}$ и её квадрата $\mathcal{M}^2$. В частности, в четырёхмерном случае имеем

$$
\begin{aligned}
e_1(\mathcal{M}^2) &= e_1^2(\mathcal{M}) - 2e_2(\mathcal{M}), && (3.34)\\
e_2(\mathcal{M}^2) &= e_2^2(\mathcal{M}) - 2e_1(\mathcal{M})e_3(\mathcal{M}) + 2e_4(\mathcal{M}), && (3.35)\\
e_3(\mathcal{M}^2) &= e_3^2(\mathcal{M}) - 2e_2(\mathcal{M})e_4(\mathcal{M}), && (3.36)\\
e_4(\mathcal{M}^2) &= e_4^2(\mathcal{M}). && (3.37)
\end{aligned}
$$

Корни из единичной матрицы

В конечном счёте нас интересует случай $\mathcal{M}^2 = g^{-1}\eta$, причём для многих задач g мало отличается от η. Поэтому рассмотрим для начала простой случай $\mathcal{M}^2 = \mathbb{I}$. Уравнения (3.34) – (3.37) принимают вид

$$
\begin{aligned}
4 &= e_1^2(\sqrt{\mathbb{I}}) - 2e_2(\sqrt{\mathbb{I}}),\\
6 &= e_2^2(\sqrt{\mathbb{I}}) - 2e_1(\sqrt{\mathbb{I}})e_3(\sqrt{\mathbb{I}}) + 2e_4(\sqrt{\mathbb{I}}),\\
4 &= e_3^2(\sqrt{\mathbb{I}}) - 2e_2(\sqrt{\mathbb{I}})e_4(\sqrt{\mathbb{I}}),\\
1 &= e_4^2(\sqrt{\mathbb{I}}).
\end{aligned}
$$

Их довольно просто проанализировать и получить перечисленные ниже решения.

Одно решение совершенно очевидно:

$$
e_1(\sqrt{\mathbb{I}}) = \pm 4, \quad e_2(\sqrt{\mathbb{I}}) = 6, \quad e_3(\sqrt{\mathbb{I}}) = \pm 4, \quad e_4(\sqrt{\mathbb{I}}) = 1,
$$

оно отвечает тривиальному квадратному корню

$$
\sqrt{\mathbb{I}} = \pm \begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1 \end{pmatrix}.
$$

Почти всегда именно оно и используется, но им возможности не исчерпываются.

Другое решение имеет вид

$$e_1(\sqrt{\mathbb{I}}) = 0, \quad e_2(\sqrt{\mathbb{I}}) = -2, \quad e_3(\sqrt{\mathbb{I}}) = 0, \quad e_4(\sqrt{\mathbb{I}}) = 1,$$

представляя квадратный корень

$$\sqrt{\mathbb{I}} = \pm \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

а также все его преобразования подобия, поскольку

$$\left(\mathcal{C} \cdot \sqrt{\mathbb{I}} \cdot \mathcal{C}^{-1}\right)^2 = \mathcal{C} \cdot (\sqrt{\mathbb{I}})^2 \cdot \mathcal{C}^{-1} = \mathbb{I}$$

для любой невырожденной матрицы $\mathcal{C}$.

Наконец, существует и третье решение:

$$e_1(\sqrt{\mathbb{I}}) = \mp 2, \quad e_2(\sqrt{\mathbb{I}}) = 0, \quad e_3(\sqrt{\mathbb{I}}) = \pm 2, \quad e_4(\sqrt{\mathbb{I}}) = -1.$$

Оно представляет корень

$$\sqrt{\mathbb{I}} = \pm \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

и все его преобразования подобия.

Можно убедиться, что в стандартном подходе возмущения вокруг второго и третьего решений плохо определены. В самом деле, легко проверить, что, если прибавлять малое вомущение к матрице $\begin{pmatrix} \mathbb{I} & \mathbb{O} \\ \mathbb{O} & -\mathbb{I} \end{pmatrix}$, то в линейном приближении ее квадрат никогда не получает ненулевых элементов во внедиагональных блоках. Соответственно, любое сколь угодно малое возмущение единичной матрицы, не уважающее блочно-диагональную структуру этого корня, приводит к невозможности так извлечь квадратный корень из полученной матрицы, чтобы он был малым возмущением исходного.

Причина очень проста. Единичная матрица не имеет никаких предпочтительных направлений в своём векторном пространстве. Извлечение же квадратного корня необычного вида предполагает необходимость выбрать два взаимно дополнительных подпространства, которые будут собственными для матрицы квадратного корня, и при этом с разными знаками собственных значений. Но если мы вводим возмущение, не коммутирующее с $\sqrt{\mathbb{I}}$, то тем самым оно определяет фиксированные собственные вектора для возмущённой матрицы, не уважая при этом произвольный выбор собственных подпространств, сделанный при записи квадратного корня. И несоответствие может отвечать сколь угодно большим углам. Поэтому и гладкое изменение такого квадратного корня становится невозможным (не говоря уже о том, что от континуальной свободы выбора корня возмущение оставляет лишь дискретную).

Новая формулировка теории

Итак, мы записываем действие теории в обычном виде

$$S = \int d^4x\sqrt{-g}\left(R + m^2\sum_{n=0}^{4}\beta_n\mathfrak{e}_n\right), \tag{3.38}$$

но при этом определяем величины $\mathfrak{e}_i$ как решения алгебраических уравнений

$$\begin{aligned}
e_1(g^{-1}\eta) &= \mathfrak{e}_1^2 - 2\mathfrak{e}_2, & (3.39)\\
e_2(g^{-1}\eta) &= \mathfrak{e}_2^2 - 2\mathfrak{e}_1\mathfrak{e}_3 + 2\mathfrak{e}_4, & (3.40)\\
e_3(g^{-1}\eta) &= \mathfrak{e}_3^2 - 2\mathfrak{e}_2\mathfrak{e}_4, & (3.41)\\
e_4(g^{-1}\eta) &= \mathfrak{e}_4^2, & (3.42)
\end{aligned}$$

нигде явно не вводя матрицы $\sqrt{g^{-1}\eta}$.

Заметим, что это принципиально отличается от нашей работы со вспомогательными полями из раздела 3.3, поскольку вспомогательные поля в последней – по сути, и есть компоненты матрицы квадратного корня. Вместе с тем, представленный здесь метод позволяет обойти именно проблематичную континуальную симметрию, которая столь же непосредственно относится и к определению вспомогательных полей. Но для начала мы хотим показать как он работает в обычном случае.

3.5.2 Предел Фирца-Паули в рамках нового метода

Применим уравнения (3.39) – (3.42), чтобы вопроизвести действие Фирца-Паули. Используя определения (3.7), (3.8), (3.9), и если угодно (3.10), получаем для левых частей наших уравнений степенные разложения:

$$\begin{aligned}
e_1(g^{-1}\eta) &= 4 - h^\mu_\mu + h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3), \\
e_2(g^{-1}\eta) &= 6 - 3h^\mu_\mu + \frac{1}{2}(h^\mu_\mu)^2 + \frac{5}{2}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3), \\
e_3(g^{-1}\eta) &= 4 - 3h^\mu_\mu + (h^\mu_\mu)^2 + 2h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3), \\
e_4(g^{-1}\eta) &= 1 - h^\mu_\mu + \frac{1}{2}(h^\mu_\mu)^2 + \frac{1}{2}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3).
\end{aligned}$$

Нас сейчас интересуют возмущения вокруг тривиального решения $\sqrt{\mathbb{I}} = \mathbb{I}$, и поэтому мы кладём

$$\mathfrak{e}_1 = 4 + \delta\mathfrak{e}_1, \quad \mathfrak{e}_2 = 6 + \delta\mathfrak{e}_2, \quad \mathfrak{e}_3 = 4 + \delta\mathfrak{e}_3, \quad \mathfrak{e}_4 = 1 + \delta\mathfrak{e}_4.$$

Конечно же, уравнение (3.42) может быть легко разрешено относительно $\mathfrak{e}_4$ с любой необходимой точностью, и мы имеем

$$\mathfrak{e}_4 = 1 - \frac{1}{2}h^\mu_\mu + \frac{1}{8}(h^\mu_\mu)^2 + \frac{1}{4}h_{\mu\nu}h^{\mu\nu} + \mathcal{O}(h^3),$$

естественно воспроизводя формулу (3.16).

Рассмотрим оставшиеся три уравнения с точностью до первого порядка:

$$\begin{aligned} h^\mu_\mu &= 2\delta\mathfrak{e}_2 - 8\delta\mathfrak{e}_1, \\ 2h^\mu_\mu &= 8\delta\mathfrak{e}_1 + 8\delta\mathfrak{e}_3 - 12\delta\mathfrak{e}_2, \\ 9h^\mu_\mu &= 2\delta\mathfrak{e}_2 - 8\delta\mathfrak{e}_3, \end{aligned}$$

откуда легко находим

$$\begin{aligned} \delta\mathfrak{e}_1 &= -\frac{1}{2}h^\mu_\mu, \\ \delta\mathfrak{e}_2 &= -\frac{3}{2}h^\mu_\mu, \\ \delta\mathfrak{e}_3 &= -\frac{3}{2}h^\mu_\mu. \end{aligned}$$

Подставляя найденные решения обратно, получим для поправок второго порядка:

$$\begin{aligned} h_{\mu\nu}h^{\mu\nu} - \frac{1}{4}(h^\mu_\mu)^2 &= 8\delta\mathfrak{e}_1^{(2)} - 2\delta\mathfrak{e}_2^{(2)}, \\ 2h_{\mu\nu}h^{\mu\nu} - \frac{1}{2}(h^\mu_\mu)^2 &= 12\delta\mathfrak{e}_2^{(2)} - 8\delta\mathfrak{e}_1^{(2)} - 8\delta\mathfrak{e}_3^{(2)}, \\ 5h_{\mu\nu}h^{\mu\nu} + \frac{7}{4}(h^\mu_\mu)^2 &= 8\delta\mathfrak{e}_3^{(2)} - 2\delta\mathfrak{e}_2^{(2)} \end{aligned}$$

с очевидным решением

$$\begin{aligned} \delta\mathfrak{e}_1^{(2)} &= \frac{3}{8}h_{\mu\nu}h^{\mu\nu}, \\ \delta\mathfrak{e}_2^{(2)} &= h_{\mu\nu}h^{\mu\nu} + \frac{1}{8}(h^\mu_\mu)^2, \\ \delta\mathfrak{e}_3^{(2)} &= \frac{7}{8}h_{\mu\nu}h^{\mu\nu} + \frac{1}{4}(h^\mu_\mu)^2. \end{aligned}$$

Как видим, формулы (3.13) – (3.16) успешно воспроизведены. Подстановка в действие (3.38) даёт теорию Фирца-Паули.

3.5.3 Об уравнениях движения

Сделаем несколько замечаний о применении нового метода в более общих ситуациях. Для краткости обозначим $e_i(g^{-1}\eta) \equiv p_i$, а симметрические полиномы квадратного корня будем теперь обозначать простой буквой e, что дает

$$\sum_{k+m=2n} (-1)^k e_k e_m = (-1)^n p_n, \tag{3.43}$$

или более явно в четырёх измерениях:

$$\begin{aligned} e_1^2 - 2e_2 &= p_1, &(3.44)\\ e_2^2 - 2e_1e_3 + 2e_4 &= p_2, &(3.45)\\ e_3^2 - 2e_2e_4 &= p_3, &(3.46)\\ e_4^2 &= p_4. &(3.47) \end{aligned}$$

Лагранжиан взаимодействия имеет вид $\sum \beta_i e_i$, а вариацию можно выполнить как

$$\frac{\partial}{\partial g^{\mu\nu}}\left(\sqrt{-g}e_i\right) = \frac{1}{2}\sqrt{-g}\left(-e_i g_{\mu\nu} + 2\sum_m \frac{\partial e_i}{\partial p_m}\frac{\partial p_m}{\partial g^{\mu\nu}}\right). \tag{3.48}$$

В случае, когда корректно определена теория в терминах матриц, эти уравнения очевидно эквивалентны уравнениям (3.27) в теминах $\mathbb{Y}$. Првеpить это явно, однако, не так просто.

* * *

Рассмотрим для иллюстрации двумерный случай. В таком случае имеем (3.43):

$$\begin{aligned} e_1^2 - 2e_2 &= p_1,\\ e_2^2 &= p_2, \end{aligned}$$

и следовательно

$$\frac{\partial p}{\partial e} = 2\begin{pmatrix} e_1 & -1 \\ 0 & e_2 \end{pmatrix}; \qquad \frac{\partial e}{\partial p} = \frac{1}{2e_1e_2}\begin{pmatrix} e_2 & 1 \\ 0 & e_1 \end{pmatrix}.$$

Заметим, что теория возмущений будет работать просто и однозначно только если $\det\left(\frac{\partial p}{\partial e}\right) \neq 0$. Легко видеть, что это так для стандартного корня из единичной матрицы, но не для необычного с одной $+1$ и одной -1. В следующем разделе мы увидим, что необычные корни связаны именно с ситуациями вырождения якобиана преобразования между e и p, поэтому и теория возмущений оказывается нетривиальной.

Простым вычислением получаем вклад в уравнения движения от вариации двумерного β_1 потенциала

$$\frac{\partial}{\partial g^{\mu\nu}}\left(\sqrt{-g}e_1\right) = \frac{1}{2}\sqrt{-g}\left(\frac{e_2+p_1}{e_1e_2}\eta_{\mu\nu} - \frac{1}{e_1e_2}\eta_{\mu\alpha}g^{\alpha\beta}\eta_{\beta\nu} - e_1 g_{\mu\nu}\right).$$

Умножая на g^{-1}, мы видим, что надо убедиться (тогда получим $\mathbb{Y}_{(1)}$), что $\mathcal{M}^2 = g^{-1}\eta$ для матрицы

$$\mathcal{M} = \frac{e_2+p_1}{e_1e_2}g^{-1}\eta - \frac{1}{e_1e_2}(g^{-1}\eta)^2 = \frac{1}{e_1e_2}\left(e_2 g^{-1}\eta + e_2^2\mathbb{I}\right),$$

где в последнем равенстве мы использовали теорему Гамильтона-Кэли, $\mathbb{X}^2 - p_1\mathbb{X} + p_2\mathbb{I} = \mathbb{O}$, для 2×2 матрицы $\mathbb{X} = g^{-1}\eta$. Проверить требуемое утверждение несложно прямым возведением в квадрат с последующим применением теоремы Гамильтона-Кэли ещё раз.

* * *

В более высоких размерностях всё становится очень громоздко, и мы опускаем соответствующие выражения. Приведём лишь, для последующих ссылок, значение якобиана преобразования между симметрическими полиномами в четырёхмерном

$$\det\left(\frac{\partial p}{\partial e}\right)_{4\times 4} = 2e_4(e_1e_2e_3 - e_3^2 - e_1^2 e_4) \tag{3.49}$$

и трёхмерном

$$\det\left(\frac{\partial p}{\partial e}\right)_{3\times 3} = 2e_3(e_1e_2 - e_3) \qquad (3.50)$$

случаях.

3.6 Возмущения вокруг нестандартных корней

Теперь рассмотрим возможности применения нового метода там, где стандартный подход не работает – для возмущений вокруг нестандартных корней с континуальной свободой. Отметим сразу, что в случае нестандартных корней требуется большая осторожность с биметрической симметрией (3.25). При нечётном числе отрицательных собственных значений корня имеем $\det\left(\sqrt{\eta^{-1}g}\right) < 0$, поэтому надо писать $\beta_i \leftrightarrow -\beta_{D-i}$ вместо $\beta_i \leftrightarrow \beta_{D-i}$, если считаем $\sqrt{-g}$ положительным. Это не великая сложность, но, к сожалению, на ней проблемы не заканчиваются. Однако, обо всём по порядку.

3.6.1 Замечания о множестве квадратных корней

Как мы уже выяснили, если все жордановы клетки матрицы имеют различные собственные значения, то в выборе квадратного корня присутствует только дискретная симметрия, соответствующая выбору знака каждого из чисел $\sqrt{\lambda}$. При наличии же совпадающих собственных значений возможна континуальная симметрия, которая возникает при выборе разных квадратных корней для совпадающих собственных значений.

На эту свободу можно ещё смотреть с той точки зрения, что преобразования подобия реализуют присоединённое представление группы $GL(n,\mathbb{C})$, и при этом континуальная свобода означает, что неподвижная подгруппа исходной матрицы $\mathfrak{S}_{\mathbb{X}} \equiv \{\mathcal{S} \in GL(n,\mathbb{C}) : \mathfrak{L}_{\mathcal{S}}\mathbb{X} = \mathbb{X}\}$ оказалась шире, чем неподвижная подгруппа выбранного квадратного корня, $\mathfrak{S}_{\sqrt{\mathbb{X}}} \neq \mathfrak{S}_{\mathbb{X}}$. Тогда множество корней с фиксированным выбором знаков при $\sqrt{\lambda}$ приобретает естественную структуру многообразия, а точнее однородного пространства $\mathfrak{S}_{\mathbb{X}}/\mathfrak{S}_{\sqrt{\mathbb{X}}}$.

Условия вещественности

До сих пор мы не учитывали требований вещественности. Между тем, это существенный аспект. Например, для вещественной матрицы $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ нормальная форма имеет комплексный вид $\begin{pmatrix} a+ib & 0 \\ 0 & a-ib \end{pmatrix}$. В этом случае мы вынуждены выбрать квадратные корни $\sqrt{a \pm ib}$ комплексно сопряженными друг другу, чтобы обратное преобразование подобия могло вернуть полученную матрицу к вещественному виду.

Также надо иметь в виду, что если вещественная матрица $\mathbb{X}$ имеет отрицательные собственные значения, то корни из них чисто мнимые. Однако же, если они идут парами с идентичными жордановыми клетками (то есть имеется чётное число жордановых клеток любого данного размера с данным отрицательным собственным числом), то мнимые числа $\sqrt{\lambda}$ можно выбрать в виде комплексно сопряженных пар, и тогда квадратный корень будет иметь вещественное представление. См. также работу [130] с более подробным обсуждением геометрического и физического смысла условий вещественности.

Проблема с вариациями

В стандартном подходе предполагается, что $\delta\sqrt{\mathbb{X}} \equiv \sqrt{\mathbb{X} + \delta\mathbb{X}} - \sqrt{\mathbb{X}}$ существует и корректно определён как бесконечно малая вариация матрицы $\sqrt{\mathbb{X}}$. Более того, для возмущений вокруг решения с $g^{-1}f = \mathbb{I}$, можно вообще разложить все выражения в простой ряд Тейлора, предполагая, что выбран тривиальный корень из единичной матрицы. Однако, в противном случае $\delta\mathbb{X}$ вообще говоря не коммутирует с $\sqrt{\mathbb{I}}$, поэтому не удаётся записать привычное разложение $\sqrt{\mathbb{I} + \delta\mathbb{X}}$ по степеням $\delta\mathbb{X}$.

В случае дискретной симметрии это не более чем техническая сложность (впрочем, вообще не возникающая в нашем методе): надо просто найти правильный способ построить степенное разложение. В частности, линейная поправка удовлетворяет уравнению типа Сильвестра

$\sqrt{\mathbb{X}}\cdot\delta\sqrt{\mathbb{X}}+\delta\sqrt{\mathbb{X}}\cdot\sqrt{\mathbb{X}} = \delta\mathbb{X}$, и проблема состоит в том, чтобы решить его. Оказывается, что теорию возмущений можно построить в явном виде с помощью подходящих переопределений полевых переменных [134].

Континуальная симметрия, как мы уже обсуждали, более проблематична. Рассмотрим, ради явного примера, единичную 2×2 матрицу $\mathbb{I} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$. У неё есть два изолированных квадратных корня, $\sqrt{\mathbb{I}}_{++} = \mathbb{I} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$ и $\sqrt{\mathbb{I}}_{--} = -\mathbb{I} = \begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}$, а также два (связных) двумерных многообразия корней $\sqrt{\mathbb{I}}_{ab} = \mathbb{I} = \begin{pmatrix}a & b_1\\ b_2 & -a\end{pmatrix}$, где $a^2 = 1 - b_1 b_2$. Однако, даже бесконечно малое возмущение оставляет только четыре изолированных корня. В самом деле, для $\mathbb{M}_\epsilon = \begin{pmatrix}1 & 0\\ 0 & 1+\epsilon\end{pmatrix}$ имеем $\sqrt{\mathbb{M}_\epsilon}_{++} = \begin{pmatrix}1 & 0\\ 0 & \sqrt{1+\epsilon}\end{pmatrix}$, $\sqrt{\mathbb{M}_\epsilon}_{--} = \begin{pmatrix}-1 & 0\\ 0 & -\sqrt{1+\epsilon}\end{pmatrix}$, $\sqrt{\mathbb{M}_\epsilon}_{+-} = \begin{pmatrix}1 & 0\\ 0 & -\sqrt{1+\epsilon}\end{pmatrix}$ и $\sqrt{\mathbb{M}_\epsilon}_{-+} = \begin{pmatrix}-1 & 0\\ 0 & \sqrt{1+\epsilon}\end{pmatrix}$.

Интересно отметить, что применение уравнения Сильвестра даёт результат только в том случае, если спектры $\sqrt{\mathbb{X}}$ и $-\sqrt{\mathbb{X}}$ не пересекаются. Очевидно, что это условие нарушается ровно тогда, когда присутствует континуальная симметрия (для двух совпадающих собственных значений выбраны корни разного знака), как и должно быть. Мы теперь переходим к применению нашего метода к этим случаям.

3.6.2 "Игрушечный" пример: двумерный случай

Будем для простоты работать с потенциалом, записанным как $\pm\sum_i \beta_{n-i} e_i(\eta^{-1}g)$, где общий знак определяется чётностью или нечётностью числа отрицательных собственных значений, см. замечания в самом начале раздела 3.6. Для нестандартных корней $\det\left(\frac{\partial p}{\partial e}\right) = 0$, и теория возмущений приобретает необычные черты неаналитичности. Поэтому начнём с простейшего двумерного случая, где всё можно сделать явно и понять, что к чему.

Рассмотрим квадратный корень

$$\sqrt{\mathbb{I}} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Другими словами, в качестве фонового решения выбираем

$$e_1 = 0, \quad e_2 = -1.$$

Тогда получается

$$p_1 = 2 + e_1(\mathcal{H}), \quad p_2 = 1 + e_1(\mathcal{H}) + e_2(\mathcal{H})$$

с матрицей возмущения $\mathcal{H} \equiv \eta^{-1} \cdot \delta g$. Уравнения для симметрических полиномов могут быть решены точно. И при нашем выборе фона они дают

$$\begin{aligned} e_2 = -\sqrt{p_2} &= -\sqrt{1 + e_1(\mathcal{H}) + e_2(\mathcal{H})}, \\ e_1 = \pm\sqrt{p_1 + 2e_2} &= \pm\sqrt{2 + e_1(\mathcal{H}) - 2\sqrt{1 + e_1(\mathcal{H}) + e_2(\mathcal{H})}}. \end{aligned}$$

Видим, что есть два решения для e_1. Более того, в низшем порядке имеем

$$e_1 = \pm\frac{1}{2}\sqrt{\left(e_1(\mathcal{H})\right)^2 - 4e_2(\mathcal{H})} + \mathcal{O}\left(\mathcal{H}^2\right),$$

что зависит от $\mathcal{H}$ как однородная функция первой степени, но не аналитично в нуле. К тому же возникла неоднозначность: имеется два решения. Причина проста. Для общего возмущения $\mathcal{H}$ два собственных значения различны, и знак e_1 зависит от того, какое из них (большее или меньшее) с каким знаком взято.

Понятно, что аналогичные черты должны сохраняться и в более сложных случаях. Число решений должно быть равно количеству способов распределить знаки "плюс" и "минус" между собственными значениями. Мы это проверим явно в трёх и четырёх измерениях.

3.6.3 Трёхмерный случай

В трёх измерениях единственный нетривиальный корень из единичной матрицы равен (с точностью до произвольных преобразований подобия):

$$\sqrt{\eta^{-1}g}\,\Big|_{g=\eta} = \sqrt{\mathbb{I}} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.51}$$

а потенциал взаимодействия принимает вид

$$\mathcal{L}_{\text{int}} = \sqrt{-g}\sum_{n=0}^{2}\beta_n \cdot e_n\left(\sqrt{g^{-1}\eta}\right) = -\sum_{n=0}^{2}\beta_n \cdot e_{3-n}\left(\sqrt{\eta^{-1}g}\right).$$

Уравнения для симметрических полиномов (3.43) дают

$$e_1^2 - 2e_2 \;=\; p_1, \tag{3.52}$$

$$e_2^2 - 2e_1e_3 \;=\; p_2, \tag{3.53}$$

$$e_3^2 \;=\; p_3, \tag{3.54}$$

где $p_i \equiv e_i\left(\eta^{-1}g\right) = e_i\left(\mathbb{I} + \mathcal{H}\right)$:

$$\begin{aligned} p_1 &= 3 + h_1, \\ p_2 &= 3 + 2h_1 + h_2, \\ p_3 &= 1 + h_1 + h_2 + h_3, \end{aligned}$$

а через $h_i \equiv e_i(\mathcal{H})$ обозначены элементарные симметрические полиномы матрицы $\mathcal{H}^{\mu}_{\nu} = \eta^{\mu\alpha} \cdot \delta g_{\alpha\nu}$.

Будем строить теорию возмущений в виде

$$e_k = \sum_{i\geq 0} e_k^{(i)}, \tag{3.55}$$

где $e_k^{(i)}$ обозначает поправку i-ого порядка к k-ому полиному. Нулевые порядки определяются выбранным квадратным корнем из единичной матрицы:

$$e_1^{(0)} = 1, \quad e_2^{(0)} = -1, \quad e_3^{(0)} = -1,$$

так что детерминант в формуле (3.50) очевидно обращается в нуль.

Тем не менее, третий полином (детерминант метрики), конечно же, может быть найден с помощью формулы (3.54):

$$\begin{aligned} e_3 = -1 - \frac{1}{2}h_1 - \frac{1}{2}\left(h_2 - \frac{1}{4}h_1^2\right) - \frac{1}{2}\left(h_3 - \frac{1}{2}h_1h_2 + \frac{1}{8}h_1^3\right) \\ + \frac{1}{4}\left(h_1h_3 + \frac{1}{2}h_2^2 - \frac{3}{4}h_1^2h_2 + \frac{5}{32}h_1^4\right) + \mathcal{O}\left(\mathcal{H}^5\right). \end{aligned} \quad (3.56)$$

Подставим теперь разложения (3.55) и (3.56) в систему уравнений (3.52), (3.53), сохраняя только члены первого порядка:

$$\begin{aligned} 2e_1^{(1)} - 2e_2^{(1)} &= h_1, \\ 2e_1^{(1)} - 2e_2^{(1)} + 2 \cdot \frac{1}{2}h_1 &= 2h_1. \end{aligned}$$

Система, очевидно, вырождена. Мы получаем

$$e_2^{(1)} = e_1^{(1)} - \frac{1}{2}h_1, \quad (3.57)$$

с одной переменной первого порядка, оставшейся неопределённой уравнениями первого порядка.

Учитывая полученное соотношение (3.57), уравнения (3.52), (3.53) в квадратичном порядке снова дают два одинаковых соотношения:

$$\left(e_1^{(1)}\right)^2 + 2e_1^{(2)} - 2e_2^{(2)} = 0,$$

из которых получаем

$$e_2^{(2)} = e_1^{(2)} + \frac{1}{2}\left(e_1^{(1)}\right)^2, \quad (3.58)$$

снова оставив половину переменных данного порядка неопределёнными.

Обратимся теперь к третьему порядку. Первое уравнение (3.52) принимает вид

$$e_2^{(3)} = e_1^{(3)} + e_1^{(1)} e_1^{(2)},$$

и, с его учётом, можно преобразовать второе уравнение (3.53) к виду

$$\left(e_1^{(1)}\right)^3 - \frac{1}{2} h_1 \left(e_1^{(1)}\right)^2 + \left(h_2 - \frac{1}{4} h_1^2\right) e_1^{(1)} + h_3 - \frac{1}{2} h_1 h_2 + \frac{1}{8} h_1^3 = 0, \quad (3.59)$$

который наконец позволяет полностью определить все поправки первого порядка.

Это кубическое уравнение общего вида (три коэффициента зависят от трёх независимых параметров h_i). Наличие трёх решений соответствует трём возможным выборам того, какое из трёх собственных значений возмущающей матрицы будет взято со знаком "минус".

В этом месте полезно убедиться, что всё самосогласованно. Пусть λ_i, $i = 1, 2, 3$ – собственные значения $\mathcal{H}$. Имеем $h_1 = \sum\limits_i \lambda_i$, $h_2 = \sum\limits_{i<j} \lambda_i \lambda_j$, $h_3 = \sum\limits_{i<j<k} \lambda_i \lambda_j \lambda_k = \lambda_1 \lambda_2 \lambda_3$. Используя теорему Виета, легко проверить, что решениями уравнения (3.59) являются $\frac{1}{2}(-\lambda_1 + \lambda_2 + \lambda_3)$, $\frac{1}{2}(-\lambda_2 + \lambda_3 + \lambda_1)$ и $\frac{1}{2}(-\lambda_3 + \lambda_1 + \lambda_2)$. И такими действительно должны быть поправки первого порядка к e_1 (следу) при собственных значениях $\pm\sqrt{1 + \lambda_i}$ и нашем выборе знаков.

После того, как мы воспроизвели всю естественную свободу в определении возмущения, система уравнений (3.52), (3.53) становится невырожденной в следующих порядках, определяя по две величины на каждом шагу. В частности, в четвёртом порядке уравнение (3.52) дает

$$e_2^{(4)} = e_1^{(4)} + e_1^{(1)} e_1^{(3)} + \frac{1}{2} \left(e_1^{(2)}\right)^2. \quad (3.60)$$

В то же время, второе уравнение (3.53) содержит $e_1^{(2)}$ и квадратично и линейно, но после учета соотношения (3.60) оно оказывается чисто линейным уравнением, позволяющим однозначно определить $e_1^{(2)}$. Явные выражения слишком громоздки и мало поучительны, чтобы их приводить.

Теперь можно вычислить первую вариацию лагранжиана:

$$\delta^{(1)}\mathcal{L}_{\text{int}} = \frac{1}{2}\beta_0 h_1 - \beta_1 \left(e_1^{(1)} - \frac{1}{2}h_1\right) - \beta_2 e_1^{(1)}.$$

Чтобы пространство Минковского было решением, надо положить $\delta^{(1)}\mathcal{L}_{\text{int}} = 0$. Поскольку $e_1^{(1)}$, вообще говоря, иррационально при рациональных значениях h_i, придётся независимо обращать в нуль две части, с h_1 и с $e_1^{(1)}$:

$$\beta_0 = -\beta_1, \quad \beta_2 = -\beta_1. \tag{3.61}$$

Интересно отметить, что если подставить пространство Минковского с нашим выбором (3.51) квадратного корня $\sqrt{g^{-1}\eta}$ в наивные уравнения движения (3.27),

$$\mathbb{O} = \sum_{i=0}^{2}(-1)^i \beta_i \mathbb{Y}_{(i)} = \begin{pmatrix} \beta_0 + 2\beta_1 + \beta_2 & 0 & 0 \\ 0 & \beta_0 - \beta_2 & 0 \\ 0 & 0 & \beta_0 - \beta_2 \end{pmatrix},$$

то получается то же самое условие (3.61) на коэффициенты β.

Посмотрим теперь на квадратичный лагранжиан для возмущений,

$$\begin{aligned}\delta^{(2)}\mathcal{L}_{\text{int}} = \frac{1}{2}\beta_0\left(h_2 - \frac{1}{4}h_1^2\right) - \beta_1\left(e_1^{(2)} + \frac{1}{2}\left(e_1^{(1)}\right)^2\right) - \beta_2 e_1^{(2)} \\ = -\frac{\beta_1}{2}\cdot\left(\left(e_1^{(1)}\right)^2 + h_2 - \frac{1}{4}h_1^2\right),\end{aligned}$$

где мы учли условие (3.61) на коэффициенты β. Как видим, он содержит корень $e_1^{(1)}$ кубического уравнения (3.59), и соответственно неаналитичен по малому $\mathcal{H}$. Однако структура его интересна.

Если занумеровать собственные значения $\mathcal{H}$ так, что $e_1^{(1)} = \frac{1}{2}(-\lambda_1 + \lambda_2 + \lambda_3)$, то получается

$$\delta^{(2)}\mathcal{L}_{\text{int}} = -\frac{\beta_1}{2}\cdot\lambda_2\lambda_3.$$

Как уже упоминалось, три ветви теории (по числу решений кубического уравнения) отвечают трём способам выбрать одно (отрицательное) собственное значение из трёх, которое опущено в произведении.

3.6.4 Четыре измерения: первый случай

В четырёх измерениях система уравнений (3.43) имеет вид

$$e_1^2 - 2e_2 = 4 + h_1, \tag{3.62}$$

$$e_2^2 - 2e_1e_3 + 2e_4 = 6 + 3h_1 + h_2, \tag{3.63}$$

$$e_3^2 - 2e_2e_4 = 4 + 3h_1 + 2h_2 + h_3, \tag{3.64}$$

$$e_4^2 = 1 + h_1 + h_2 + h_3 + h_4, \tag{3.65}$$

и у нас есть две возможности выбора $\sqrt{\mathbb{I}}$: с одним отрицательным собственным значением и с двумя.

Начнём рассмотрение с первой возможности

$$\sqrt{\eta^{-1}g}\,\Big|_{g=\eta} = \sqrt{\mathbb{I}} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.66}$$

которая очень похожа на трёхмерный случай. Фоновые значения переменных равны

$$e_1^{(0)} = 2, \quad e_2^{(0)} = 0, \quad e_3^{(0)} = -2, \quad e_4^{(0)} = -1,$$

и легко видеть, что якобиан (3.49) обращается в нуль.

Как всегда, последний полином (3.65) легко найти,

$$e_4 = -1 - \frac{1}{2}h_1 - \frac{1}{2}\left(h_2 - \frac{1}{4}h_1^2\right) - \frac{1}{2}\left(h_3 - \frac{1}{2}h_1h_2 + \frac{1}{8}h_1^3\right) - \frac{1}{2}\left(h_4 - \frac{1}{2}h_1h_3 - \frac{1}{4}h_2^2 + \frac{3}{8}h_1^2h_2 - \frac{5}{64}h_1^4\right) + \mathcal{O}\left(\mathcal{H}^5\right).$$

Оставшиеся уравнения (3.62) - (3.64) дают в первом порядке

$$\begin{aligned}
4e_1^{(1)} - 2e_2^{(1)} &= h_1, \\
4e_1^{(1)} - 4e_3^{(1)} - h_1 &= 3h_1, \\
-4e_3^{(1)} + 2e_2^{(1)} &= 3h_1.
\end{aligned}$$

Система оказывается вырожденной, и всё, что можно определить из неё – это

$$\begin{aligned}
e_2^{(1)} &= 2e_1^{(1)} - \frac{1}{2}h_1, \\
e_3^{(1)} &= e_1^{(1)} - h_1,
\end{aligned}$$

оставляя одну переменную неопределённой.

Во втором порядке оказывается аналогичная вырожденность, и мы находим

$$\begin{aligned}
e_2^{(2)} &= 2e_1^{(2)} + \frac{1}{2}\left(e_1^{(1)}\right)^2, \\
e_3^{(2)} &= e_1^{(2)} + \frac{1}{2}\left(e_1^{(1)}\right)^2 - \frac{1}{2}\left(h_2 - \frac{1}{4}h_1^2\right).
\end{aligned}$$

Вырожденность уходит только в четвёртом порядке, с появлением общего (по числу независимых параметров) уравнения четвёртой степени для $e_1^{(1)}$. Мы не приводим весьма громоздких деталей, к тому же этот случай полностью аналогичен трёхмерному (и, по-видимому, любой размерности с одним изменённым знаком).

В лагранжиане взаимодействия

$$\mathcal{L}_{\text{int}} = \sqrt{-g}\sum_{n=0}^{3}\beta_n \cdot e_n\left(\sqrt{g^{-1}\eta}\right) = -\sum_{n=0}^{3}\beta_n \cdot e_{4-n}\left(\sqrt{\eta^{-1}g}\right)$$

вклад первого порядка даёт

$$\delta^{(1)}\mathcal{L}_{\text{int}} = \frac{1}{2}\beta_0 h_1 - \beta_1\left(e_1^{(1)} - h_1\right) - \beta_2\left(2e_1^{(1)} - \frac{1}{2}h_1\right) - \beta_3 e_1^{(1)}.$$

Опять же, требуем $\delta^{(1)}\mathcal{L}_{\text{int}} = 0$, чтобы пространство Минковского было решением. И снова две части, с h_1 и с $e_1^{(1)}$, должны обратиться в нуль независимо,

$$\beta_0 + 2\beta_1 + \beta_2 = 0, \quad \beta_1 + 2\beta_2 + \beta_3 = 0. \tag{3.67}$$

Как и в $3D$, можно формально подставить пространство Минковского с выбором корня (3.66) в наивные уравнения движения

$$\sum_{i=0}^{3} (-1)^i \beta_i \mathbb{Y}_{(i)} = \begin{pmatrix} \beta_0 + 3\beta_1 + 3\beta_2 + \beta_3 & 0 \\ 0 & (\beta_0 + \beta_1 - \beta_2 - \beta_3) \cdot \mathbb{I}_{3\times 3} \end{pmatrix} = \mathbb{O}$$

и получить правильное условие (3.67) для параметров β.

Посмотрим теперь на квадратичный лагранжиан для возмущений:

$$\begin{aligned}
\delta^{(2)}\mathcal{L}_{\text{int}} = \frac{1}{2}\beta_0 \left(h_2 - \frac{1}{4}h_1^2\right) - \beta_1 \left(e_1^{(2)} + \frac{1}{2}\left(e_1^{(1)}\right)^2 - \frac{1}{2}\left(h_2 - \frac{1}{4}h_1^2\right)\right) \\
- \beta_2 \left(2e_1^{(2)} + \frac{1}{2}\left(e_1^{(1)}\right)^2\right) - \beta_3 e_1^{(2)} \\
= -\frac{\beta_1 + \beta_2}{2} \cdot \left(\left(e_1^{(1)}\right)^2 + h_2 - \frac{1}{4}h_1^2\right).
\end{aligned}$$

Интересно, что если пронумеровать собственные знеачения $\mathcal{H}$ так, что $e_1^{(1)} = \dfrac{1}{2}(-\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4)$, то получается

$$\delta^{(2)}\mathcal{L}_{\text{int}} = -\frac{\beta_1 + \beta_2}{2} \cdot (\lambda_2\lambda_3 + \lambda_1\lambda_2 + \lambda_1\lambda_3).$$

Структура этого действия повторяет ту, что приводит к действию Фирца-Паули, но только с укороченным набором собственных значений.

Если бы мы взяли на матричном языке матрицу (3.66) в качестве квадратного корня, то для ограниченного класса возмущений, для которого $h_{1i} = 0$ при $i \geqslant 2$, возмущение коммутировало бы с выбранным корнем из единичной матрицы, и тогда можно было бы построить ряд Тейлора для возмущённой матрицы, получив квадратичный лагранжиан в виде $\sum\limits_{i,j\geqslant 2} h_{ij}h^{ij} - (\sum\limits_{i\geqslant 2} h_i^i)^2$. Тот же ответ воспроизводится и у нас. Однако, для произвольного возмущения появляется неаналитичность в виде решения уравнения четвёртого порядка, и число независимых решений отвечает разным способам выбрать одно собственное число из четырёх.

Любопытный факт, что наивные уравнения движения дают корректный результат, когда приводящий к ним подход вообще не должен работать, возможно, связан с тем, что рассматриваемая ситуация является пределом космологических пространств с матрицей вида $\mathrm{diag}\,(N(t), a(t), \ldots, a(t))$, вокруг которой возмущения хорошо определены для корня $\mathrm{diag}\,(-N, a, \ldots, a)$ при $N \neq a$. Изотропный предел оказывается сингулярным, с появлением описанной выше неаналитичности, но на языке матриц он вообще корректно не определён (ср. с работой [125]).

3.6.5 Четыре измерения: второй случай

Другой возможный выбор квадратного корня в четырёхмерном случае может быть представлен матрицей

$$\sqrt{\eta^{-1}g}\,\Big|_{g=\eta} = \sqrt{\mathbb{I}} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.68}$$

для которой фоновые значения симметрических полиномов равны

$$e_1^{(0)} = 0, \quad e_2^{(0)} = -2, \quad e_3^{(0)} = 0, \quad e_4^{(0)} = 1.$$

Конечно, при них обращается в нуль детерминант в формуле (3.49). Но последний полином (3.47) легко определяется как

$$e_4 = 1 + \frac{1}{2}h_1 + \frac{1}{2}\left(h_2 - \frac{1}{4}h_1^2\right) + \frac{1}{2}\left(h_3 - \frac{1}{2}h_1 h_2 + \frac{1}{8}h_1^3\right) \\ + \frac{1}{2}\left(h_4 - \frac{1}{2}h_1 h_3 - \frac{1}{4}h_2^2 + \frac{3}{8}h_1^2 h_2 - \frac{5}{64}h_1^4\right) + \mathcal{O}\left(\mathcal{H}^5\right).$$

Уравнения (3.44) – (3.46) в первом порядке оказываются особенно просты и вырожденны:

$$\begin{aligned}
-2e_2^{(1)} &= h_1, \\
-4e_2^{(1)} + h_1 &= 3h_1, \\
-2e_2^{(1)} + 2h_1 &= 3h_1.
\end{aligned}$$

Из них мы можем только заключить, что

$$e_2^{(1)} = -\frac{1}{2}h_1,$$

оставляя две другие переменные неизвестными.

Во втором порядке уравнения (3.44) – (3.46) дают

$$\begin{aligned}
\left(e_1^{(1)}\right)^2 - 2e_2^{(2)} &= 0, \\
-4e_2^{(2)} - 2e_1^{(1)}e_3^{(1)} &= 0, \\
\left(e_3^{(1)}\right)^2 - 2e_2^{(2)} &= 0,
\end{aligned}$$

и мы получаем

$$\begin{aligned}
e_2^{(2)} &= \frac{1}{2}\left(e_1^{(1)}\right)^2, \\
e_3^{(1)} &= -e_1^{(1)}.
\end{aligned}$$

В третьем порядке первое уравнение (3.44) принимает вид

$$e_3^{(2)} = e_1^{(1)}e_1^{(2)},$$

и с его учетом второе (3.45) и третье (3.46) уравнения оказываются эквивалентными другу и дают нетривиальный результат:

$$2e_1^{(1)}\left(e_1^{(2)} + e_3^{(2)}\right) = h_3 - \frac{1}{2}h_1h_2 + \frac{1}{8}h_1^3 - \frac{1}{2}h_1\left(e_1^{(1)}\right)^2. \tag{3.69}$$

В этом месте можно начать беспокоиться о самосогласованности вычислений. Что произойдет, если $e_1^{(1)} = 0$? В самом деле, имея четыре собственных значения λ_i в матрице возмущения, при нашем выборе знаков возможны шесть (число выборов двух чисел из четырёх) значений величины $e_1^{(1)}$ вида $\frac{1}{2}\left(\lambda_1+\lambda_2-\lambda_3-\lambda_4\right)$. Может ли полином третьего порядка $h_3-\frac{1}{2}h_1h_2+\frac{1}{8}h_1^3$ без остатка делиться на все шесть из них? Оказывается, может, поскольку эти шесть значений являются тремя парами величин, отличающихся друг от друга лишь знаком. И можно легко проверить, что

$$
\begin{aligned}
&h_3-\frac{1}{2}h_1h_2+\frac{1}{8}h_1^3\\
&\qquad=\frac{1}{8}\left(\lambda_1+\lambda_2-\lambda_3-\lambda_4\right)\left(\lambda_1+\lambda_3-\lambda_2-\lambda_4\right)\left(\lambda_1+\lambda_4-\lambda_2-\lambda_3\right),
\end{aligned}
$$

и при этом

$$
e_1^{(2)}+e_3^{(2)}=\frac{1}{2}\left(\lambda_3\lambda_4-\lambda_1\lambda_2\right). \tag{3.70}
$$

В четвёртом порядке вырожденность системы уравнений (3.44) - (3.46) снимается, и первое уравнение (3.44) принимает вид

$$
e_2^{(4)}=e_1^{(1)}e_1^{(3)}+\frac{1}{2}\left(e_1^{(2)}\right)^2. \tag{3.71}
$$

Два других уравнения очень громоздки. Мы не будем их явно выписывать. Однако, оказывается, что если вычесть второе (3.45) уравнение из третьего (3.46) и использовать первое (3.71), то получается замечательное выражение, которое позволяет наконец полностью определить первый порядок:

$$
h_4-\frac{1}{4}\left(\left(e_1^{(1)}\right)^2+h_2-\frac{1}{4}h_1^2\right)^2+\left(e_1^{(2)}+e_3^{(2)}\right)^2=0. \tag{3.72}
$$

Вместе с предыдущим результатом (3.69) для $e_1^{(2)}+e_3^{(2)}$, оно приводит к кубическому уравнению для $\left(e_1^{(1)}\right)^2$, как и должно было быть.

Более того, принимая опять, что $e_1^{(1)}=\frac{1}{2}\left(\lambda_1+\lambda_2-\lambda_3-\lambda_4\right)$, легко получаем

$$
\frac{1}{2}\left(\left(e_1^{(1)}\right)^2+h_2-\frac{1}{4}h_1^2\right)=\frac{1}{2}\left(\lambda_3\lambda_4+\lambda_1\lambda_2\right) \tag{3.73}
$$

и, учитывая ещё (3.70), видим, что правильные (в смысле соответствия собственным числам) значения $e_1^{(1)}$ действительно решают уравнение (3.72).

Для лагранжиана же взаимодействия

$$\mathcal{L}_{\text{int}} = \sqrt{-g}\sum_{n=0}^{3}\beta_n \cdot e_n\left(\sqrt{g^{-1}\eta}\right) = \sum_{n=0}^{3}\beta_n \cdot e_{4-n}\left(\sqrt{\eta^{-1}g}\right)$$

в первом порядке получаем

$$\delta^{(1)}\mathcal{L}_{\text{int}} = \frac{1}{2}\beta_0 h_1 - \beta_1 e_1^{(1)} - \frac{1}{2}\beta_2 h_1 + \beta_3 e_1^{(1)},$$

и из условия $\delta^{(1)}\mathcal{L}_{\text{int}} = 0$ выводим

$$\beta_0 = \beta_2, \quad \beta_1 = \beta_3. \tag{3.74}$$

В этом месте нас, возможно, уже не должно удивлять, что наивное уравнение движения $\sum\limits_{i=0}^{3}(-1)^i\beta_i\mathbb{Y}_{(i)} = 0$ формально даёт

$$\begin{pmatrix} (\beta_0+\beta_1-\beta_2-\beta_3)\cdot\mathbb{I}_{2\times 2} & \mathbb{O} \\ \mathbb{O} & (\beta_0-\beta_1-\beta_2+\beta_3)\cdot\mathbb{I}_{2\times 2}\end{pmatrix} = \mathbb{O}$$

с полученным выше результатом (3.74) для параметров β.

Для возмущений лагранжиан второго порядка принимает вид

$$\begin{aligned}\delta^{(2)}\mathcal{L}_{\text{int}} &= \frac{1}{2}\beta_0\left(h_2 - \frac{1}{4}h_1^2\right) + \beta_1 e_3^{(2)} + \frac{1}{2}\beta_2\left(e_1^{(1)}\right)^2 + \beta_3 e_1^{(2)} \\ &= \frac{\beta_1}{2}\cdot\left(\frac{h_3 - \frac{1}{2}h_1h_2 + \frac{1}{8}h_1^3}{e_1^{(1)}} - \frac{1}{2}h_1 e_1^{(1)}\right) + \frac{\beta_2}{2}\cdot\left(\left(e_1^{(1)}\right)^2 + h_2 - \frac{1}{4}h_1^2\right),\end{aligned}$$

который с учетом (3.70) и (3.73) эквивалентен замечательному выражению

$$\delta^{(2)}\mathcal{L}_{\text{int}} = \frac{\beta_2+\beta_1}{2}\cdot\lambda_3\lambda_4 + \frac{\beta_2-\beta_1}{2}\cdot\lambda_1\lambda_2.$$

По сути, полученная структура воспроизводит действие Фирца-Паули, но по отдельности для двух групп индексов. Если бы рассматривалось возмущение, коммутирующее с выбранным корнем из единичной матрицы ($h_{ij} = 0$ для $i \leqslant 2$ и $j \geqslant 3$ и наоборот), то получилось бы два слагаемых типа Фирца-Паули, $\sum\limits_{i,j\leqslant 2} h_{ij}h^{ij} - (\sum\limits_{i\leqslant 2} h^i_i)^2$ и $\sum\limits_{i,j\geqslant 3} h_{ij}h^{ij} - (\sum\limits_{i\geqslant 3} h^i_i)^2$. То же самое воспроизводит в этом случае и наша формула.

В целом же имеется шесть возможных ветвей решения, по числу выборов двух чисел из четырёх. При этом алгебраическая сложность выражений не превосходит таковой для корней уравнения четвёртого порядка, поскольку всё определяется собственными значениями матрицы 4×4. Напомним, кстати, что решение общего уравнения четвёртого порядка сводится к решению кубического и квадратного уравнений.

Отметим также, что подобный квадратный корень можно тоже пытаться использовать в космологии, но для регуляризации теории возмущений необходим уже пространственно анизотропный фон. В противном случае космологическая теория возмущений в стандартных терминах вообще не будет корректно определена (ср. опять с работой [125]).

3.6.6 Обсуждение

Таким образом, мы описали структуру квадратных корней из матриц, лежащих в основе современных теорий массивной гравитации. Это определённо не самый простой и приятный объект для работы из тех, что встречаются в теоретической физике. При этом оказывается, что дискретная свобода в выборе корня не приводит к принципиальным трудностям, хотя технически работа может оказаться веьма непростой. Континуальная же свобода делает вообще модель плохо сформулированной, и описание возмущений становится невозможным.

Наш метод работы с элементарными симметрическими полиномами позволяет обойти обе эти сложности, хотя во втором случае в теории возмущений появляются неаналитичности неприятного вида. Однако они неразрывно связаны со свойствами тех величин, что входят в действие, и от них нельзя избавиться в рамках этой модели. По крайней мере, новый метод позволяет хотя бы в принципе обсуждать эти ситуации.

Свойства новой теории возмущений вокруг нестандартных корней мы установили на низкоразмерных примерах (впрочем, включающих физическую размерность). Однако, кажется вероятным, что за ними могут стоять интересные математические структуры, которые ещё предстоит установить. В частности, представляется правдоподобным, что в любой размерности, если обозначить через $\mathcal{I}_-$ и $\mathcal{I}_+$ множества индексов, отвечающих отрицательным и положительным собственным значениям $\sqrt{\mathbb{I}}$ соответственно, то квадратичный потенциал вокруг пространства Минковского примет вид суммы двух слагаемых типа Фирца-Паули:

$$C_1 \cdot \sum_{i<j;\ i,j\in\mathcal{I}_-} \lambda_i\lambda_j + C_2 \cdot \sum_{i<j;\ i,j\in\mathcal{I}_+} \lambda_i\lambda_j.$$

Интересной задачей на будущее будет установить, как это всё соотносится с тетрадным формализмом в массивной гравитации [102].

3.7 Проблемы и обобщения массивной гравитации

На самом деле, одновременно с большим прогрессом в построении бездуховых моделей массивной гравитации [100], стало понятно, что, как всегда, у теории есть множество трудностей. Было много обсуждений в литературе, связанных с проблемой нарушения причинности [135], которую, впрочем, вообще не так легко определить в биметрических картинах. Отметим однако, что в современной теории гравитации вообще не так редко рассматривают модели с нарушением причинности [136]. Существуют также аргументы в пользу того, что квантовые поправки должны охранять наш мир от нарушений причинности [137]. Мы не будем далее обсуждать вопросы причинности и квантования [138].

С нашей точки зрения более важной задачей является построение космологических моделей. К сожалению, оказалось, что массивная гравитация, хоть и является очень интересной возможностью, но не оправдала всех возложенных поначалу на неё надежд. В частности, оказалось очень сложно построить жизнеспособные космологические модели. При опорной метрике Минковского не существует однородных космологических моделей [139], кроме случая отрицательной пространственной кривизны [140], который всё равно оказывается неустойчивым (дух в векторном секторе, поскольку действие для векторных возмущений начинается с третьего порядка; это не дух Боулвара-Дезера) [141]. Переход к биметрической теории улучшает дело [142], но не решает всех проблем. И самым надёжным вариантом оказывается космология, в которой вторая метрика почти отключена значением своей массы Планка [143].

Вместе с тем массивная гравитация является теоретически очень интересной моделью, которая, по крайней мере в некоторых биметрических режимах [143], не противоречит наблюдательным данным, обладает потенциально очень многообещающей динамикой, и даже, вполне возможно, массивный партнёр гравитона мог бы взять на себя роль Тёмной Материи [144]. Отказываться от такой возможности было бы явно преждевременно. Но, с другой стороны, хочется выяснить, можно ли построить космологическую модель, не обращаясь к полностью биметрической конструкции с подавленными эффектами второй метрики [143].

Поэтому рассматривалось множество возможностей расширить модель, от переменой массы гравитона, параметризованной некоторым скалярным полем [145] до модификаций типа $f(R)$ [146]. Очень интересной возможностью оказался вариант расширения с помощью квазидилатона [147], в котором матрица $\sqrt{g^{-1}f}$ умножена на экспоненциальную функцию от значений дополнительного скалярного поля (квазидилатона), а когда и такие космологии столкнулись с проблемами устойчивости возмущений [148], был предложен расширенный квазидилатон [149].

Большинство работ по квазидилатону написано на языке матрицы $\mathbb{I}-\sqrt{g^{-1}f}$ вместо $\sqrt{g^{-1}f}$ (напомним, что это исторически первая запись теории [113], которая удобна для исследований в рамках теории возмущений). Действие расширенного квазидилатона в этой картине выглядит так:

$$S=\frac{M_{Pl}^2}{2}\int d^4x\sqrt{-g}\left(R[g]-\frac{\omega}{M_{Pl}^2}\partial_\mu\sigma\partial^\mu\sigma+2m^2\left(e_2+\alpha_3\,e_3+\alpha_4\,e_4\right)\right), \tag{3.75}$$

где под e_i-ыми подразумеваются элементарные симметрические полиномы от матрицы

$$K^\mu{}_\nu=\delta^\mu{}_\nu-e^{\sigma/M_{Pl}}\left(\sqrt{g^{-1}\tilde{f}}\right)^\mu{}_\nu \tag{3.76}$$

с новой опорной метрикой, равной

$$\tilde{f}_{\mu\nu}=f_{\mu\nu}-\frac{a_\sigma}{M_{Pl}^2m^2}e^{-2\sigma/M_{Pl}}\partial_\mu\sigma\partial_\nu\sigma. \tag{3.77}$$

Именно изменение опорной метрики отличает расширенный квазидалатон от обычного: обращение a_σ в нуль приводит к модели обычного квазидилатона.

В качестве основной опорной метрики $f_{\mu\nu}$, как правило, берётся пространство Минковского: $f_{\mu\nu}=\eta_{\mu\nu}$. Однако же часто при этом формально восстанавливают калибровочную инвариантность введением полей Штюкельберга ϕ^a:

$$f_{\mu\nu}=\eta_{ab}\partial_\mu\phi^a\partial_\nu\phi^b, \tag{3.78}$$

которые по своему смыслу описывают преобразование координат от тех, в которых метрика имеет вид Минковского.

3.7.1 Космологические решения с расширенным квазидилатоном

Для целей следующего раздела кратко опишем космологии в моделях с расширенным квазидилатоном. При этом нас будет интересовать пространственно плоский случай

$$\mathrm{d}s_g^2=-N^2(t)\,dt^2+a^2(t)\delta_{ij}\,dx^idx^j, \tag{3.79}$$

$$\mathrm{d}s_{\tilde{f}}^2=-n^2(t)\,dt^2+\delta_{ij}\,dx^idx^j \tag{3.80}$$

с пространственно однородным квазидилатоном $\sigma = \sigma(t)$. На языке полей Штюкельберга можно написать

$$\phi_0 = \phi_0(t), \quad \phi_i = x_i, \tag{3.81}$$

$$n^2 = \dot{\phi}_0^2 + \frac{a_\sigma}{M_{\rm Pl}^2 m^2} e^{-2\sigma/M_{Pl}} \dot{\sigma}^2. \tag{3.82}$$

Уравнения движения при этом принимают вид (см., например, [150])

$$3H^2 = \Lambda_A + \frac{\omega \dot{\sigma}^2}{2M_{Pl}^2 N^2}, \tag{3.83}$$

$$\frac{2\dot{H}}{N} = \frac{(1-r)\dot{\Lambda}_A}{3HN - 3\dot{\sigma}/M_{Pl}} - \frac{\omega\dot{\sigma}^2}{M_{Pl}^2 N^2}, \tag{3.84}$$

$$\partial_t \left(\frac{1}{n} m^2 M_{Pl}^2 a^4 J(A-1) A\, \dot{\phi}_0 \right) = 0, \tag{3.85}$$

$$m^2 M_{Pl} N^3 A (3(r-1)A(-2+\alpha_3(A-1)) + J(-3 + r(-1+4A))) \\ = \omega(3HN^2\dot{\sigma} + N\ddot{\sigma} - \dot{N}\dot{\sigma}), \tag{3.86}$$

где введены следующие стандартные обозначения:

$$\begin{aligned} A(t) &\equiv \frac{e^{\sigma/M_{Pl}}}{a}, \\ H(t) &\equiv \frac{\dot{a}}{aN}, \\ r(t) &\equiv \frac{na}{N}, \\ J(t) &\equiv 3 + 3(1-A)\,\alpha_3 + (1-A)^2\,\alpha_4, \\ \Lambda_A &\equiv m^2(A-1)[J + (A-1)(\alpha_3\,(A-1) - 3)]. \end{aligned}$$

3.8 Космологические возмущения в модели с расширенным квазидилатоном

Мы теперь хотим показать, что теории с квазидилатоном не просто космологически неустойчивы, но и страдают от присутствия духа

Боулвара-Дезера (на примере космологических возмущений). К сожалению, речь здесь пойдёт об очень громоздких, но весьма прямолинейных вычислениях, который проводились с использованием вычислительной техники. Даже явное выписывание квадратичного действия может занимать несколько страниц, поэтому мы ограничимся формулировкой наиболее важных результатов без вычислительных деталей.

Отметим, что история этого вопроса запутанна. В первой версии препринта Мукоямы [151] содержалось утверждение об отсутствии духа и даже было приведено непертурбативное доказательство, основанное на некорректном выборе калибровки, которое однако подверглось критике в работе [152], впрочем тоже не вполне убедительной. В работе [150] были сделаны утверждения об устойчивости возмущений, но в таких приближениях, которые на самом деле не позволяют сделать уверенного заключения (вырожденность гессиана, полученная в ультрафиолетовом пределе, может просто указывать на разную скорость роста ненулевых собственных значений с импульсом). И наконец, когда мы уже готовили к печати нашу работу [10*], появилась статья [153] с независимым исследованием космологических возмущений, в которой (наряду с другими результатами) был сделан вывод о присутствии духа Боулвара-Дезера, подтверждающийся и нашими более обстоятельными вычислениями. Вскоре после появления работы [153] вышла вторая версия препринта Мукоямы [154], в которой уже доказывалось, что дух в модели есть.

3.8.1 Духовая мода в космологических возмущениях

Рассмотрим космологические возмущения в моделях с расширенным квазидилатоном. Поскольку мы интересуемся скалярным сектором, содержащим дух Боулвара-Дезера, мы игнорируем векторные и тензорные возмущения метрики. Мы использовали следующую стандартную параметризацию возмущений метрики:

$$
\begin{aligned}
\delta g_{00} &= -2\,N^2\,\frac{\Phi}{M_{Pl}},\\
\delta g_{0i} &= N\,a\,\partial_i \frac{B}{M_{Pl}},\\
\delta g_{ij} &= a^2\,\left[2\delta_{ij}\frac{\psi}{M_{Pl}} + \left(\partial_i\partial_j - \frac{\delta_{ij}}{3}\partial^k\partial_k\right)\frac{E}{M_{\rm Pl}}\right],
\end{aligned}
$$

которая, с точностью до переобозначений и принятого в этих исследованиях явного выделения обезразмеривающих масс Планка, отличается от рассмотрений из раздела 1.2.6 только удалением следовой части из E, что часто бывает удобно.

Рассмотрим квадратичное действие вокруг описанных выше космологических решений, в пространственном фурье-образе с импульсом k. Полное выражение чрезвычайно громоздко. Однако нас интересует лишь гессиан кинетического слагаемого после удаления нединамических переменных. Выпишем относящиеся к делу части этого выражения (все слагаемые, квадратичные по скоростям, а также все слагаемые, содержащие нединамические переменные B и Φ):

$$
\begin{aligned}
\mathcal{L}^{(2)} = &-\frac{k^2a^3AB^2Q}{2(-1+A)^2(1+r)} - 2k^2a^2BH\Phi - a^3\Lambda_A\Phi^2\\
&+\frac{1}{2}a^3\left(\omega + \frac{a^2(-1+A)Ja_\sigma\dot{\phi}_0^2}{Ar^3}\right)\dot{\delta\sigma}^2 - \frac{1}{3}k^4a^2B\dot{E} + \frac{1}{12}k^4a^3\dot{E}^2\\
&-\omega a^3\Phi\dot{\delta\sigma}\dot{\sigma} + \left(2k^2a^2B + 6a^3H\Phi\right)\dot{\psi} - 3a^3\dot{\psi}^2 + \ldots, \qquad (3.87)
\end{aligned}
$$

где

$$
Q \equiv -m^2J(A-1) + (\Lambda_A + m^2(A-1)^2)A
$$

и, только для упрощения выражений, мы положили функцию шага равной единице, $N=1$, а также опустили множители M_{Pl}.

Поля Φ и B можно определить, решив алгебраические уравнения. Подставив соответствующие решения в квадратичное действие (3.87)), мы получаем гессиан

$$\det \mathcal{H} \equiv \begin{vmatrix} \frac{\partial^2 \mathcal{L}}{\partial \dot{\psi}^2} & \frac{\partial^2 \mathcal{L}}{\partial \dot{\psi}\, \partial \dot{E}} & \frac{\partial^2 \mathcal{L}}{\partial \dot{\psi}\, \partial (\delta \dot{\sigma})} \\ \frac{\partial^2 \mathcal{L}}{\partial \dot{E}\, \partial \dot{\psi}} & \frac{\partial^2 \mathcal{L}}{\partial \dot{E}^2} & \frac{\partial^2 \mathcal{L}}{\partial \dot{E}\, \partial (\delta \dot{\sigma})} \\ \frac{\partial^2 \mathcal{L}}{\partial (\delta \dot{\sigma})\, \partial \dot{\psi}} & \frac{\partial^2 \mathcal{L}}{\partial (\delta \dot{\sigma})\, \partial \dot{E}} & \frac{\partial^2 \mathcal{L}}{\partial (\delta \dot{\sigma})^2} \end{vmatrix} = -\frac{k^4\, \omega\, a^{13}(-1+A)\, J \dot{\sigma}^2\, Q\, a_\sigma \dot{\phi}_0^2}{2r^3 \left(2k^2(-1+A)^2 H^2 (1+r) - a^2 A Q \Lambda_A\right)}. \tag{3.88}$$

Легко видеть, что $\det \mathcal{H} \neq 0$, кроме случаев $a_\sigma = 0$ (простой квазидилатон), или в так называемом пределе позднего времени (ветвь $J = 0$ [149]), который на самом деле никогда не достигается. Тем самым в теории есть дух.

3.8.2 Подход с использованием полей Штюкельберга

Аналогично можно провести вычисления в терминах полей Штюкельберга, но удобнее оказывается работать в (эквивалентной) β-картине [115], которой мы в основном и пользовались в данной Диссертации:

$$S = \frac{M_{Pl}^2}{2} \int d^4x \sqrt{-g} \left(R[g] - \frac{\omega}{M_{Pl}^2} \partial_\mu \sigma \partial^\mu \sigma + 2m^2 \sum_{n=0}^{4} \beta_n\, e_n(\mathcal{M}) \right), \tag{3.89}$$

где

$$\mathcal{M}^\mu{}_\nu = e^{\sigma/M_{Pl}} \left(\sqrt{g^{-1} \tilde{f}} \right)^\mu{}_\nu . \tag{3.90}$$

В этом подходе рассмотрим только возмущения квазидалатона и полей Штюкельберга. Слагаемые квадратичного действия, содержащие $\delta\dot{\sigma}$

и $\delta\dot\phi_a$ таковы:

$$\mathcal{L}^{(2)} = \frac{e^{-3\sigma/M_{\rm Pl}}}{2a^3m^2M_{\rm Pl}^2n^3N^2(an+N)}$$
$$\times\Bigg(-2e^{\frac{2\sigma}{M_{\rm Pl}}}m^2M_{\rm Pl}N(an+N)\,\Theta\,\delta\dot\sigma\,\dot\sigma\,\delta\dot\phi_0\,\dot\phi_0\,a_\sigma$$
$$+(an+N)\,\delta\dot\sigma^2\left(a^3e^{\frac{3\sigma}{M_{\rm Pl}}}m^2M_{\rm Pl}^2n^3\,\omega+e^{\frac{2\sigma}{M_{\rm Pl}}}m^2M_{\rm Pl}^2n^2N\,\Theta\,a_\sigma-N\,\Theta\,\dot\sigma^2a_\sigma^2\right)$$
$$+e^{\frac{4\sigma}{M_{\rm Pl}}}m^4M_{\rm Pl}^2N\left((an+N)\delta\dot\phi_0{}^2\left(n^2-\dot\phi_0{}^2\right)\Theta-an^3\delta\dot\phi_i{}^2\,\Xi\right)\Bigg)+\dots, \quad (3.91)$$

где

$$\Theta = a^3\beta_1+3a^2e^{\frac{\sigma}{M}}\beta_2+3ae^{\frac{2\sigma}{M}}\beta_3+e^{\frac{3\sigma}{M}}\beta_4,$$
$$\Xi = a^3\beta_1+2a^2e^{\frac{\sigma}{M_{\rm Pl}}}\beta_2+ae^{\frac{2\sigma}{M_{\rm Pl}}}\beta_3.$$

В обозначениях предыдущего подраздела имеем

$$\mathcal{L}^{(2)} = \frac{1}{2N^2}\Bigg(w\delta\dot\sigma^2+\frac{m^2N}{a^3A^3}\Bigg(\frac{a^5A^4Q\delta\dot\phi_i{}^2}{m^2(A-1)^2(an+N)}$$
$$+\frac{a^4(-1+A)A^4\,J\,\delta\dot\phi_0{}^2\left(n^2-\dot\phi_0{}^2\right)}{n^3}$$
$$-\frac{2a^2(-1+A)A^2\,J\,\delta\dot\sigma\,\dot\sigma\,\delta\dot\phi_0\,\dot\phi_0\,a_\sigma}{m^2Mn^3}$$
$$+\frac{(-1+A)J\delta\dot\sigma^2a_\sigma\left(a^2A^2m^2M^2n^2-\dot\sigma^2a_\sigma\right)}{m^4M^2n^3}\Bigg)\Bigg)+\dots, \quad (3.92)$$

а соответствующий гессиан

$$\det\frac{\partial\mathcal{L}^{(2)}}{\partial\{\delta\dot\sigma,\delta\dot\phi_a\}} = \frac{\omega\,a_\sigma\,a^5A^2JQ^3\dot\sigma^2}{(-1+A)^5M_{\rm Pl}^2n^3N^6(an+N)^3}\neq 0, \quad (3.93)$$

откуда очевидно присутствие духа Боулвара-Дезера.

3.8.3 Заключительные замечания

Тем самым мы подтверждаем присутствие духа Боулвара-Дезера в теории расширенного квазидилатона, и этот вывод на сегодня принимается основными специалистами по этому виду массивной гравитации. Отметим, что проблемой исходного "доказательства" отсутствия духа [151] было использование $\phi^0 = -e^{-\sigma}$ в качестве калибровки. Но, поскольку у двух скалярных полей поверхности постоянных значений вообще говоря не совпадают, то это конечно же не калибровка, в условие, уменьшающее число степеней свободы.

Таким образом, скалярно-тензорные модификации не обеспечили возможности построения жизнеспособной космологической модели в рамках чисто массивной (не полностью биметрической) гравитации, не говоря уже о решении космологических проблем и загадок. (Впрочем, можно рассматривать квазидилатон с расширением, но без собственного кинетического слагаемого [155].) Биметрические варианты выглядят более перспективными, и с нашей точки зрения сама дРГТ структура взаимодействия несомненно заслуживает дальнейшего теоретического изучения.

Глава 4

Другие биметрические теории гравитации

Здесь мы рассматриваем более общие биметрические теории, не связанные с массивной гравитацией. Полученные в данной главе результаты опубликованы в наших работах [11*], [12*], [13*].

С современной точки зрения работа с биметрическими теориями за пределами дРГТ версий может вызывать недоумение, поскольку есть аргументы в пользу того, что общая теория относительности является единственным жизнеспособным кандидатом на роль кинетического слагаемого для частиц спина 2. С точки зрения массивной гравитации подобные аргументы представлены в [156, 157].

Тем не менее, любые no-go теоремы хороши ровно настолько, насколько и принятые при их доказательстве предположения. Часто в таких работах неявно предполагается, что гамильтонов анализ проходит обычным образом с нединамическими переменными шага и сдвига, и так далее. Вместе с тем, различные биметрические формализмы находят интересные приложения, вплоть до биметрической реализации парадигмы МОНД [158], которая, кстати, недавно получила интересное развитие [159].

Кроме того, вообще говоря, нет никаких причин, почему метрическая и аффинная структуры (грубо говоря, мера и параллельный перенос) на многообразии должны быть связаны друг с другом условием Леви-

Чивита. Связность может быть независимой переменной, или переменной, которая связана с метрикой каким-нибудь другим способом. Одной из интересных идей в этом направлении стало использование вспомогательной метрики для порождения связности [160] в форме Леви-Чивита.

В наших работах [11*] и [12*] эта идея была развита, причём оказалось, что при наличии совершенно независимой второй метрики широкий класс возможных теорий действительно оказывается содержащим духи. Однако наложение дополнительных условий на связь двух метрик может приводить к жизнеспособным моделям. Также в работе [13*] мы рассмотрели с этой точки зрения интересный класс теорий [161] с соотношением между метриками, зависящим от кривизны. Оказалось, что этот класс, стого говоря, нуждается в доопределении, но может иметь интересные связи с нелокальными теориями гравитации.

4.1 Биметрический вариационный принцип

В этом разделе мы рассматриваем теории, заданные действием

$$S = \int d^4x \sqrt{-g} g^{\mu\nu} \bar{R}_{\mu\nu} + S_{matter}(g, \ldots), \tag{4.1}$$

где $\bar{R}_{\mu\nu}$ – тензор Риччи независимой метрики $\bar{g}_{\mu\nu}$, которая определяет свою аффинную связность, а метрика g определяет меру интегрирования, способ перемещения значков, а также взаимодействие с материей. (В принципе, можно рассмотреть даже возможность наличия антисимметричной части у $\bar{g}_{\mu\nu}$, и это описано в нашей работе [11*], но здесь мы опустим эту возможность, поскольку она выглядит геометрически менее мотивированной.) Подобная модель была названа в нашей работе [11*] , биметрическим вариационным принципом.

Для изучения спектра теории рассмотрим предел, в котором обе метрики близки к метрике Минковского. При этом нам будет удобнее рассматривать не флуктуации обеих метрик, а флуктуации физической метрики и отклонения вспомогательной метрики от нее:

$$\bar{g}_{\mu\nu} = g_{\mu\nu} + \bar{h}_{\mu\nu}$$

где $\bar{h}_{\mu\nu}$ рассматривается как малое возмущение.

Заметим, что обратная (вспомогательная) метрика может быть найдена как

$$\bar{g}^{\mu\nu} = \left[\left(g + \bar{h}\right)^{-1}\right]^{\mu\nu} = \left[\left(\mathbb{I} + g^{-1}\bar{h}\right)^{-1}\right]^{\mu}_{\alpha} g^{\alpha\nu} = \left[\sum_{n=0}^{\infty}\left(-g^{-1}\bar{h}\right)^{n}\right]^{\mu}_{\alpha} g^{\alpha\nu}, \tag{4.2}$$

где $\left(-g^{-1}\bar{h}\right)^{0} \equiv \mathbb{I}$, а требование того, чтобы степенной ряд был сходящимся, придаёт точный смысл малости $\bar{h}$.

Находим теперь коэффициенты связности

$$\begin{aligned}\bar{\Gamma}^{\mu}_{\alpha\beta} &= \frac{1}{2}\bar{g}^{\mu\nu}\left(\partial_{\alpha}\bar{g}_{\nu\beta} + \partial_{\beta}\bar{g}_{\nu\alpha} - \partial_{\nu}\bar{g}_{\alpha\beta}\right) \\ &= \frac{1}{2}\left[\sum_{n=0}^{\infty}\left(-g^{-1}\bar{h}\right)^{n}\right]^{\mu}_{\rho}\left(2\Gamma^{\rho}_{\alpha\beta} + g^{\rho\nu}\left(\partial_{\alpha}\bar{h}_{\nu\beta} + \partial_{\beta}\bar{h}_{\nu\alpha} - \partial_{\nu}\bar{h}_{\alpha\beta}\right)\right).\end{aligned}$$

В нулевом порядке по $\bar{h}$ получается, конечно же, $\Gamma^{\mu}_{\alpha\beta}$. Аккуратное рассмотрение следующих слагаемых показывает, что

$$\bar{\Gamma}^{\mu}_{\alpha\beta} = \Gamma^{\mu}_{\alpha\beta} + \frac{1}{2}\left[\sum_{n=0}^{\infty}\left(-g^{-1}\bar{h}\right)^{n}\right]^{\mu}_{\rho} g^{\rho\nu}\left(\nabla_{\alpha}\bar{h}_{\nu\beta} + \nabla_{\beta}\bar{h}_{\nu\alpha} - \nabla_{\nu}\bar{h}_{\alpha\beta}\right), \tag{4.3}$$

где ковариантные производные относятся к леви-чивитовской связности метрики g. Как видим, получается интересное свойство, что

$$\delta\Gamma^{\mu}_{\alpha\beta} = \frac{1}{2}\bar{g}^{\mu\nu}\left(\nabla_{\alpha}\bar{h}_{\nu\beta} + \nabla_{\beta}\bar{h}_{\nu\alpha} - \nabla_{\nu}\bar{h}_{\alpha\beta}\right),$$

которое, впрочем, как нетрудно убедиться прямой проверкой, справедливо и без предположения о сходимости степенного ряда для обратной метрики. Из полученного выражения в частности получаем

$$\delta\Gamma^{\mu}_{\alpha\mu} = \frac{1}{2}\bar{g}^{\mu\nu}\nabla_{\alpha}\bar{h}_{\mu\nu}.$$

Заметим, что если оборвать ряд для обратной метрики на некотором конечном порядке N, то сумма ряда не будет симметричной по перестановке индексов. Однако, антисимметричная часть всегда имеет следующий порядок малости по $\bar{h}$. В самом деле,

$$\left[\sum_{n=0}^{N}\left(-g^{-1}\bar{h}\right)^n\right]^{\mu}_{\alpha} g^{\alpha\nu} = \left[\left[\sum_{n=0}^{\infty}\left(-g^{-1}\bar{h}\right)^n\right]^{\mu}_{\alpha} - \left[\sum_{n=N+1}^{\infty}\left(-g^{-1}\bar{h}\right)^n\right]^{\mu}_{\alpha}\right] g^{\alpha\nu}$$
$$= \left[\mathbb{I} - \left(-g^{-1}\bar{h}\right)^{N+1}\right]\left[\sum_{n=0}^{\infty}\left(-g^{-1}\bar{h}\right)^n\right] g^{-1} = \left[\mathbb{I} - \left(-g^{-1}\bar{h}\right)^{N+1}\right]^{\mu}_{\alpha} g^{\alpha\nu}.$$

Далее, необходимо найти тензор Риччи

$$\bar{R}_{\mu\nu} = \partial_\alpha \bar{\Gamma}^{\alpha}_{\mu\nu} - \partial_\nu \bar{\Gamma}^{\alpha}_{\mu\alpha} + \bar{\Gamma}^{\alpha}_{\beta\alpha}\bar{\Gamma}^{\beta}_{\mu\nu} - \bar{\Gamma}^{\alpha}_{\mu\beta}\bar{\Gamma}^{\beta}_{\nu\alpha}.$$

Подставляя $\bar{\Gamma}^{\alpha}_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu} + \delta\Gamma^{\alpha}_{\mu\nu}$, получаем

$$\bar{R}_{\mu\nu} = R_{\mu\nu} + \nabla_\alpha \delta\Gamma^{\alpha}_{\mu\nu} - \nabla_\nu \delta\Gamma^{\alpha}_{\mu\alpha} + \delta\Gamma^{\alpha}_{\beta\alpha}\delta\Gamma^{\beta}_{\mu\nu} - \delta\Gamma^{\alpha}_{\mu\beta}\delta\Gamma^{\beta}_{\nu\alpha}.$$

Слагаемые с ковариантными производными могут быть отброшены при подстановке в действие, ибо дадут лишь поверхностные члены. Впрочем, при необходимости они тоже могут быть учтены в замкнутой форме, если принять во внимание, что

$$\nabla_\mu \left[\sum_{n=0}^{\infty}(-\mathfrak{h})^n\right]^{\alpha}_{\beta} = -\left[\sum_{m=0}^{\infty}(-\mathfrak{h})^m\right]^{\alpha}_{\rho}\left(\nabla_\mu \bar{h}^{\rho}_{\sigma}\right)\left[\sum_{k=0}^{\infty}(-\mathfrak{h})^k\right]^{\sigma}_{\beta},$$

где матрица $\mathfrak{h}$ обозначает поле $\bar{h}^{\mu}_{\nu}$.

В результате получается действие

$$S=\int d^4x\sqrt{-g}\left(R+g^{\mu\nu}\left(\delta\Gamma^{\alpha}_{\beta\alpha}\delta\Gamma^{\beta}_{\mu\nu}-\delta\Gamma^{\alpha}_{\mu\beta}\delta\Gamma^{\beta}_{\nu\alpha}\right)\right)=\int d^4x\sqrt{-g}R$$
$$+\int d^4x\frac{\sqrt{-g}}{4}\left[\sum_{n=0}^{\infty}\left(-g^{-1}\bar{h}\right)^n\right]^{\alpha}_{\kappa}g^{\kappa\rho}\left[\sum_{m=0}^{\infty}\left(-g^{-1}\bar{h}\right)^m\right]^{\beta}_{\gamma}g^{\gamma\sigma}\times$$
$$\Bigg(\nabla_\beta\bar{h}_{\alpha\rho}\left(2\nabla_\mu\bar{h}^{\mu}_{\sigma}-\nabla_\sigma\bar{h}^{\mu}_{\mu}\right)$$
$$-\ \left(\nabla_\mu\bar{h}_{\beta\rho}+\nabla_\beta\bar{h}_{\mu\rho}-\nabla_\rho\bar{h}_{\mu\beta}\right)g^{\mu\nu}\left(\nabla_\nu\bar{h}_{\alpha\sigma}+\nabla_\alpha\bar{h}_{\nu\sigma}-\nabla_\sigma\bar{h}_{\alpha\nu}\right)\Bigg).\quad(4.4)$$

При разложении вокруг двойного пространства Минковского в квадратичном приближении в действии для $\bar{h}$ достаточно заменить метрику g метрикой Минковского. Тем самым в этом секторе квадратичное действие уже получено (надо лишь выкинуть лишние слагаемые в степенных рядах). Квадратичное же действие для h получается стандартным образом (общая теория относительности). Для полноты напомним, как это происходит.

Квадратичное действие для общей теории относительности вокруг пространства Минковского может быть получено как

$$S^{(2)}_{GR}=\int d^4x\left[\sqrt{-g}g^{\mu\nu}(\delta^{(2)}R_{\mu\nu})+(\delta^{(1)}(\sqrt{-g}g^{\mu\nu}))(\delta^{(1)}R_{\mu\nu})\right].\qquad(4.5)$$

Первое слагаемое равно

$$\delta^{(2)}R_{\mu\nu}=\delta\Gamma^{\alpha}_{\beta\alpha}\delta\Gamma^{\beta}_{\mu\nu}-\delta\Gamma^{\alpha}_{\mu\beta}\delta\Gamma^{\beta}_{\nu\alpha}+\text{surface terms}$$

и по форме совпадает с квадратичным действием для $\bar{h}$, а второе слагаемое принимает вид

$$\frac{1}{2}\left(-h^{\mu\nu}+\frac{1}{2}h^{\beta}_{\beta}\eta^{\mu\nu}\right)\left(\partial^2_{\nu\alpha}h^{\alpha}_{\mu}+\partial^2_{\mu\alpha}h^{\alpha}_{\nu}-\partial^{2\alpha}_{\ \ \alpha}h_{\mu\nu}-\partial^2_{\mu\nu}h^{\alpha}_{\alpha}\right)$$

и, на самом деле, после интегрирований по частям становится равным первому слагаемому, умноженному на -2,

$$-\frac{1}{2}\left(\partial_\alpha h_{\mu\nu}\partial^\alpha h^{\mu\nu} - 2\partial^\alpha h_{\mu\nu}\partial^\nu h^\mu_\alpha + 2\partial_\alpha h^{\alpha\mu}\partial_\mu h^\beta_\beta - \partial_\mu h^\alpha_\alpha \partial^\mu h^\beta_\beta\right).$$

Собирая оба вклада вместе, можно получить квадратичное действие для общей теории относительности

$$S^{(2)}_{GR} = -\frac{1}{4}\int d^4x \left(\partial_\alpha h_{\mu\nu}\partial^\alpha h^{\mu\nu} - 2\partial^\alpha h_{\mu\nu}\partial^\nu h^\mu_\alpha + 2\partial_\alpha h^{\alpha\mu}\partial_\mu h^\beta_\beta - \partial_\mu h^\alpha_\alpha \partial^\mu h^\beta_\beta\right).$$

В результате для биметрического вариационного приниципа получается следующее квадратичное действие:

$$S^{(2)} = \frac{1}{4}\int d^4x \left(\left(\partial_\alpha \bar{h}_{\mu\nu}\partial^\alpha \bar{h}^{\mu\nu} - 2\partial^\alpha \bar{h}_{\mu\nu}\partial^\nu \bar{h}^\mu_\alpha + 2\partial_\alpha \bar{h}^{\alpha\mu}\partial_\mu \bar{h}^\beta_\beta - \partial_\mu \bar{h}^\alpha_\alpha \partial^\mu \bar{h}^\beta_\beta\right)\right.$$
$$\left. - \left(\partial_\alpha h_{\mu\nu}\partial^\alpha h^{\mu\nu} - 2\partial^\alpha h_{\mu\nu}\partial^\nu h^\mu_\alpha + 2\partial_\alpha h^{\alpha\mu}\partial_\mu h^\beta_\beta - \partial_\mu h^\alpha_\alpha \partial^\mu h^\beta_\beta\right)\right). \quad (4.6)$$

Как видим, независимо от выбранной сигнатуры (или общего знака перед действием) есть два типа гравитонов, один номальный, другой – дух.

Это и не удивительно. Опуская тензорную структуру, наш лагранжиан имеет вид $(\ddot{x} + \dot{x}^2)y$, вторая переменная при этом оказывается динамической за счёт интегрирования по частям, которое приводит к лагранжиану $-\dot{x}\dot{y} + \dot{x}^2 y$. Легко видеть, что подобная квадратичная форма (где одна из скоростей входит лишь в виде произведения с другой) всегда (вне зависимости от конкретных коэффициентов) оказывается после диагонализации знаконеопределённой, то есть содержит и нормальные частицы и духи.

Основываясь только на квадратичном действии, можно надеяться на исправление ситуации с помощью подходящей линейной комбинации с действием Эйнштейна-Гильберта для проблематичной метрики. В следующем разделе мы увидим, что эта надежда не оправдана.

4.2 АДМ анализ для биметрических теорий

Итак, мы хотим проверить, могут ли более сложные формы биметрического вариационного принципа оказаться более жизнеспособными. Однако для ответа на этот вопрос полезно уже построить полноценный АДМ формализм. В этом разделе мы покажем, как это может быть сделано, следуя нашей работе [12*].

4.2.1 Простейшая форма биметрического вариационного принципа

Из формы действия очевидно, что нам потребуются компоненты тензора Риччи $R_{\mu\nu} \equiv R^{\alpha}_{\mu\alpha\nu}$ в АДМ переменных, даже для анализа простейшей модели предыдущего раздела. Используя соотношения (1.36)–(1.39) из первой главы, легко получить

$$R_{ij} = \overset{(3)}{R}_{ij} - \frac{1}{N}\dot{K}_{ij} - \frac{1}{N}\overset{(3)}{\bigtriangledown}_i\overset{(3)}{\bigtriangledown}_j N + K_{ij}K^k_k - 2K_{ik}K^k_j + \frac{1}{N}\mathcal{L}ie_{\vec{N}}K_{ij}, \quad (4.7)$$

$$n^{\mu}R_{\mu i} = \overset{(3)}{\bigtriangledown}_i K^j_j - \overset{(3)}{\bigtriangledown}_j K^j_i, \quad (4.8)$$

$$n^{\mu}n^{\nu}R_{\mu\nu} = \frac{1}{N}\gamma^{ij}\dot{K}_{ij} + \frac{1}{N}\overset{(3)}{\triangle}N + K_{ij}K^{ij} - \frac{1}{N}\gamma^{ij}\mathcal{L}ie_{\vec{N}}K_{ij}. \quad (4.9)$$

Нам теперь будет удобнее обозначать метрику, определяющую связность, буквой g (без черты наверху). Обозначим поэтому физическую метрику буквой $\hat{g}$ (со шляпкой), для более явного отличия от предыдущего раздела.

Предполагая, что обе метрики допускают $(3+1)$-разложение в одних и тех же координатах, мы можем вычислить величину $\hat{g}^{\mu\nu}R_{\mu\nu}$, обозначив АДМ переменные метрики $\hat{g}_{\mu\nu}$, например, через M, M^i и $\hat{\gamma}_{ij}$. Для этого потребуется в формулах (4.7)–(4.9) перейти к явным компонентам тензора Риччи (вместо сверток с вектором единичной нормали), используя

$$\begin{aligned} R_{0i} &= Nn^{\mu}R_{\mu i} + N^j R_{ij}, \\ R_{00} &= N^2 n^{\mu}n^{\nu}R_{\mu\nu} + 2NN^i n^{\mu}R_{\mu i} + N^i N^j R_{ij}. \end{aligned}$$

Подставляя результаты в интересующее нас действие, получаем для лагранжевой плотности $\sqrt{-\hat{g}}\hat{g}^{\mu\nu}R_{\mu\nu}$

$$
\begin{aligned}
&= M\sqrt{\hat{\gamma}}\left[-\frac{N^2}{M^2}\left(\frac{1}{N}\gamma^{ij}\dot{K}_{ij} + K_{ij}K^{ij} + \frac{1}{N}\overset{(3)}{\triangle}N - \frac{1}{N}\gamma^{ij}\mathcal{L}ie_{\vec{N}}K_{ij}\right)\right. \\
&\qquad + 2\frac{N}{M^2}(M^i - N^i)\left(\overset{(3)}{\nabla}_i K^j_j - \overset{(3)}{\nabla}_j K^j_i\right) \\
&\qquad + \left(\hat{\gamma}^{ij} - \frac{\left(M^i - N^i\right)\left(M^j - N^j\right)}{M^2}\right)\times \\
&\left.\left(\overset{(3)}{R}_{ij} - \frac{1}{N}\dot{K}_{ij} - \frac{1}{N}\overset{(3)}{\nabla}_i\overset{(3)}{\nabla}_j N + K_{ij}K^k_k - 2K_{ik}K^k_j + \frac{1}{N}\mathcal{L}ie_{\vec{N}}K_{ij}\right)\right], \quad (4.10)
\end{aligned}
$$

где по договорённости индексы тензоров внешней кривизны опускаются и поднимаются с помощью метрики γ^{ij}.

Вводя новые переменные $a \equiv \frac{N}{M}$ и $a^i \equiv \frac{M^i - N^i}{M}$ и приводя подобные слагаемые в уравнении (4.10), получаем

$$
\begin{aligned}
&\sqrt{-\hat{g}}\hat{g}^{\mu\nu}R_{\mu\nu} \\
&\quad = \sqrt{\gamma}\cdot\sqrt{\frac{\hat{\gamma}}{\gamma}}\left[\left(\frac{1}{a}\left(\hat{\gamma}^{ij} - a^i a^j\right) + a\gamma^{ij}\right)\left(\mathcal{L}ie_{\vec{N}}K_{ij} - \overset{(3)}{\nabla}_i\overset{(3)}{\nabla}_j N - \dot{K}_{ij}\right)\right. \\
&\qquad + 2Na^i\left(\overset{(3)}{\nabla}_i K^j_j - \overset{(3)}{\nabla}_j K^j_i\right) + \frac{N}{a}\left(\hat{\gamma}^{ij} - a^i a^j\right)\overset{(3)}{R}_{ij} \\
&\left. + N\left(\frac{1}{a}\left(\hat{\gamma}^{ij} - a^i a^j\right)\gamma^{kl} - a\gamma^{ik}\gamma^{jl} - \frac{2}{a}\left(\hat{\gamma}^{ik} - a^i a^k\right)\gamma^{jl}\right)K_{ij}K_{kl}\right], \quad (4.11)
\end{aligned}
$$

где мы явно вынесли общий множитель $\sqrt{\gamma}$, чтобы упростить, когда это потребуется, интегрирование по частям в слагаемых, содержащих ковариантные производные $\overset{(3)}{\nabla}_i$.

Мы видим, что вторая производная по времени появляется только в одном месте, а именно $-\sqrt{\gamma}\sqrt{\frac{\hat{\gamma}}{\gamma}}\frac{1}{a}\left(\hat{\gamma}^{ij} - a^i a^j + a^2\gamma^{ij}\right)\dot{K}_{ij}$. Определим (пространственный) тензор χ^{ij} как следующую комбинацию пространственных метрик:

$$\chi^{ij} \equiv \sqrt{\frac{\hat{\gamma}}{\gamma}}\frac{1}{a}\left(\hat{\gamma}^{ij} - a^i a^j + a^2\gamma^{ij}\right). \tag{4.12}$$

Можно провести интегрирование по частям по следующей схеме:

$$-\sqrt{\gamma}\chi^{ij}\dot{K}_{ij} \longrightarrow \sqrt{\gamma}K_{ij}\dot{\chi}^{ij} + \sqrt{\gamma}K_{ij}\chi^{ij}\overset{(3)}{\nabla}_k N^k - \sqrt{\gamma}N\chi^{ij}\gamma^{kl}K_{ij}K_{kl}. \tag{4.13}$$

В результате χ^{ij} становится динамической, но её скорость входит в действие только линейно, в произведении с другой скоростью – это в точности проблема духа, с которой мы встретились в предыдущем разделе.

Используя соотношение (4.13), легко переписать лагранжиан (4.11) в виде

$$\begin{aligned} S = \int dt d^3x \sqrt{\gamma}\Bigg[& 2\chi^{ij}K_{jk}\overset{(3)}{\nabla}_i N^k - K_{ij}N^k\overset{(3)}{\nabla}_k\chi^{ij} - \chi^{ij}\overset{(3)}{\nabla}_i\overset{(3)}{\nabla}_j N + K_{ij}\dot{\chi}^{ij} \\ & + 2\gamma^{kj}K_{ij}\overset{(3)}{\nabla}_k\left(\sqrt{\frac{\hat{\gamma}}{\gamma}}Na^i\right) - 2\gamma^{ij}K_{ij}\overset{(3)}{\nabla}_k\left(\sqrt{\frac{\hat{\gamma}}{\gamma}}Na^k\right) \\ & + N\left(\chi^{ij} - a\sqrt{\frac{\hat{\gamma}}{\gamma}}\gamma^{ij}\right)\overset{(3)}{R}_{ij} \\ & + N\left(\left(a\sqrt{\frac{\hat{\gamma}}{\gamma}}\gamma^{ik} - 2\chi^{ik}\right)\gamma^{jl} - a\sqrt{\frac{\hat{\gamma}}{\gamma}}\gamma^{ij}\gamma^{kl}\right)K_{ij}K_{kl}\Bigg], \end{aligned} \tag{4.14}$$

где $\sqrt{\hat{\gamma}}$ должна пониматься как функция независимых перменных γ_{ij}, χ^{ij}, a и a^i. Заметим, что мы проитегрировали по частям некоторые слагаемые, содержащие ковариантные производные $\overset{(3)}{\nabla}$, чтобы избавиться от производных внешних кривизн K_{ij} (это будет удобно для вычисления канонических импульсов), а также проделали некоторые несложные тождественные преобразования.

Все временные производные от a и a^i оказались собраны в слагаемые с $\dot{\chi}^{ij}$. Соответственно, если принять за независимые переменные γ_{ij}, χ^{ij}, a и a^i, то a и a^i оказываются нединамическими. Однако же, действие зависит от них нелинейно, поскольку они нелинейно входят в величину $\sqrt{\hat{\gamma}}$. В самом деле, $\frac{\hat{\gamma}}{\gamma} = \det\left(\hat{\gamma}^{ik}\gamma_{kj}\right)$, и из определения (4.12) заключаем, что $\sqrt{\frac{\hat{\gamma}}{\gamma}}$ является решением нелинейного алгебраического уравнения

$$\frac{\hat{\gamma}}{\gamma} = \det\left(\sqrt{\frac{\gamma}{\hat{\gamma}}}a\chi^i_j + a^i a_j - a^2\delta^i_j\right),$$

где индексы поднимаются метрикой γ^{ij}.

Соответственно, вариации по отношению к N и N^i приводят к появлению связей в физическом секторе (проявление калибровочной инвариантности), в то время как нединамические переменные a и a^i производят лишь уравнения на самих себя. Соответственно имеем 12 переменных, входящих в действие со скоростями (6 компонент γ_{ij} и 6 компонент χ^{ij}), и 4 калибровочные симметрии. Общее число степеней свободы равно 8. Это и не удивительно для биметрической теории общего положения: две поляризации безмассового гравитона и шесть степеней свободы от второго симметричного пространственного тензора.

Гамильтониан

Вычислим также гамильтониан модели. Для этого надо найти канонические импульсы

$$\pi^{(\chi)}_{ij} \equiv \frac{\partial\mathcal{L}}{\partial\dot{\chi}^{ij}} = \sqrt{\gamma}K_{ij} = \frac{\sqrt{\gamma}}{2N}\left(\overset{(3)}{\nabla}_i N_j + \overset{(3)}{\nabla}_j N_i - \dot{\gamma}_{ij}\right) \tag{4.15}$$

и

$$\pi^{ij}_{(\gamma)} \equiv \frac{\partial\mathcal{L}}{\partial\dot{\gamma}_{ij}} = -\frac{\sqrt{\gamma}}{2N}\Bigg[\chi^{ik}\overset{(3)}{\nabla}_k N^j + \chi^{jk}\overset{(3)}{\nabla}_k N^i - N^k\overset{(3)}{\nabla}_k\chi^{ij} + \dot{\chi}^{ij} + \gamma^{kj}\overset{(3)}{\nabla}_k\left(\sqrt{\frac{\hat{\gamma}}{\gamma}}Na^i\right) + \gamma^{ki}\overset{(3)}{\nabla}_k\left(\sqrt{\frac{\hat{\gamma}}{\gamma}}Na^j\right) - 2\gamma^{ij}\overset{(3)}{\nabla}_k\left(\sqrt{\frac{\hat{\gamma}}{\gamma}}Na^k\right) + N\left(A^{ijkl} + A^{klij}\right)K_{kl}\Bigg], \tag{4.16}$$

где нами введена величина

$$A^{ijkl} \equiv \left(a\sqrt{\frac{\hat{\gamma}}{\gamma}}\gamma^{ik} - 2\chi^{ik} \right)\gamma^{jl} - a\sqrt{\frac{\hat{\gamma}}{\gamma}}\gamma^{ij}\gamma^{kl}.$$

Из уравнения (4.15) легко находятся скорости переменных γ,

$$\dot{\gamma}_{ij} = \overset{(3)}{\nabla}_i N_j + \overset{(3)}{\nabla}_j N_i - \frac{2N}{\sqrt{\gamma}}\pi^{(\chi)}_{ij},$$

или иными словами $K_{ij} = \frac{1}{\sqrt{\gamma}}\pi^{(\chi)}_{ij}$. Аналогично можно найти $\dot{\chi}$, используя уравнение (4.16). Мы не приводим здесь этого громоздкого выражения, и вместо этого запишем получающийся гамильтониан (опять выполнены некоторые интегрирования по частям, чтобы явно выделить переменные шага и сдвига в виде множителей Лагранжа):

$$\begin{aligned} H = \pi^{ij}_{(\gamma)}\dot{\gamma}_{ij} + \pi^{(\chi)}_{ij}\dot{\chi}^{ij} - \mathcal{L} = -\int dtd^3x\sqrt{\gamma}\Bigg\{ N \Bigg[2\frac{\pi^{ij}_{(\gamma)}\pi^{(\chi)}_{ij}}{\gamma} \\ + \left(\left(a\sqrt{\frac{\hat{\gamma}}{\gamma}}\gamma^{ik} - 2\chi^{ik} \right)\gamma^{jl} - a\sqrt{\frac{\hat{\gamma}}{\gamma}}\gamma^{ij}\gamma^{kl} \right)\frac{\pi^{(\chi)}_{ij}\pi^{(\chi)}_{kl}}{\gamma} \\ + \left(\chi^{ij} - a\sqrt{\frac{\hat{\gamma}}{\gamma}}\gamma^{ij} \right)\overset{(3)}{R}_{ij} + 2\sqrt{\frac{\hat{\gamma}}{\gamma}}a^i\gamma^{kj}\left(\overset{(3)}{\nabla}_i\frac{\pi^{(\chi)}_{kj}}{\sqrt{\gamma}} - \overset{(3)}{\nabla}_j\frac{\pi^{(\chi)}_{ki}}{\sqrt{\gamma}} \right) - \overset{(3)}{\nabla}_i\overset{(3)}{\nabla}_j\chi^{ij} \Bigg] \\ + N^k\left(2\gamma_{kj}\overset{(3)}{\nabla}_i\frac{\pi^{ij}_{(\gamma)}}{\sqrt{\gamma}} - 2\overset{(3)}{\nabla}_i\left(\chi^{ij}\frac{\pi^{(\chi)}_{jk}}{\sqrt{\gamma}} \right) - \frac{\pi^{(\chi)}_{ij}}{\sqrt{\gamma}}\overset{(3)}{\nabla}_k\chi^{ij} \right) \Bigg\}. \end{aligned} \tag{4.17}$$

Как и положено репараметризационно-инвариантной теории, гамильтониан оказывается линейной комбинацией связей. Однако, при изучении возмущений вокруг какого-либо фона, разумеется, приобретает разумный смысл эволюция во времени, а с нею и значения энергия. При этом энергия будет неограниченной ни снизу ни сверху, поскольку импульсы $\pi^{ij}_{(\gamma)}$ входят лишь в комбинации $\pi^{ij}_{(\gamma)}\pi^{(\chi)}_{ij}$ – тот же самый эффект, что был виден и в лагранжевом подходе.

4.2.2 Включение слагаемого Эйнштейна-Гильберта

Из соображений линеаризованной гравитации можно было надеяться, что избежать духов позволит действие

$$S = \int d^4x \sqrt{\hat{g}} \hat{g}^{\mu\nu} \left(\alpha \hat{R}_{\mu\nu} + R_{\mu\nu} \right)$$

при подходящем выборе коэфиициента α.

Однако, с учётом нелинейных эффектов ситуация становится только хуже. В самом деле, теперь мы имеем лагранжеву плотность

$$\begin{aligned} \sqrt{-\hat{g}} \hat{g}^{\mu\nu} \left(\alpha \hat{R}_{\mu\nu} + R_{\mu\nu} \right) &= \alpha \sqrt{\hat{\gamma}} \left(\overset{(3)}{\hat{R}} + \left(\hat{\gamma}^{ik} \hat{\gamma}^{jl} - \hat{\gamma}^{ij} \hat{\gamma}^{kl} \right) \hat{K}_{ij} \hat{K}_{kl} \right) \\ + \sqrt{\gamma} \cdot \sqrt{\frac{\hat{\gamma}}{\gamma}} \Bigg[&\left(\frac{1}{a} \left(\hat{\gamma}^{ij} - a^i a^j \right) + a \gamma^{ij} \right) \left(\mathcal{L}ie_{\overrightarrow{N}} K_{ij} - \overset{(3)}{\nabla}_i \overset{(3)}{\nabla}_j N - \dot{K}_{ij} \right) \\ &+ 2 N a^i \left(\overset{(3)}{\nabla}_i K^j_j - \overset{(3)}{\nabla}_j K^j_i \right) + \frac{N}{a} \left(\hat{\gamma}^{ij} - a^i a^j \right) \overset{(3)}{R}_{ij} \\ + N \left(\frac{1}{a} \left(\hat{\gamma}^{ij} - a^i a^j \right) \gamma^{kl} - a \gamma^{ik} \gamma^{jl} - \frac{2}{a} \left(\hat{\gamma}^{ik} - a^i a^k \right) \gamma^{jl} \right) & K_{ij} K_{kl} \Bigg] . \end{aligned} \qquad (4.18)$$

Важное отличие от действия (4.11) заключается в том, что метрика $\hat{\gamma}_{ij}$ с самого начала имеет независимые кинетические слагаемые. Поэтому скорости величин a и a^i нельзя поглотить в производную по времени от поля χ. Поля a и a^i становятся четырьмя новыми степенями свободы, причём их скорости входят в лагранжиан линейно (в произведениях на другие скорости). Тем самым в модели присутствуют духи. Заметим, что это заведомо не чистая калибровка, поскольку, например, для решений с $\hat{g}_{\mu\nu} = g_{\mu\nu}$ имеем $a = 1$ и $a^i = 0$ при любом выборе координат. Полное число степеней свободы получается равным $6+6+4-4 = 12$. Вычисление гамильтониана оказывается довольно громоздким, и мы опускаем его.

* * *

Рассмотроим также (эквивалентное) действие из предыдущего раздела, просуммировав в нём бесконечный ряд $g^{\mu\nu} = \left[\sum_{n=0}^{\infty} \left(-\hat{g}^{-1} h \right)^n \right]^{\mu}_{\alpha} \hat{g}^{\alpha\nu}$

для обратной метрики. В наших обозначениях оно принимает вид

$$S = \int d^4x\sqrt{-\hat{g}}\left\{\hat{R} + \frac{1}{4}g^{\alpha\rho}g^{\beta\sigma}\left[\left(\hat{\bigtriangledown}_\beta h_{\alpha\rho}\right)\hat{g}^{\mu\kappa}\left(2\hat{\bigtriangledown}_\mu h_{\sigma\kappa} - \hat{\bigtriangledown}_\sigma h_{\mu\kappa}\right)\right.\right.$$
$$\left.\left. - \left(\hat{\bigtriangledown}_\mu h_{\beta\rho} + \hat{\bigtriangledown}_\beta h_{\mu\rho} - \hat{\bigtriangledown}_\rho h_{\mu\beta}\right)\hat{g}^{\mu\nu}\left(\hat{\bigtriangledown}_\nu h_{\alpha\sigma} + \hat{\bigtriangledown}_\alpha h_{\nu\sigma} - \hat{\bigtriangledown}_\sigma h_{\alpha\nu}\right)\right]\right\} \quad (4.19)$$

где $h_{\mu\nu} \equiv g_{\mu\nu} - \hat{g}_{\mu\nu}$, но можно смело подставить $g_{\mu\nu}$ вместо $h_{\mu\nu}$, поскольку $\hat{g}_{\mu\nu}$ ковариантно постоянна по отношению к производной $\hat{\bigtriangledown}_\alpha$.

Очевидно, что $\dot{h}_{00}^2$, $\dot{h}_{00}\dot{h}_{0i}$ и $\dot{h}_{0i}\dot{h}_{0j}$ выпадают из действия, но $\dot{h}_{00}\dot{h}_{ij}$ и $\dot{h}_{0k}\dot{h}_{ij}$ (и конечно $\dot{h}_{ij}\dot{h}_{kl}$) остаются. Явное вычисление показывает, что есть только следующие два слагаемых в кинетической функции для h_{00} и h_{0i}:

$$\frac{1}{4}\left(2g^{00}g^{0i}\hat{g}^{0j} + g^{00}g^{ij}\hat{g}^{00} - g^{00}g^{00}\hat{g}^{ij} - 2g^{0i}g^{0j}\hat{g}^{00}\right)\dot{h}_{00}\dot{h}_{ij},$$
$$\frac{1}{4}\left(4g^{00}g^{jk}\hat{g}^{0i} + 2g^{00}g^{0k}\hat{g}^{ij} - 2g^{0k}g^{ij}\hat{g}^{00} - 4g^{0i}g^{jk}\hat{g}^{00}\right)\dot{h}_{0k}\dot{h}_{ij}.$$

Когда $\hat{g}_{\mu\nu} = g_{\mu\nu}$, эти кинетические слагаемые пропадают в полном согласии с линейным анализом, который нашёл только две моды спиральности два. С другой стороны, есть также хорошее соответствие с АДМ анализом, в котором $\dot{a}$ и $\dot{a}^i$ тоже вошли в действие линейно.

Может показаться странным, что простое изменение относительного коэффициента при части со слагаемыми $\left(\hat{\bigtriangledown}h\right)^2$ в действии (4.19) приводит к такому потрясающему эффекту – появление новых степеней свободы, как было описано выше, ведь это действие выглядит как просто поле $h_{\mu\nu}$, взаимодействующее с общей теорией относительности. Однако для симметричного тензора эта точки зрения слишком наивна. Его ковариантные производные содержат производные от метрики $\hat{g}$ в коэффициентах связности, и поэтому в них имеется сложное кинетическое смешивание двух метрик.

4.2.3 Обобщения с нелинейными функциями

Следующая по сложности возможность – попробовать нелинейную функцию смешанной скалярной кривизны,

$$S = \int d^4x \sqrt{-\hat{g}}\; f\left(\mathcal{R}\right), \tag{4.20}$$

где $\mathcal{R} \equiv \hat{g}^{\mu\nu} R_{\mu\nu}$. Эта идея проваливается весьма замечательным образом.

Применим стандартный для $f\left(R\right)$ гравитации трюк и перепишем действие (4.20) в эквивалентной форме

$$S = \int d^4x \sqrt{-\hat{g}} \left[\hat{g}^{\mu\nu} f'(\phi) R_{\mu\nu} + f(\phi) - \phi f'(\phi)\right], \tag{4.21}$$

со вспомогательным скаляром ϕ. Как обычно, решение алгебраического уравнения движения для ϕ и подстановка полученного ответа в действие (4.21) приводит к исходному действию (4.20).

Множитель $f'(\phi)$ при кривизне можно поглотить в переопределение метрики $\hat{g}_{\mu\nu}$. Однако отличие от обычной $f\left(R\right)$ гравитации заключается в том, что скалярное поле не приобретает при этом кинетического слагаемого, поскольку нам не приходится пересчитывать тензор Риччи, зависящий от другой метрики. Итак, вводя новую метрику $\hat{\hat{g}}_{\mu\nu} \equiv f'(\phi) \cdot \hat{g}_{\mu\nu}$, получаем действие

$$S = \int d^4x \sqrt{-\hat{\hat{g}}} \left(\hat{\hat{g}}^{\mu\nu} R_{\mu\nu} + \frac{f(\phi) - \phi f'(\phi)}{f'^2(\phi)}\right), \tag{4.22}$$

в которое должно быть подставлено значение поля ϕ в минимуме потенциала $V(\phi) = \frac{f(\phi) - \phi f'(\phi)}{f'^2(\phi)}$. Таким образом, мы получили исходную теорию с добавлением космологической постоянной. Заметим, что если $f(0) = 0$, то возможно решение с $\phi = 0$, которое полностью эквивалентно исходной модели (без добавления космологической постоянной). Для простейшей же функции $f(\phi) = \phi + \alpha\phi^2$ это вообще единственное решение.

Отметим, правда, что описанная выше эквивалентность в точности справедлива при отсутствии полей материи. После конформного преобразования в гравитационном секторе материя будет взаимодействовать с той же метрикой, что и раньше, но в новых переменных это $\frac{\hat{\hat{g}}_{\mu\nu}}{f'(\phi)}$. Впрочем, учитывая, что в данной модели поле ϕ принимает постоянное значение в минимуме потенциала, этот эффект лишь перенормирует гравитационную постоянную.

Можно надеяться на использование более сложных моделей, например, комбинируя теперь уже нелинейное действие (4.20) с членом Эйнштейна-Гильберта для метрики $\hat{g}_{\mu\nu}$. Это приведёт, разумеется, к тому, что поле ϕ станет динамическим, и вообще действие (4.18) будет модифицировано. Однако легко видеть, что это не изменит присутствия проблематичных динамических полей a и a^i.

Также возможно использование следующих (смешанных) инвариантов кривизны, таких как $\hat{g}^{\mu\alpha}\hat{g}^{\nu\beta}R_{\mu\nu}R_{\alpha\beta}$. Однако, как хорошо известно, и очевидно из АДМ анализа, подобные слагаемые содержат квадратичные по $\dot{K}$ члены, неминуемо приводя к духам Остроградского.

4.2.4 Один класс (скалярно-тензорных) моделей

Как видим, оказывается невозможным построить жизнеспособный биметрический вариационный принцип, опираясь на совершенно произвольную вспомогательную метрику. Однако можно пытаться строить аналогичные формализмы, ограничивая допустимый класс вариаций этой метрики, или можно рассматривать другие типы связности.

Построим пример успешной модели такого сорта. Рассмотрим вспомогательные коэффициенты связности вида

$$\Gamma^{\alpha}_{\mu\nu} = \hat{\Gamma}^{\alpha}_{\mu\nu} + \delta\Gamma^{\alpha}_{\mu\nu}, \tag{4.23}$$

где

$$\delta\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2}\left[a(\rho, X)\left(\delta^{\alpha}_{\mu}\partial_{\nu}\rho + \delta^{\alpha}_{\nu}\partial_{\mu}\rho\right) - b(\rho, X)\hat{g}_{\mu\nu}\partial^{\alpha}\rho\right] \tag{4.24}$$

при $X \equiv (\partial_{\mu}\rho)(\partial^{\mu}\rho)$ со скалярным полем ρ.

Учитывая, что

$$R_{\mu\nu} = \hat{R}_{\mu\nu} + \hat{\nabla}_{\alpha}\delta\Gamma^{\alpha}_{\mu\nu} - \hat{\nabla}_{\nu}\delta\Gamma^{\alpha}_{\mu\alpha} + \delta\Gamma^{\alpha}_{\beta\alpha}\delta\Gamma^{\beta}_{\mu\nu} - \delta\Gamma^{\alpha}_{\mu\beta}\delta\Gamma^{\beta}_{\nu\alpha}. \tag{4.25}$$

можно написать действие (4.1) в виде

$$S = \int d^4 \sqrt{-\hat{g}} \left[\hat{R} + \frac{3a^2 - 12ab + 3b^2}{4} \hat{g}^{\mu\nu} (\partial_\mu \rho)(\partial_\nu \rho) \right]. \tag{4.26}$$

Случай $a = b$ при зависимости только от ρ, но не от X, отвечает конформному преобразованию метрики некоторой функцией поля ρ. При этом получается скалярно-тензорная теория с каноническим скалярным полем, а сама модель может рассматриваться как биметрическая в нашем смысле, но с ограничением, что две метрики отличаются друг от друга конформным преобразованием. В общем же случае мы получаем аналогичную теорию типа k-эссенции, которая обладает здоровым знаком при кинетическом члене при условии $(a - b)^2 < 2ab$. (Впрочем, свойства устойчивости зависят ещё и от того, как устроены функции a и b по своему аргументу X.) Тем самым возможен новый взгляд на модели k-эссенции.

Можно также обобщить модель, рассмотрев при нашей связности (4.23, 4.24) нелинейную функцию $\mathcal{R}$ в действии (4.20). Положим для простоты $a = b = 1$ (обычное конформное преобразование). Подставляя $\hat{g}^{\mu\nu} R_{\mu\nu} = \hat{g}^{\mu\nu} \left(\hat{R}_{\mu\nu} - \frac{3}{2} (\partial_\mu \rho)(\partial_\nu \rho) \right)$ в действие (4.21) и преобразовывая метрику, получаем

$$S = \int d^4x \sqrt{-\hat{\hat{g}}} \left(\hat{\hat{R}} - \frac{3}{2} (\partial_\mu \rho)(\partial^\mu \rho) \right.$$
$$\left. - \frac{3}{2} \left(\partial_\mu \log f'(\phi)\right) \left(\partial^\mu \log f'(\phi)\right) + \frac{f(\phi) - \phi f'(\phi)}{f'^2(\phi)} \right),$$

причём материя будет взаимодействовать с $\hat{g}_{\mu\nu} = \frac{\hat{\hat{g}}_{\mu\nu}}{f'(\phi)}$. Получается, по сути, обычная теория типа $f(R)$, но с дополнительным (скрытым) безмассовым скаляром, который при желании можно пытаться связать с тёмным излучением (dark radiation).

* * *

Использование скаляров при модификации связности – это один из самых удобных способов получения новых степеней свободы, поскольку естественная запись (4.24) всегда содержит производные от поля, и оно становится динамическим, и при этом, как правило, существуют необременительные условия, позволяющие избежать появления духов. Это общая ситуация, так же бывает не только с неметричностью, но и с кручением. Векторные поля сложнее, поскольку в простейшем виде, при $\delta\Gamma^{\alpha}_{\mu\nu} = \delta^{\alpha}_{\mu}A_{\nu} + \delta^{\alpha}_{\nu}A_{\mu} - g_{\mu\nu}A^{\alpha}$, они оказываются нединамическими. Тем не менее, их тоже можно использовать, см. работу [162].

4.3 Пример нелокальной гравитации из биметрического подхода

В этом разделе мы рассмотрим пример ограниченных биметрических теорий, но с условиями, которые сами зависят от кривизны. Условия при этом будут рассматриваться как конформные (C-модель) так и дисформные, с тензором Риччи (D-модель). Поскольку соотношения, накладываемые на переменные, содержат производные, требуется большая осторожность при обращении с ними. Этот аспект, похоже, не был в должной мере осознан авторами работ [161, 163], в которых впервые эти модели были предложены. Мы поначалу тоже будем действовать наивно и увидим, к каким проблемам это приводит.

4.3.1 C- и D-теории

Рассмотрим (в произвольной размерности) действие уже знакомого нам вида

$$S = \int d^n x \sqrt{-g} g^{\mu\nu} \hat{R}_{\mu\nu}, \tag{4.27}$$

в котором тензор Риччи со шляпкой соответствует леви-чивитовской связности вспомогательной метрики

$$\hat{g}_{\mu\nu} = C(\mathbf{R}) g_{\mu\nu} + D(\mathbf{R}) \hat{R}_{\mu\nu}. \tag{4.28}$$

В самом общем случае функции C и D зависят от матрицы

$$\mathbf{R}^{\mu}_{\nu} \equiv g^{\mu\alpha}\hat{R}_{\alpha\nu}$$

посредством любых ее скалярных инвариантов: $\mathrm{Tr}\mathbf{R} = g^{\mu\nu}\hat{R}_{\mu\nu} \equiv \mathcal{R}$ и $\mathrm{Tr}\mathbf{R}^2$ и так далее, а также, при желании, и любых их производных, таких как $\Box\mathcal{R}$ и $\mathrm{Tr}[(\nabla\mathrm{Tr}\nabla\mathbf{R})\mathbf{R}]$ и так далее. Заметим, что исходная версия [161] относилась к весьма частному случаю $C = C(\mathcal{R})$, $D = 0$, который дальше активно исследовался, например, в работе [164]). Однако такая версия способна приводить лишь к скалярно-тензорным модификациям, а общие модели [163] могут модифицировать тензорный пропагатор, что само по себе интересно.

Определим также матрицу

$$\hat{\mathbf{R}}^{\mu}_{\nu} \equiv \hat{g}^{\mu\alpha}\hat{R}_{\alpha\nu}$$

и постараемся выразить действие в виде функции от одной из метрик. Мы имеем

$$g_{\mu\nu} = \frac{\hat{g}_{\mu\alpha}}{C}\left(\delta^{\alpha}_{\nu} - D\hat{R}^{\alpha}_{\nu}\right)$$

или, в матричной записи,

$$g = \frac{1}{C}\cdot\hat{g}\left(I - D\hat{\mathbf{R}}\right),$$

а для обратной метрики –

$$g^{-1} = C\cdot\left(I - D\hat{\mathbf{R}}\right)^{-1}\hat{g}^{-1}. \tag{4.29}$$

Действие тогда принимает вид

$$S = \int d^n x\sqrt{-\hat{g}}\cdot\frac{\sqrt{\det\left(\mathrm{I} - \mathrm{D}\hat{\mathbf{R}}\right)}}{C^{\frac{n-2}{2}}}\cdot\mathrm{Tr}\left(\left(\mathrm{I} - \mathrm{D}\hat{\mathbf{R}}\right)^{-1}\hat{\mathbf{R}}\right). \tag{4.30}$$

Чтобы действительно выразить действие в терминах одной только метрики $\hat{g}_{\mu\nu}$, надо пересчитать аргументы функций C и D к величинам только со шляпкой. Чтобы это сделать, умножим уравнение (4.29) для g^{-1} на $\hat{R}_{\mu\nu}$, и получим уравнение, определяющее $\mathbf{R}$ в терминах $\hat{\mathbf{R}}$:

$$\mathbf{R} = C(\mathbf{R}) \cdot \Big(I - D(\mathbf{R}) \cdot \hat{\mathbf{R}}\Big)^{-1} \hat{\mathbf{R}}\,. \tag{4.31}$$

Это уравнение может – по крайней мере, в принципе – быть разрешено относительно $\mathbf{R}$. Подставив ответ в действие (4.30), получаем желаемый результат.

Этот результат обобщает так называемую "C-картину" из работы [161]. Как легко видеть, в частном случае $C = C(\mathcal{R})$, $D = 0$ мы тоже получаем теорию типа $f(R)$.

Такая простая форма связана конечно (помимо простоты конформного соотношения) с тем, что мы выражали действие через ту метрику, которая зависит от производных другой метрики (в кривизне). Если бы мы захотели пойти в обратную сторону, то пришлось бы решать дифференициальные уравнения и, по крайней мере наивно, была бы получена нелокальная теория. Впрочем, сделать это точно весьма непросто.

4.3.2 Вычисление $\mathcal{R}$ во всех порядках в нелокальной картине

Рассмотрим переход к нелокальной картине по теории возмущений в $C(\mathcal{R})$-модели (при $D = 0$). Разложим определяющую функцию в ряд

$$C(\mathcal{R}) \equiv 1 + \sum_{i=1}^{\infty} c_{,i} \mathcal{R}^i$$

по степеням смешанной кривизны. Легко видеть (просто конформное преобразование, см. [163]), что

$$\mathcal{R} = R - (n-1)\frac{\Box C}{C} - (n-1)(n-6)\frac{(\partial C)^2}{4C^2}. \tag{4.32}$$

Очевидно, в первом порядке имеем

$$\mathcal{R} \approx R - (n-1)\Box c_{,1}\mathcal{R},$$

откуда

$$\mathcal{R} \approx \frac{1}{1+(n-1)\Box c_{,1}} R, \tag{4.33}$$

где дробь должна пониматься как нелокальный оператор, действующий на R (и мы имеем в виду, что $\Box$ может не коммутировать с $c_{,1}$). Разумеется, тут возникают все обычные проблемы с определением нелокальных операторов, но мы их игнорируем.

Желая пойти во второй порядок, положим

$$\mathcal{R} = \frac{1}{1+(n-1)\Box c_{,1}} R + X_2$$

в соотношении (4.32), где X_2 обозначает поправку второго порядка. В этом порядке имеем

$$\begin{aligned} X_2 = {} & \frac{1}{1+(n-1)\Box c_{,1}} \times \\ & \left[(n-1)\left(c_{,1}\frac{1}{1+(n-1)\Box c_{,1}} R \right) \Box \left(c_{,1}\frac{1}{1+(n-1)\Box c_{,1}} R \right) \right. \\ & \left. \quad - \frac{(n-1)(n-6)}{4} \left(\partial_\mu \left(c_{,1}\frac{1}{1+(n-1)\Box c_{,1}} R \right) \right)^2 \right] \\ & \qquad - (n-1)\Box c_{,2} \left(\frac{1}{1+(n-1)\Box c_{,1}} R \right)^2 . \end{aligned} \tag{4.34}$$

Легко понять общую структуру старших порядков:

$$X_n = \frac{1}{1+(n-1)\Box c_{,1}} \cdot \mathfrak{F}\left(X_i,\ \Box X_i,\ (\partial_\mu X_i)^2 \right) - (n-1)\Box c_{,n} \left(\mathcal{R}^1 \right)^n ,$$

где $i = 1, \ldots, n-1$, а $\mathcal{R}^1$ – результат первого порядка, полученный нами выше (4.33). Следовательно, с точностью до определения нелокального оператора $(1+(n-1)\Box c_{,1})^{-1}$, получаем полное разложение в ряд величины $\mathcal{R}$ как функции R .

Заметим, что если при вычислении действия будет позволено проинтегрировать по частям в предпоследнем слагаемом из выражения (4.34), а также проигнорировать поверхностный вклад с $c_{,2}$, то квадратичное действие примет вид

$$S = \int d^n x \sqrt{-g} \cdot \frac{1}{1 + (n-1)\Box c_{,1}} \left(R + \frac{(n-1)(n-2)}{4} \left(c_{,1} \frac{1}{1+(n-1)\Box c_{,1}} R \right) \Box \left(c_{,1} \frac{1}{1+(n-1)\Box c_{,1}} R \right) \right), \tag{4.35}$$

а если $c_{,1}$ – просто число (без собственных нелокальностей), то дробь, стоящую общим множителем, можно рассматривать как единицу с добавлением операторов, дающих поверхностные слагаемые.

4.3.3 Линеаризованная теория

Рассмотрим теперь теорию, заданную формулами (4.27, 4.28) в линеаризованном виде вокруг пространства Минковского. Рассмотрение будет проведено в обеих картинах.

Локальная $\hat{g}$-картина

Перейдём сперва к картине в метрике со шляпкой. Раскладываем определяющие функции в ряд вокруг пространства Минковского:

$$C(\mathbf{R}) = 1 + c_{,1}\mathcal{R} + \cdots, \qquad D(\mathbf{R}) = d_0 + \cdots .$$

Для линейных флуктуаций ($g \equiv \eta + h$) имеем

$$\hat{h}_{\mu\nu} = h_{\mu\nu} + c_{,1}\mathcal{R}\eta_{\mu\nu} + d_0 \hat{R}_{\mu\nu}, \tag{4.36}$$

а для кривизн

$$\hat{R}_{\mu\nu} = \frac{1}{2}\left(\partial^2_{\mu\alpha}\hat{h}^\alpha_\nu + \partial^2_{\nu\alpha}\hat{h}^\alpha_\mu - \Box\hat{h}_{\mu\nu} - \partial^2_{\mu\nu}\hat{h}^\alpha_\alpha \right) + \cdots$$

и

$$\mathcal{R} = \partial^2_{\mu\nu}\hat{h}^{\mu\nu} - \Box\hat{h}^{\mu}_{\mu} + \cdots \,,$$

где индексы поднимаются метрикой Минковского $\eta_{\mu\nu}$.

Квадратичное действие вычисляется как

$$S = \int d^n x \left(\eta^{\mu\nu}\delta^{(2)}\hat{R}_{\mu\nu} + \delta^{(1)}\left(\sqrt{-\hat{g}}g^{\mu\nu}\right)\cdot\delta^{(1)}\hat{R}_{\mu\nu}\right),$$

где $\delta^{(n)}(X)$ – вклад n-ого порядка к величине X. Используя элементарные соотношения

$$\delta^{(1)}\left(\sqrt{-g}g^{\mu\nu}\right) = -h^{\mu\nu} + \frac{1}{2}\eta^{\mu\nu}h^{\alpha}_{\alpha}$$

и

$$\delta^{(1)}\left(\sqrt{-\hat{g}}\hat{g}^{\mu\nu}\right) = \delta^{(1)}\left(\sqrt{-g}g^{\mu\nu}\right) - \frac{1}{2}\eta^{\mu\nu}\left((n-2)c_{,1} + d_0\right)\hat{R} + d_0\hat{R}^{\mu\nu},$$

и принимая во внимание, что

$$\eta^{\mu\nu}\delta^{(2)}\hat{R}_{\mu\nu} + \delta^{(1)}\left(\sqrt{-\hat{g}}\hat{g}^{\mu\nu}\right)\cdot\delta^{(1)}\hat{R}_{\mu\nu} = \delta^{(2)}\left(\sqrt{-\hat{g}}\hat{R}\right),$$

получаем окончательный ответ с квадратичной точностью

$$S = \int d^n x\sqrt{-\hat{g}}\left(\hat{R} - \frac{(n-2)c_{,1} + d_0}{2}\hat{R}^2 + d_0\hat{R}^{\mu\nu}\hat{R}_{\mu\nu}\right), \tag{4.37}$$

который доказывает присутствие духов всегда, кроме чистой C-теории, то есть $D = 0$.

Случай нелокальных C и D функций

Поскольку рассматриваемая тема в любом случае носит достаточно спекулятивный характер, мы не постесняемся отметить, что можно было бы рассмотреть нелокальные определяющие функции, так что $c_{,i} = c_{,i}(\Box)$ и $d_{,i} = d_{,i}(\Box)$. При этом квадратичное действие можно записать как

$$S = \int d^n x \sqrt{-\hat{g}} \left(\hat{R} - \hat{R} \frac{(n-2)c_{,1}(\Box) + d_0(\Box)}{2} \hat{R} + \hat{R}_{\mu\nu} d_0(\Box) \hat{R}^{\mu\nu} \right).$$

Используя результаты работы [165], можно убедиться, что при условии $d_0 = -3c_{,1} + \frac{2}{\Box}$ (полагая $n = 4$) получаемый пропагатор $\Pi = \frac{2\Pi_{GR}}{3c_{,1}k^2}$ является нелокальной модификацией пропагатора общей теории относительности, не приводящей к новым степеням свободы, если уравнение $c_{,1}(-k^2)k^2 = 0$ не имеет решений. В частности, при $c_{,1} = \frac{1}{\Box} e^{\Box}$ можно получить нелокальность экспоненциального типа, которая иногда рассматривается [165] как средство избавления от ультрафиолетовых расходимостей квантовой гравитации.

Нелокальная g-картина для C-моделей

Снова обратимся к соотношению (4.36) между метриками. Для удобства можно записать его виде $\hat{g}_{\mu\nu} = e^{2\rho} g_{\mu\nu}$, где $C = e^{2\rho}$. Тогда получается

$$\mathcal{R} = R - (n-1)(n-2)(\partial\rho)^2 - 2(n-1)\Box\rho.$$

Несложно найти ρ в первом порядке по возмущениям. Для этого запишем сперва

$$\hat{h}_{\mu\nu} = h_{\mu\nu} + c_{,1}\eta_{\mu\nu} \left(\partial^2_{\alpha\beta} \hat{h}^{\alpha\beta} - \Box \hat{h}^\alpha_\alpha \right), \tag{4.38}$$

и заметим, что изменяется только следовая часть:

$$\hat{h}_{\mu\nu} = h_{\mu\nu} + \frac{1}{n}\eta_{\mu\nu} \left(\hat{h}^\alpha_\alpha - h^\alpha_\alpha \right),$$

а соответственно мы можем найти

$$\hat{h}^\mu_\mu = \frac{1}{1 + (n-1)c_{,1}\Box} \left(h^\mu_\mu + n c_{,1} \partial^2_{\mu\nu} h^{\mu\nu} - c_{,1}\Box h^\alpha_\alpha \right), \tag{4.39}$$

что при нашей точности эквивалентно конформному растяжению с

$$\rho = \frac{1}{2n}(\hat{h}^\alpha_\alpha - h^\alpha_\alpha) = \frac{c_{,1}\left(\partial^2_{\alpha\beta} h^{\alpha\beta} - \Box h^\alpha_\alpha\right)}{2(1 + (n-1)c_{,1}\Box)} = \frac{c_{,1}}{2(1 + (n-1)c_{,1}\Box)} R.$$

Соответственно, действие равно

$$S=\int d^nx\sqrt{-g}\left(R-(n-1)(n-2)(\partial\rho)^2-2(n-1)\Box\rho\right)$$

или, используя интегрирование по частям с сохранением всех поверхностных слагаемых, получаем

$$S=\int d^nx\sqrt{-g}\left(R+R\frac{(n-1)(n-2)c_{,1}^2\Box}{4(1+(n-1)c_{,1}\Box)^2}R\right.$$
$$\left.-\,\partial_\mu\left(R\frac{(n-1)(n-2)c_{,1}^2\partial^\mu}{4(1+(n-1)c_{,1}\Box)^2}R\right)-\frac{(n-1)c_{,1}\Box}{1+(n-1)c_{,1}\Box}R\right),$$

а если поверхностные слагаемые могут быть отброшены (в противном случае, очевидно, надо было использовать C-соотношение до второго порядка, включая вклад с $c_{,2}$), то получаем

$$S=\int d^nx\sqrt{-g}\left(R-(n-1)(n-2)(\partial\rho)^2\right)=$$
$$\int \mathrm{d}^nx\sqrt{-\det(\mathrm{g})}\left(R+R\frac{(n-1)(n-2)c_{,1}^2\Box}{4(1+(n-1)c_{,1}\Box)^2}R\right). \quad (4.40)$$

С точностью до поверхностных слагаемых это действие эквивалентно (4.35), если в операторе $\frac{1}{1+(n-1)c_{,1}\Box}=\sum\limits_{n=0}^{\infty}\left(-(n-1)c_{,1}\Box\right)^n$ можно отбросить все $n\geqslant 1$ члены как дающие поверхностые слагаемые. (На самом деле, это конечно не безобидно.)

Случай нелокальной функции C

Если мы захотим теперь, как в другой картине, рассмотреть нелокальную функцию C, то в случае появления особенностей типа $c_{,1}\sim\frac{1}{\Box}$ отбрасывание "поверхностных" слагаемых даже наивно будет неуместно. В частности, квадратичное действие в g-картине требует учёта коэффициента $c_{,2}$. (Конечно же, при переходе к $\hat{g}$-картине тоже происходит отбрасывание полной дивергенции, только оно, возможно, менее заметно, поскольку абсолютно стандартно: по сути, это линейная часть скалярной кривизны $\partial^2_{\mu\nu}\hat{h}^{\mu\nu} - \Box\hat{h}^{\mu}_{\mu}$.)

Если требуется самосогласованно перейти к одной из картин без отбрасывания полных дивергенций, требуется большая точность. Мы не будем на этом останавливаться, но приведём необходимые формулы во втором порядке:
обратная метрика

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} + h^{\mu\alpha}h^{\nu}_{\alpha} + \mathcal{O}\left(h^3\right), \tag{4.41}$$

детерминант метрики

$$\sqrt{-g} = 1 + \frac{1}{2}h^{\mu}_{\mu} - \frac{1}{4}h_{\mu\nu}h^{\mu\nu} + \frac{1}{8}\left(h^{\mu}_{\mu}\right)^2 + \mathcal{O}\left(h^3\right), \tag{4.42}$$

коэффициенты связности

$$\begin{aligned}\hat{\Gamma}^{\alpha}_{\mu\nu} = \frac{1}{2}&\left(\partial_\mu\hat{h}^{\alpha}_{\nu} + \partial_\nu\hat{h}^{\alpha}_{\mu} - \partial^{\alpha}\hat{h}_{\mu\nu}\right) \\ &- \frac{1}{2}\hat{h}^{\alpha\beta}\left(\partial_\mu\hat{h}_{\beta\nu} + \partial_\nu\hat{h}_{\beta\mu} - \partial_\beta\hat{h}_{\mu\nu}\right) + \mathcal{O}\left(\hat{h}^3\right),\end{aligned} \tag{4.43}$$

смешанная скалярная кривизна

$$\begin{aligned}\mathcal{R} \equiv g^{\mu\nu}\hat{R}_{\mu\nu} = \partial^2_{\mu\nu}\hat{h}^{\mu\nu} - \Box\hat{h}^{\mu}_{\mu} - \frac{1}{2}h^{\mu\nu}&\left(\partial^2_{\mu\alpha}\hat{h}^{\alpha}_{\nu} + \partial^2_{\nu\alpha}\hat{h}^{\alpha}_{\mu} - \Box\hat{h}_{\mu\nu} - \partial^2_{\mu\nu}\hat{h}^{\alpha}_{\alpha}\right) \\ + \frac{1}{4}(\partial_\mu\hat{h}_{\alpha\beta})(\partial^\mu\hat{h}^{\alpha\beta}) - \frac{1}{2}&(\partial_\mu\hat{h}_{\alpha\beta})(\partial^\alpha\hat{h}^{\mu\beta}) \\ + \frac{1}{2}(\partial_\mu\hat{h}^{\mu\alpha})(\partial_\alpha\hat{h}^{\beta}_{\beta}) - \frac{1}{4}&(\partial_\mu\hat{h}^{\alpha}_{\alpha})(\partial^\mu\hat{h}^{\beta}_{\beta}) \\ - \partial_\alpha\left(\hat{h}^{\alpha\beta}\left(\partial_\mu\hat{h}^{\mu}_{\beta} - \frac{1}{2}\partial_\beta\hat{h}^{\mu}_{\mu}\right)\right) + \frac{1}{2}\partial^\mu&\left(\hat{h}^{\alpha\beta}\partial_\mu\hat{h}_{\alpha\beta}\right) + \mathcal{O}\left(\left(h,\hat{h}\right)^3\right),\end{aligned} \tag{4.44}$$

и, наконец, лагранжева плотность

$$
\begin{aligned}
\sqrt{-\det(\mathrm{g})}\mathcal{R} &\approx \left(1+\frac{1}{2}h^{\alpha}_{\alpha}\right)\mathcal{R} \\
&= -\frac{1}{2}(\partial_\mu \hat{h}_{\alpha\beta})\left(\partial^\mu h^{\alpha\beta}-\frac{1}{2}\partial^\mu \hat{h}^{\alpha\beta}\right)+(\partial_\mu \hat{h}_{\alpha\beta})\left(\partial^\alpha h^{\mu\beta}-\frac{1}{2}\partial^\alpha \hat{h}^{\mu\beta}\right) \\
&\quad -\frac{1}{2}\left((\partial_\mu h^{\mu\alpha})(\partial_\alpha \hat{h}^{\beta}_{\beta})+(\partial_\mu \hat{h}^{\mu\alpha})(\partial_\alpha h^{\beta}_{\beta})-(\partial_\mu \hat{h}^{\mu\alpha})(\partial_\alpha \hat{h}^{\beta}_{\beta})\right) \\
&\quad +\frac{1}{2}(\partial_\mu \hat{h}^{\alpha}_{\alpha})\left(\partial^\mu h^{\beta}_{\beta}-\frac{1}{2}\partial^\mu \hat{h}^{\beta}_{\beta}\right) \\
&\quad -\partial_\alpha\left(\hat{h}^{\alpha\beta}\partial_\mu \hat{h}^{\mu}_{\beta}+h^{\mu\beta}\partial_\mu \hat{h}^{\alpha}_{\beta}-\frac{1}{2}\left(\hat{h}^{\alpha\beta}+h^{\alpha\beta}\right)\partial_\beta \hat{h}^{\mu}_{\mu}\right) \\
&\quad +\frac{1}{2}\partial^\mu\left(\left(\hat{h}^{\alpha\beta}+h^{\alpha\beta}\right)\partial_\mu \hat{h}_{\alpha\beta}\right)+\frac{1}{2}\partial_\mu\left(h^{\alpha}_{\alpha}\left(\partial_\nu \hat{h}^{\nu\mu}-\partial^\mu \hat{h}^{\nu}_{\nu}\right)\right) \\
&\quad +\partial^2_{\mu\nu}\hat{h}^{\mu\nu}-\Box \hat{h}^{\mu}_{\mu}+\mathcal{O}\left(\left(h,\hat{h}\right)^3\right),
\end{aligned}
\tag{4.45}
$$

которая при $h=\hat{h}$, как легко видеть, переходит в обычное квадратичное действие для гравитации.

Соотношение между метриками тоже надо учитывать до второго порядка. Оно принимает вид

$$
\hat{h}_{\mu\nu}=\left(1+c_{,1}\delta\mathcal{R}^{(1)}\right)h_{\mu\nu}+\left(c_{,1}\delta\mathcal{R}^{(2)}+\frac{1}{2}c_{,2}\left(\delta\mathcal{R}^{(1)}\right)^2\right)\eta_{\mu\nu}+\mathcal{O}\left(\left(h,\hat{h}\right)^3\right).
\tag{4.46}
$$

4.3.4 Уравнения движения

Итак, мы показали, как вычисления в принципе могут быть проделаны. Теперь пришла пора расплачиваться за легкомыслие, проявленное при работе с определяющим соотношением, содержащим производные. Действительно, уже можно было заметить, что действия, получаемые в разных картинах не вполне эквивалентны.

Особенно ярко это проявляется, конечно, в тех случаях где мы действовали совсем смело, используя нелокальные функции. В частности, если взять $c_{,1}=\frac{2}{3\Box}$ и $D=0$, то наивно получится $\mathcal{L}=\frac{1}{3}R+\frac{2}{81}R\frac{1}{\Box}R$ в одной картине и $\mathcal{L}=\hat{R}-\frac{2}{3}\hat{R}\frac{1}{\Box}\hat{R}$ в другой. Действия похожие, но на самом деле различные, даже с точки зрения числа полюсов пропагатора в соответствии с работой [165] (напомним, что для $\hat{g}$-картины этот выбор функций отвечает отсутствию лишних полюсов).

Выясним теперь, насколько велико будет различие, если коэффициенты $c_{,i}$ являются обычными числами. Для этого полезно посмотреть на уравнения движения.

Сперва проварииируем действие (4.37) в локальной картине:

$$\left(1-(n-2)c_{,1}\hat{R}\right)\hat{R}_{\mu\nu}+(n-2)c_{,1}\left(\hat{\nabla}_\mu\hat{\nabla}_\nu-\hat{g}_{\mu\nu}\hat{\Box}\right)\hat{R} \\ -\frac{1}{2}\left(\hat{R}-\frac{(n-2)c_{,1}}{2}\hat{R}^2\right)\hat{g}_{\mu\nu}=0,$$

что на линейном уровне превращается в

$$\hat{R}_{\mu\nu}-\frac{1}{2}\hat{R}\eta_{\mu\nu}+(n-2)c_{,1}\left(\partial_\mu\partial_\nu-\eta_{\mu\nu}\Box\right)\hat{R}=0,$$

или на языке явных компонент метрики:

$$\partial^2_{\mu\alpha}\hat{h}^\alpha_\nu+\partial^2_{\nu\alpha}\hat{h}^\alpha_\mu-\Box\hat{h}_{\mu\nu}-\partial^2_{\mu\nu}\hat{h}^\alpha_\alpha \\ +\left(2(n-2)c_{,1}\left(\partial^2_{\mu\nu}-\eta_{\mu\nu}\Box\right)-\eta_{\mu\nu}\right)\cdot\left(\partial^2_{\alpha\beta}\hat{h}^{\alpha\beta}-\Box\hat{h}^\alpha_\alpha\right)=0. \quad (4.47)$$

Выберем гармоническую калибровку для $\hat{h}$:

$$\partial_\mu\hat{h}^{\mu\nu}=\frac{1}{2}\partial^\nu\hat{h}^\mu_\mu. \quad (4.48)$$

Тогда из (4.38) легко следует

$$\hat{h}_{\mu\nu}=h_{\mu\nu}-\frac{1}{2}c_{,1}\eta_{\mu\nu}\Box\hat{h}^\alpha_\alpha$$

или

$$\hat{h}^\mu_\mu=\frac{1}{1+\frac{n}{2}c_{,1}\Box}\cdot h^\mu_\mu,$$

а в картине без шляпок это соответствует калибровке

$$\partial_\mu h^{\mu\nu} = \frac{1}{2}\partial^\nu \frac{1 + c_{,1}\Box}{1 + \frac{n}{2}c_{,1}\Box} h^\mu_\mu. \tag{4.49}$$

Применим гармоническую калибровку (4.48) в полученном уравнении движения (4.47):

$$\Box \hat{h}_{\mu\nu} + \frac{1}{2}\left(2(n-2)c_{,1}\left(\partial^2_{\mu\nu} - \eta_{\mu\nu}\Box\right) - \eta_{\mu\nu}\right) \cdot \Box \hat{h}^\alpha_\alpha = 0.$$

Бесследовая часть удовлетворяет волновому уравнению с источником, зависящим от следа, а для следа уравнение принимает вид

$$(n-2)\left(1 + 2(n-1)c_{,1}\Box\right) \cdot \Box \hat{h}^\mu_\mu = 0,$$

или, эквивалентно, в терминах переменной без шляпки

$$(n-2)\frac{1 + 2(n-1)c_{,1}\Box}{1 + \frac{n}{2}c_{,1}\Box} \cdot \Box h^\mu_\mu = 0. \tag{4.50}$$

Совпадает ли это с тем, что можно получить, непосредственно варьируя действие (4.40) нелокальной картины?

Варьируя действие (4.40) нелокальной картины (мы игнорируем любые трудности; по поводу варьирования нелокальных функционалов см. например [166]), получаем

$$R_{\mu\nu} - \frac{1}{2}R\eta_{\mu\nu} - \frac{(n-1)(n-2)c^2_{,1}\Box}{2\left(1 + (n-1)c_{,1}\Box\right)^2}\left(\partial_\mu\partial_\nu - \eta_{\mu\nu}\Box\right) R = 0,$$

что в терминах метрики даёт

$$\begin{aligned} &\partial^2_{\mu\alpha} h^\alpha_\nu + \partial^2_{\nu\alpha} h^\alpha_\mu - \Box h_{\mu\nu} - \partial^2_{\mu\nu} h^\alpha_\alpha \\ &\quad - \left[\eta_{\mu\nu} + \frac{(n-1)(n-2)c^2_{,1}\Box}{\left(1 + (n-1)c_{,1}\Box\right)^2}\left(\partial_\mu\partial_\nu - \eta_{\mu\nu}\Box\right)\right]\left(\partial_{\alpha\beta}h^{\alpha\beta} - \Box h^\alpha_\alpha\right) = 0. \end{aligned}$$

Применяя ранее выбранную калибровку (4.49), имеем

$$\partial^2_{\mu\nu}\frac{1+c_{,1}\Box}{1+\frac{n}{2}c_{,1}\Box}h^\alpha_\alpha - \Box h_{\mu\nu} - \partial^2_{\mu\nu}h^\alpha_\alpha$$
$$+\left[\eta_{\mu\nu} + \frac{(n-1)(n-2)c_{,1}^2\Box}{(1+(n-1)c_{,1}\Box)^2}\left(\partial_\mu\partial_\nu - \eta_{\mu\nu}\Box\right)\right]\frac{1+(n-1)c_{,1}\Box}{2+nc_{,1}\Box}\Box h^\alpha_\alpha = 0$$

и, снова выделяя следовую часть, получаем

$$(n-2)\frac{1+2(n-1)c_{,1}\Box}{(2+nc_{,1}\Box)\cdot(1+(n-1)c_{,1}\Box)}\cdot\Box h^\mu_\mu = 0. \qquad (4.51)$$

Отличие от уравнения (4.50) заключается в нелокальном операторе $\frac{1}{2(1+(n-1)c_{,1}\Box)}$, который происходит от замены переменных $\hat{h}\to h$, содержащей производные. Так бывает всегда. Если мы делаем замену переменных $\phi\to Q\phi$ в действии с некоторым оператором Q, то уравнение движения $\frac{\delta S}{\delta(Q\phi)} = Q^{-1}\frac{\delta S}{\delta\phi}$ умножается на Q^{-1}.

Заметим, что при варьировании действия (4.35) вместо (4.40) мы получили бы еще одну степень нелокального оператора:

$$(n-2)\frac{1+2(n-1)c_{,1}\Box}{(2+nc_{,1}\Box)\cdot(1+(n-1)c_{,1}\Box)^2}\cdot\Box h^\mu_\mu = 0\,. \qquad (4.52)$$

благодаря общему множителю $\frac{1}{(1+(n-1)c_{,1}\Box)}$ в действии. Это ещё одно проявление того факта, что надо уважительно обращаться с классами функций, в которых происходит варьирование, и с полными дивергенциями, когда имеем дело с подобными моделями.

4.3.5 "Игрушечный" пример и обсуждение

Проиллюстрируем встретившиеся сложности на примере простой модели с двумя скалярными полями $\phi(x)$ и $\psi(x)$:

$$S = \int d^n x\cdot(1+\psi(x))\left(\Box\phi(x) - (\partial\phi(x))^2\right)$$

при наложении дополнительного условия

$$\phi = \psi + c_{,1}\left(\Box\phi - (\partial\phi)^2\right) + c_{,1}\psi\Box\phi + c_{,2}(\Box\phi)^2 + \ldots,$$

которое моделирует C-соотношение в применении к "метрикам" $1+\phi$ и $1+\psi$.

Для получения квадратичного действия в ϕ-картине, достаточно найти ψ в линейном приближении. Подставляя

$$\psi = \phi - c_{,1}\Box\phi + \ldots,$$

получаем

$$S = -\int d^n x \cdot \left(2(\partial\phi)^2 + c_{,1}(\Box\phi)^2\right),$$

что даёт уравнение движения

$$2\Box\phi - c_{,1}\Box^2\phi = 0,$$

в котором старшие производные возникли из-за производных в соотношении между полями.

Если нельзя отбрасывать поверхностные слагаемые, то противоположный переход требует определить ϕ с квадратичной точностью:

$$\phi = \frac{1}{1-c_{,1}\Box}\left(\psi + c_{,1}\left(\psi\frac{\Box}{1-c_{,1}\Box}\psi + \left(\partial\frac{\psi}{1-c_{,1}\Box}\right)^2\right)\right.$$
$$\left. + c_{,2}\left(\frac{\Box}{1-c_{,1}\Box}\psi\right)^2 + \ldots\right).$$

Квадратичное действие при этом равно

$$S = \int d^4x \left(\psi \frac{\Box}{1 - c_{,1}\Box}\psi - \left(\partial \frac{\psi}{1 - c_{,1}\Box} \right)^2 \right.$$
$$+ \frac{\Box}{1 - c_{,1}\Box} \left(\psi + c_{,1} \left(\psi \frac{\Box}{1 - c_{,1}\Box}\psi + \left(\partial \frac{\psi}{1 - c_{,1}\Box} \right)^2 \right) \right.$$
$$\left.\left. + +c_{,2} \left(\frac{\Box}{1 - c_{,1}\Box}\psi \right)^2 \right) \right).$$

Пренебрегая же поверхностными членами и производя наивную вариацию, получаем уравнение движения

$$\left(\frac{\Box}{(1 - c_{,1}\Box)^2} + \frac{\Box}{1 - c_{,1}\Box} \right) \psi = 0,$$

которое легко превращается в

$$\frac{1}{1 - c_{,1}\Box}(2\Box\phi - c_{,1}\Box^2)\phi = 0.$$

Разница между картинами чётко соответствует тому оператору, который осуществляет переход между ψ и ϕ. Так у нас получилось и в гравитационной модели. Если же разрешить зависимость типа $c_{,i} \propto \frac{1}{\Box}$, то очевидно, что даже $c_{,2}$-слагаемое в ψ-действии не может быть отброшено.

* * *

Эквивалентность между картинами оказывается зависящей от тонких вопросов, связанных с граничными условиями и определением нелокальных операторов. Взятые как таковые, C- и D-модели, на самом деле, плохо сформулированы и нуждаются в доопределении. Но с другой стороны, в перспективе они могут давать новый подход к нелокальным моделям гравитации.

Глава 5

Разное

В этой главе собраны различные результаты, которые не подпадают под конкретные темы предыдущих трёх глав, но тоже служат общей большой цели. Результаты данной главы опубликованы в статьях [14*], [15*], [16*], [17*], [18*]. Наиболее важные с нашей точки зрения результаты содержатся в работах [14*] и [15*], и мы их выносим в положения на защиту.

В разделе 5.1 мы изучаем обобщения теории относительности, формулируемые на телепараллельном языке с точки зрения локальной лоренц-инвариантности и роли спин связности. В наши дни такие теории являются очень популярными при построении космологических моделей ($f(T)$ гравитация). Однако в основаниях теории до сих пор царит путаница, которую мы проясняем в нашей работе [14*].

В разделе 5.2 рассмотрен вариант модифицированной теории гравитации, который моделирует эффекты тёмной материи, – миметическая гравитация. Это очень молодое направление, возникшее в статье [167], и наша статья [15*] была второй в ряду работ по mimetic dark matter.

В разделе 5.3 случай парадигмы МОНД [47] использован для обсуждения задач, стоящих перед моделями модифицированной гравитации, имеющих своей целью объяснить эффекты тёмной материи.

Разделы 5.4 и 5.5 – единственные, в которых наши исследования непосредственно касаются квантовой теории. Поскольку в работе [168] было выдвинуто предложение использовать в космологиях миров на бранах квантовые потенциалы, возникающие от квантования свободной частицы, живущей на искривлённой поверхности, мы показываем в разделе 5.4, что подобная абстрактная задача плохо определена, и потенциал зависит от метода квантования и от способа задания поверхности (интересно, что авторы статьи [168] ссылаются на нашу работу [17*], основной целью которой было как раз показать эту неоднозначность).

По всей видимости, для последовательного описания таких квантовых эффектов потребуется лучшее понимание квантовой гравитации (впрочем, при переходе к минисуперпространству в квантовой гравитации, наоборот, могут возникать подобные модели [169]). В разделе 5.5 мы обсуждаем информационный парадокс в физике Чёрных Дыр как наиболее перспективный на сегодня путь подступиться к проблемам квантования гравитации.

5.1 Телепараллельные теории гравитации

Очень интересную возможность для модификации гравитации даёт телепараллельная формулировка [170]. В частности, весьма популярными являются $f(T)$ модели. Однако многие аспекты, связанные с локальной лоренц-инвариантностью в пространстве тетрад, были плохо поняты, из-за чего возникало множество ложных ожиданий и надежд. В частности, поскольку $f(T)$ модель не содержит старших производных в действии, можно было надеяться на сохранение числа степеней свободы при переходе от телепараллельного эквивалента общей теории относительности к этому обобщению. Однако, это не так [171] из-за особенностей, связанных с локальными преобразованиями Лоренца в пространстве тетрад.

Наша задача – разобраться с этими тонкостями.

5.1.1 Ковариантная формулировка телепараллельного эквивалента

Как мы уже обсуждали в первой главе, исходная формулировка телепараллельной гравитации, полагая спин-связность равной нулю (связность Вайтценбёка), нарушает локальную лоренц-инвариантность. В этом разделе мы обсудим, как можно сделать теорию ковариантной. Для этого следует ввести явную спин-связность в формализм. Ясно, что ковариантным условием будет требование инерциальности этой спин-связности. Как было отмечено в работах [172, 173] её введение меняет действие лишь на полную производную. Однако всегда надо иметь в виду, и это не всегда достаточно чётко подчеркивается в цитированных выше статьях, что спин-связность должна быть инерциальной. Мы переходим к подробному обсуждению соответствующих деталей, которого не было в литературе до появления нашей работы [14*].

Вариация по отношению к независимой спин-связности

Мы рассматриваем действие телепараллельной гравитации и все входящие в него величины, но явно включаем в них ненулевую спин-связность. Вариации по отношению к коэффициентам спин-связности могут быть найдены точно, поскольку

$$\delta_\omega T^{\alpha}{}_{\mu\nu} = \delta\omega^{\alpha}{}_{\mu\nu} - \delta\omega^{\alpha}{}_{\nu\mu}, \tag{5.1}$$

является точным соотношением при $\delta\omega^{\alpha}{}_{\mu\nu} \equiv e^{\alpha}_{a} e^{b}_{\nu} \delta\omega^{a}{}_{\mu b}$. Имеем

$$\delta_\omega T_\mu = \delta_\omega T^{\alpha}{}_{\mu\alpha} = -\delta\omega^{\alpha}{}_{\alpha\mu}, \tag{5.2}$$

$$\delta_\omega \left(T^\mu T_\mu\right) = -2T_\alpha \delta\omega_\mu{}^{\mu\alpha} + \delta\omega_\mu{}^{\mu\alpha} \cdot \delta\omega^{\nu}{}_{\nu\alpha}, \tag{5.3}$$

$$\delta_\omega \left(T^{\alpha\mu\nu} T_{\mu\alpha\nu}\right) = 2\left(T_{\mu\alpha\nu} - T_{\alpha\mu\nu}\right)\delta\omega^{\alpha\mu\nu} - \left(\delta\omega_{\alpha\mu\nu} - 3\delta\omega_{\mu\alpha\nu}\right) \cdot \delta\omega^{\alpha\mu\nu}, \tag{5.4}$$

$$\delta_\omega \left(T^{\alpha\mu\nu} T_{\alpha\mu\nu}\right) = 4T_{\alpha\mu\nu}\delta\omega^{\alpha\mu\nu} + 2\left(\delta\omega_{\alpha\mu\nu} - \delta\omega_{\mu\alpha\nu}\right) \cdot \delta\omega^{\alpha\mu\nu}, \tag{5.5}$$

где мы использовали то, что, благодаря свойствам симметрии, есть только две независимые свертки T и $\delta\omega$, а именно $T_{\mu\alpha\nu}\delta\omega^{\alpha\mu\nu}$ и $T_{\alpha\mu\nu}\delta\omega^{\alpha\mu\nu}$, и аналогично для свёрток двух $\delta\omega$.

Рассмотрим теперь ковариантизованное действие телепараллельной гравитации

$$S = -\int d^4x \|e\| \cdot \mathbb{T}(e, \omega), \tag{5.6}$$

но не будем учитывать, что спин-связность должна быть инерциальной. Проварииируем независимо по e и ω.

Используя вариации (5.3) – (5.5) и соотношение (1.52), получаем

$$\delta_\omega S = -\int d^4x \|e\| \cdot (T^\mu{}_{\alpha\nu} + 2T_\nu \delta^\mu_\alpha)\, \delta\omega^\alpha{}_\mu{}^\nu.$$

Уравнение движения выходит очень простое

$$T^\mu{}_{\alpha\nu} + T_\nu \delta^\mu_\alpha - T_\alpha \delta^\mu_\nu = 0.$$

После взятия следа (в размерности $d \neq 2$) получаем $T_\mu = 0$, и в конечном итоге

$$T^\mu{}_{\alpha\nu} = 0.$$

Теория оказывается тривиальной, и это не то, чего мы хотим.

Телепараллельное действие с инерциальной спин-связностью

Будем действовать аккуратнее, и потребуем, чтобы спин-связность была инерциальной:

$$\omega^a{}_{\mu b} = -(\Lambda^{-1})^a_c \partial_\mu \Lambda^c_b \tag{5.7}$$

где Λ – произвольная матрица из группы Лоренца. В некоторой системе отсчёта данная спин-связность обращается в нуль, тем самым это ковариантный способ задать геометрию Вайтценбёка. В результате мы рассматриваем действие

$$S_{\mathfrak{W}'} = -\int d^4x \|e\| \cdot \mathbb{T}(e, \omega(\Lambda)) \tag{5.8}$$

с независимыми переменными e и Λ.

Ясно, что модель должна быть эквивалентна обычной телепараллельной гравитации. Формальная причина весьма проста. При вариации по отношению к полю лоренцевых матриц имеем в силу соотношения (1.50)

$$\delta_\Lambda \mathbb{T} = \delta_\Lambda R(\omega) - 2 \overset{(0)}{\bigtriangledown}_\mu (\delta_\Lambda T^\mu),$$

где $\delta_\Lambda(...) = \delta_\omega(...) \cdot \delta_\Lambda \omega$. Поскольку при этом $R(\omega(\Lambda)) \equiv 0$ по определению инерциальной связности, получаем, что $\delta_\omega S_{\mathfrak{W}'}$ является чисто поверхностным слагаемым и не приводит к нетривиальным уравнениям движения. Вариация же по отношению к тетраде проводится при фиксированной (инерциальной) спин-связности, и в системе отсчета, в которой оная обращается в нуль, даёт в точности обычные уравнения движения телепараллельной гравитации. Однако же при личном общении с некоторыми специалистами по телепараллельной гравитации выяснилось, что это утверждение не представляется им очевидным и, даже более того, вызывает сомнения. Полностью разрешить вопрос удалось с помощью приводимых ниже рассуждений.

Заметим, что вариация $\delta_\Lambda \omega$ в классе инерциальных спин-связностей производится применением к матрице Λ произвольного бесконечно малого преобразования Лоренца: $\Lambda \to (\exp \lambda) \cdot \Lambda$ и $\Lambda^{-1} \to \Lambda^{-1} \cdot \exp(-\lambda)$, где матрица $\delta_\Lambda \omega$ принадлежит алгебре Ли группы Лоренца (Осторожно! Это не то же самое преобразование, которое соответствует повороту тетрад матрицей $\exp \lambda$.), $\lambda^{ab} = -\lambda^{ba}$ для $\lambda^{ab} \equiv \eta^{bc} \lambda^a_c$. В первом порядке имеем

$$\delta {\omega^a}_{\mu b} = -(\Lambda^{-1})^a_c (\partial_\mu \lambda^c_d) \Lambda^d_b. \tag{5.9}$$

В вайтценбёковской картине, где $\Lambda = \mathbb{I}$, получаем просто $\delta {\omega^a}_{\mu b} = -\partial_\mu \lambda^a_b$. В противном случае можно посмотреть на вариацию (5.9) с другой точки зрения:

$$-\delta {\omega^a}_{\mu b} = \partial_\mu \left((\Lambda^{-1})^a_c \lambda^c_d \Lambda^d_b \right) - (\partial_\mu (\Lambda^{-1})^a_c) \lambda^c_d \Lambda^d_b - (\Lambda^{-1})^a_c \lambda^c_d (\partial_\mu \Lambda^d_b) = \mathfrak{D}_\mu \tilde{\lambda}^a_b,$$

где $\tilde{\lambda}^a_b \equiv (\Lambda^{-1})^a_c \lambda^c_d \Lambda^d_b$ – преобразование Лоренца матрицы λ к другой системе отсчета, а $\mathfrak{D}$ – лоренц-ковариантная производная с плоской спин-связностью (5.7).

Посмотрим теперь на вариацию дивергенции вектора кручения

$$\delta_\omega \left(\|e\| \overset{(0)}{\nabla}_\mu T^\mu \right) = \delta_\omega \left(\partial_\mu \left(\|e\| T^\mu \right) \right) = -\partial_\mu \left(\|e\| e^\alpha_a \delta\omega^a{}_\alpha{}^b e^\mu_b \right).$$

Можно явно применить производную в правой части, и используя $\partial_\mu \|e\| = \|e\| \cdot e^\beta_c \partial_\mu e^c_\beta$, и $\partial_\mu e^\alpha_a = -e^\beta_a e^\alpha_c \partial_\mu e^c_\beta$, а также аналогично для $\partial_\mu e^\mu_b$, получить

$$\begin{aligned} &-\partial_\mu \left(\|e\| \cdot e^\alpha_a \delta\omega^a{}_\alpha{}^b e^\mu_b \right) \\ &\quad = \|e\| \cdot \left(e^\alpha_c (\partial_\mu e^c_\beta) \delta\omega^\beta{}_\alpha{}^\mu + (e^\mu_c \partial_\mu e^c_\beta - e^\mu_c \partial_\beta e^c_\mu) \delta\omega^\alpha{}_\alpha{}^\beta - e^\alpha_a e^\mu_b \partial_\mu \delta\omega^a{}_\alpha{}^b \right), \end{aligned}$$

где $\delta\omega^\beta{}_\alpha{}^\mu \equiv e^\beta_a e^\mu_b \delta\omega^a{}_\alpha{}^b$. Благодаря антисимметрии ω и определению вайтценбёковского кручения $\overset{\mathfrak{W}}{T}{}^\mu{}_{\alpha\nu} = e^\mu_c \partial_\alpha e^c_\nu - e^\mu_c \partial_\nu e^c_\alpha$ получаем отсюда

$$\begin{aligned} &\|e\| \cdot \left(\overset{\mathfrak{W}}{T}{}^\mu{}_{\alpha\nu} + 2 \overset{\mathfrak{W}}{T}_\nu \delta^\mu_\alpha \right) \delta\omega^\alpha{}_\mu{}^\nu \\ &\qquad = 2\partial_\mu \left(\|e\| \cdot e^\alpha_a \delta\omega^a{}_\alpha{}^b e^\mu_b \right) + \|e\| \cdot e^\alpha_a e^\mu_b \left(\partial_\alpha \delta\omega^a{}_\mu{}^b - \partial_\mu \delta\omega^a{}_\alpha{}^b \right). \end{aligned} \tag{5.10}$$

Это тождество. Предположим теперь для простоты, что мы находимся в вайтценбёковской системе отсчете с $\omega = 0$, $\Lambda = I$, и $\delta\omega^a{}_\mu{}^b = -\partial_\mu \lambda^{ab}$. Тогда в левой части (5.10) получаем вариацию $\delta_\omega S_{\mathfrak{W}'}$, а правая часть обращается в нуль, если подходящие граничные условия наложены на $\delta\omega$.

Конечно, это должно быть справедливо в любой системе отсчета, поскольку действие было явно лоренц-инвариантно. Явно это можно проверить, подставив

$$T^\alpha{}_{\mu\nu} = \overset{\mathfrak{W}}{T}{}^\alpha{}_{\mu\nu} + \omega^\alpha{}_{\mu\nu} - \omega^\alpha{}_{\nu\mu}$$

в соотношение (5.10) и получив слагаемые вида $\omega \cdot \delta\omega$ в левой части. И это ровно те самые слагаемые, которые заменяют последнее слагаемое в правой части на ковариантное

$$\begin{aligned}
&\|e\| \cdot (T^{\mu}{}_{\alpha\nu} + 2T_{\nu}\delta^{\mu}_{\alpha})\, \delta\omega^{\alpha}{}_{\mu}{}^{\nu} \\
&\qquad = 2\partial_{\mu}\left(\|e\| \cdot e^{\alpha}_{a}\delta\omega^{a}{}_{\alpha}{}^{b}e^{\mu}_{b}\right) + \|e\| \cdot e^{\alpha}_{a}e^{\mu}_{b}\eta^{bc}\left(\mathfrak{D}_{\alpha}\delta\omega^{a}{}_{\mu c} - \mathfrak{D}_{\mu}\delta\omega^{a}{}_{\alpha c}\right). \quad (5.11)
\end{aligned}$$

Соотношение (5.11) – ковариантная версия соотношения (5.10). Разумеется, всё, что мы сделали, это явно проверили, что

$$\delta_{\omega}\mathbb{T} = -\delta_{\omega}\left(2\overset{(0)}{\bigtriangledown}_{\mu}T^{\mu}\right) + \delta_{\omega}R(\omega)$$

в соответствии с (1.50). Явные вычисления отнимают некоторое время, но в целом достаточно элементарны.

Подход со множителем Лагранжа

Эквивалентный подход к этим моделям заключается в том, чтобы оставить в действии произвольную спин-связность, но наложить связь $R^{a}{}_{b\mu\nu} = 0$ с помощью множителей Лагранжа:

$$S_{\mathfrak{LW}} = -\int d^4x\|e\| \cdot \left(\mathbb{T}(e,\omega) + \lambda_{a}{}^{b\mu\nu}R^{a}{}_{b\mu\nu}(\omega)\right) \qquad (5.12)$$

где на множитель Лагранжа $\lambda_{a}{}^{b\mu\nu}$ наложены следующие условия симметрии: $\lambda^{ab\mu\nu} = -\lambda^{ab\nu\mu}$ и $\lambda^{ab\mu\nu} = -\lambda^{ba\mu\nu}$.

Вариация действия (5.12) по отношению к λ даёт

$$R^{a}{}_{b\mu\nu}(\omega) = 0$$

эквивалентно условию (5.7), по крайней мере локально. Вариация по отношению к e приводит к

$$\overset{(0)}{\bigtriangledown}_{\alpha}S_{\beta}{}^{\mu\alpha} - S^{\alpha\mu\nu}\left(T_{\alpha\beta\nu} + K_{\alpha\nu\beta}\right) + \frac{1}{2}\left(\mathbb{T} + \lambda_{a}{}^{b\alpha\nu}R^{a}{}_{b\alpha\nu}\right)\delta^{\mu}_{\beta} = 0,$$

что не отличается от (1.64) благодаря связи. Наконец, вариация $\omega^a{}_\nu{}^c$ даёт уравнение на множитель Лагранжа

$$\mathfrak{D}_\mu\left(\|e\|\lambda_a{}^{b\mu\nu}\right)=0$$

или эквивалентно

$$\overset{(0)}{\nabla}_\mu\lambda_\alpha{}^{\beta\mu\nu}-K^\gamma{}_{\mu\alpha}\lambda_\gamma{}^{\beta\mu\nu}+K^\beta{}_{\mu\gamma}\lambda_\alpha{}^{\gamma\mu\nu}=0.$$

Формулировка, не зависящая от выбора спин-связности

В качестве курьёза отметим действие

$$S_{\omega-free}=-\int d^4x\|e\|\left(\mathbb{T}-e^\mu_aR^a{}_{b\mu\nu}(\omega)e^\nu_c\eta^{bc}\right), \qquad (5.13)$$

которое эквивалентно действию Эйнштейна-Гильберта, но при этом вообще не зависит от спин-связности с точностью до поверхностного слагаемого. К сожалению, разумных способов для обобщения не просматривается.

5.1.2 Расширенные телепараллельные теории

Разумеется, телепараллельная гравитация интересна не только сама по себе, но и как средство для формулирования модифицированных теорий. Простейшими примерами являются произвольный квадратичный по кручению лагранжиан $\frac{c_1}{4}T_{\alpha\beta\mu}T^{\alpha\beta\mu}+\frac{c_2}{2}T_{\alpha\beta\mu}T^{\beta\alpha\mu}-c_3T_\mu T^\mu$ (его можно ещё расширить, добавляя не сохраняющие чётность слагаемые), в также знаменитые $f(T)$ теории.

Процедура ковариантизации проходит здесь по-другому. Вариация такого лагранжиана по отношению к плоской спин-связности уже не является поверхностным слагаемым. Однако, она при этом всё-равно не даёт новых уравнений движения. Причина здесь проста. По самому своему построению модель инвариантна по отношению к одновременному преобразованию спин-связности и тетрады. При этом инфинитезимальное лоренцово вращение тетрады – это частный случай её вариации, который даёт антисимметричную часть уравнения движения. В телепараллельном эквиваленте такая вариация обращается в нуль в силу локальной лоренц-инвариантности на уровне чистых тетрад. В расширенных же теориях она не исчезает и приводит к нетривиальной антисимметричной части уравнения движения для тетрады. И, разумеется, эта часть точно совпадает с уравнением движения для спин-связности, поскольку две эти вариации, сделанные одновременно, действия не меняют. Поскольку это простое рассуждение при личном общении со специалистами по телепараллельной гравитации также не сразу встретило понимание, ниже даётся более подробное рассмотрение на уровне вариаций и, в следующем подразделе, на языке уравнений движения, причём с получением новой интересной формы последних.

Рассмотрим нашу модель

$$S_{f(T)} = -\int d^4x \|e\| \cdot f\left(\mathbb{T}(e, \omega(\Lambda)\right). \tag{5.14}$$

и проведём варьирование по инерциальной ω, используя (1.50) и (5.2):

$$\delta_\omega S_{f(T)} = -\int d^4x \|e\| \left(\overset{(0)}{\bigtriangledown}_\mu f'(\mathbb{T})\right) e^\nu_a e^\mu_b \eta^{cb} \delta\omega^a{}_{\nu c}. \tag{5.15}$$

Результат получается нетривиальный даже для чисто инерциальных спин-связностей и вариаций, даваемых формулами (5.7) и (5.9).

Но это ровно та вариация, которая получается из вращения тетрад $e^a_\mu \to \Lambda^a_b e^b_\mu$. Поскольку это лоренцево преобразование, из (1.50) для него получаем $f' \cdot \delta\mathbb{T} = -2f' \cdot \overset{(0)}{\bigtriangledown}_\mu \delta T^\mu$. Можно убедиться, что вариация вектора кручения примет ту же форму, что и $e^\nu_a e^\mu_b \eta^{cb} \delta\omega^a{}_{\nu c}$, где $\delta\omega$ – вариация спин-связности под действием данного преобразования Лоренца. Например, вокруг вайтценбёковского фона вариация спин-связности принимает вид $\delta\omega^a{}_{\alpha b} = -\partial_\alpha \lambda^a_b$ и $\delta T_\mu = e^\alpha_a (\partial_\alpha \lambda^a_b) e^b_\mu$. В то же время, инфинитезимальное вращение тетрад записывается как $\delta e^a_\mu = \lambda^a_b e^b_\mu$. В векторе кручения $\overset{\mathfrak{W}}{T}_\mu = e^\alpha_a \partial_\mu e^a_\alpha - e^\alpha_a \partial_\alpha e^a_\mu$ первый член лоренц-инвариантен (и равен $\frac{1}{\|e\|}\partial_\mu\|e\|$), в то время как вариация второго равна $-e^\alpha_a(\partial_\alpha\lambda^a_b)e^b_\mu$. Что и требовалось доказать.

Проведём те же рассуждения прямо, без обращения к (1.50). Вариация связности дает

$$\delta_\omega S_{f(T)} = -\int d^4x\|e\|f'(\mathbb{T})S_\alpha{}^{\mu\nu}\delta T^\alpha{}_{\mu\nu} = -2\int d^4x\|e\|f'(\mathbb{T})S_\alpha{}^{\mu\nu}e^\alpha_a e^b_\nu \delta\omega^a{}_{\mu b}$$

Используя (5.9), получаем

$$\delta_\omega S_{f(T)} = 2\int d^4x\|e\|f'(\mathbb{T})S_\alpha{}^{\mu\nu}e^\alpha_a(\Lambda^{-1})^a_c(\partial_\mu\lambda^c_d)\Lambda^d_b e^b_\nu.$$

Уравнение движения требует обращения этого выражения в нуль для любой $\lambda^c_d \in so(1,3)$.

Теперь провариируем по отношению к тетрадам,

$$\delta_e S_{f(T)} = -\int d^4x\,(f(\mathbb{T})\delta\|e\| + \|e\|f'(\mathbb{T})\delta\mathbb{T})\,.$$

Вариация δS должна обращаться в нуль для произвольной вариации тетрады, в частности для $e \to (\exp\tilde{\lambda})\cdot e$, или инфинитезимально $\delta e^a_\mu = \tilde{\lambda}^a_b e^b_\mu$. В таком случае $\delta\|e\| = 0$, а $\delta_e\mathbb{T} = S_\alpha{}^{\mu\nu}\delta_e T^\alpha{}_{\mu\nu}$ поскольку $\delta_e g_{\mu\nu} = 0$. Принимая во внимание, что в нашем случае $\delta_e T^\alpha{}_{\mu\nu} = e^\alpha_a\left((\mathfrak{D}_\mu\tilde{\lambda}^a_b)e^b_\nu - (\mathfrak{D}_\nu\tilde{\lambda}^a_b)e^b_\mu\right)$, имеем

$$\delta_e S_{f(T)}\Big|_{\delta e=\tilde{\lambda}e} = 2\int d^4x\|e\|f'(\mathbb{T})S_\alpha{}^{\mu\nu}e^\alpha_a(\mathfrak{D}_\mu\tilde{\lambda}^a_b)e^b_\nu.$$

Как мы показали после формулы (5.9), если положить $\tilde{\lambda}^a_b = -(\Lambda^{-1})^a_c\lambda^c_d\Lambda^d_b$, то получится

$$\mathfrak{D}_\mu\tilde{\lambda}^a_b = -(\Lambda^{-1})^a_c(\partial_\mu\lambda^c_d)\Lambda^d_b.$$

Тем самым мы видим $\delta_e S_{f(T)}\big|_{\delta e=\tilde{\lambda}e} = -\delta_\omega S_{f(T)}$. Уравнения движения, следующие из δ_ω в инерциальном классе уже содержатся в качестве частного случае в уравнениях движения для тетрады.

Ещё раз подчеркнём, что это совершенно общее свойство ковариантизированных теорий этого типа: вариация $\Lambda^a_b \to (\exp(\lambda))^a_c \Lambda^c_b$ в спин-связности приводит ровно к тому же изменению действия, только с противоположным знаком, что и вращение тетрад $e^a_\mu \to (\exp(\tilde{\lambda}))^a_c e^c_\mu$ с подходящим $\tilde{\lambda}$. Поэтому и новых уравнений из вариации спин-связности не получается.

5.1.3 Новая форма уравнений движения для $f(T)$ гравитации

Рассмотрим теперь ту же проблему на уровне уравнений движения для действия

$$S_{f(T)} = -\int d^4x \|e\| \cdot f\left(\mathbb{T}(e, \omega(\Lambda)\right). \tag{5.16}$$

Сперва рассмотрим вариацию тетрады. Получаем

$$\delta S_e = -\int d^4x \|e\| \cdot \left(f(\mathbb{T}) e^\mu_a \delta e^a_\mu + f'(\mathbb{T}) \delta \mathbb{T} \right).$$

В случае инерциальных спин-связностей, тождество (1.50) даёт

$$\mathbb{T} = -\overset{(0)}{R} - 2 \overset{(0)}{\triangledown}_\mu T^\mu,$$

и следовательно

$$\delta S_e = -\int d^4x \|e\| \cdot \left(f(\mathbb{T}) e^\mu_a \delta e^a_\mu - f'(\mathbb{T}) \delta \overset{(0)}{R} + 2 \left(\partial_\mu f'(\mathbb{T})\right) \delta(g^{\mu\nu} T_\nu) \right).$$

Учитывая, что $\delta g_{\mu\nu} = 2\eta_{ab} e^a_\mu \delta e^b_\nu$ и $\delta g^{\mu\nu} = -\left(g^{\mu\rho} e^\nu_a + g^{\nu\rho} e^\mu_a\right) \delta e^a_\rho$, получаем

$$\delta S_e = -\int d^4x \|e\| \cdot \left(f e^\mu_a \delta e^a_\mu + 2 \left(\left(\overset{(0)}{R}{}^{\mu\nu} - \overset{(0)}{\triangledown}{}^\mu \overset{(0)}{\triangledown}{}^\nu + g^{\mu\nu} \overset{(0)}{\Box} \right) f' \right) \eta_{ab} e^a_\mu \delta e^b_\nu \right.$$
$$\left. - 2 \left(\left(\partial^\mu f'\right) T_\nu + \left(\partial_\nu f'\right) T^\mu \right) e^\nu_a \delta e^a_\mu + 2 \left(\partial^\mu f'\right) \delta T_\mu \right).$$

Теперь следует проварииировать кручение (1.60)

$$\delta_e T^\alpha{}_{\mu\nu} = -e^\alpha_a T^\beta{}_{\mu\nu} \delta e^a_\beta + e^\alpha_a \left(\mathfrak{D}_\mu \delta e^a_\nu - \mathfrak{D}_\nu \delta e^a_\mu\right).$$

Отметив, что коэффициенты связности из первого слагаемого справа превращают лоренц-ковариантные производные второго слагаемого в полные ковариантные производные δe, и учитывая, что полные ковариантные производные самих тетрад обращаются в нуль, получаем замечательное соотношение, которое несмотря на свою простоту и элегантность в литературе нам не встречалось:

$$\delta_e T^\alpha{}_{\mu\nu} = \triangledown_\mu \left(e^\alpha_a \delta e^a_\nu\right) - \triangledown_\nu \left(e^\alpha_a \delta e^a_\mu\right), \tag{5.17}$$

и в частности

$$\delta_e T_\mu = \partial_\mu \left(e^\alpha_a \delta e^a_\alpha\right) - \triangledown_\alpha \left(e^\alpha_a \delta e^a_\mu\right).$$

Учитывая различие между телепараллельной связностью и связностью Леви-Чивита, получаем

$$\delta_e T_\mu = \partial_\mu \left(e^\alpha_a \delta e^a_\alpha\right) - \overset{(0)}{\triangledown}_\alpha \left(e^\alpha_a \delta e^a_\mu\right) - K^\alpha{}_{\alpha\nu} e^\nu_a \delta e^a_\mu + K^\nu{}_{\alpha\mu} e^\alpha_a \delta e^a_\nu.$$

Теперь уравнение движения вывести совсем несложно, и оно принимает сравнительно простой вид

$$f'(\mathbb{T}) \cdot \overset{(0)}{R}{}^{\mu\nu} + K^{\mu\nu\alpha} \partial_\alpha f'(\mathbb{T}) - T^\mu \partial^\nu f'(\mathbb{T}) + \frac{1}{2} f(\mathbb{T}) \cdot g^{\mu\nu} = 0, \tag{5.18}$$

где мы воспользовались тем, что $K^\alpha{}_{\alpha\nu} = -T_\nu$. Обратим внимание на то, что слагаемые со старшими производными, типичные для $f(R)$ гравитации, сократились, как и должно было быть.

Очевидно, что уравнение (5.18) содержит антисимметричную часть:

$$T^{\alpha\mu\nu} \partial_\alpha f'(\mathbb{T}) + T^\nu \partial^\mu f'(\mathbb{T}) - T^\mu \partial^\nu f'(\mathbb{T}) = 0, \tag{5.19}$$

где мы учли, что $K^{\mu\nu\alpha} - K^{\nu\mu\alpha} = T^{\alpha\mu\nu}$. Наша задача показать, что полученное уравнение совпадает с уравнением движения для спин-связности.

В самом деле, проварииируем спин-связность (5.15) в инерциальном классе, в котором $\delta\omega^a{}_\nu{}^b = \mathfrak{D}_\nu \tilde{\lambda}^{ab}$ с произвольной антисимметричной $\tilde{\lambda}$:

$$\delta_\omega S_{f(T)} = -\int d^4x \|e\| \left(\partial_\mu f'(\mathbb{T})\right) e_a^\mu e_b^\nu \mathfrak{D}_\nu \tilde{\lambda}^{ab}.$$

Интегрируя по частям, получаем

$$\delta_\omega S = \int d^4x \left(\partial_\mu f'\right) \cdot \left(\partial_\nu \left(\|e\| e_a^\mu e_b^\nu\right) - \|e\| \left(\omega^c{}_{\nu a} e_c^\mu e_b^\nu + \omega^c{}_{\nu b} e_a^\mu e_c^\nu\right)\right) \tilde{\lambda}^{ab}.$$

Используя $\partial_\nu \|e\| = \|e\| \overset{(0)}{\Gamma}{}^\alpha_{\nu\alpha}$ и обращение в нуль полной ковариантной производной тетрады, получаем

$$\delta_\omega S = -\int d^4x \|e\| \left(\partial_\mu f'\right) \cdot \left(K^\alpha{}_{\alpha\beta} e_a^\mu e_b^\beta + \Gamma^\mu_{\nu\beta} e_a^\beta e_b^\nu\right) \tilde{\lambda}^{ab}.$$

Благодаря антисимметрии $\tilde{\lambda}$ снова воспроизводим (5.19).

5.1.4 Подведение итогов

Итак, мы рассмотрели несколько вариантов ковариантизации.

1. "Вариация Вайтценбёка": ω полностью фиксирована. (В исходной формулировке положена равной нулю). Лагранжиан сам по себе не инвариантен, но есть свобода смены системы отсчёта путём изменения лагранжиана.

2. "Независимая вариация" совершенно произвольной спин-связности, см. (5.6). Приводит к тривиальной модели, $T = 0$.

3. "Инерциальная вариация": варьируем ω в классе инерциальных спин-связностей, см. (5.7) и (5.9). Независимые переменные – тетрада и матрица Лоренца. Это настоящая ковариантизация, пригодная и для телепараллельного эквивалента и для модифицированных теорий.

4. "Вариация со связями": эквивалентна предыдущей, но вместо явного задания инерциальной спин-связности используется лагранжева связь (5.12).

5. "Компенсированная вариация": полностью отщепившаяся спин-связность ω, см. (5.13). Строго говоря, это уже не телепараллельная модель. Гравитацию можно описывать как произвольную (по нашему желанию) комбинацию кривизны и кручения. По-видимому, не подходит для обобщений за пределами телепараллельного эквивалента, по крайней мере наивно.

5.2 Об одной модели (Mimetic Dark Matter) эффективной Тёмной Материи

В работе [167] была предложена интересная модель, в рамках которой метрика в действии Эйнштейна-Гильберта была взята в виде

$$g_{\mu\nu} = \tilde{g}_{\mu\nu}\tilde{g}^{\alpha\beta}(\partial_\alpha\phi)(\partial_\beta\phi) \tag{5.20}$$

со вспомогательной метрикой $\tilde{g}_{\mu\nu}$ и скалярным полем ϕ. Очевидно, что общий множитель (конформная мода) в метрике $\tilde{g}$ не несёт никакой смысловой нагрузки, и его роль передана скалярному полю.

Однако непосредственно из определения (5.20) следует, что помимо уравнения движения

$$\nabla_\mu \left((G - T)\partial^\mu\phi\right) = 0, \tag{5.21}$$

скалярное поле подвержено действию связи

$$g^{\alpha\beta}(\partial_\alpha\phi)(\partial_\beta\phi) = 1. \tag{5.22}$$

Подставляя это в уравнения Эйнштейна

$$G_{\mu\nu} - T_{\mu\nu} - (G - T)(\partial_\mu\phi)(\partial_\nu\phi) = 0, \tag{5.23}$$

где $G_{\mu\nu}$ и $T_{\mu\nu}$ – тензор Эйнштейна и тензор энергии-импульса материи соответственно, а буквы без индексов обозначают следы, мы обнаруживаем, что следовая часть уравнения удовлетворена автоматически.

Соответственно, в системе имеется дополнительная свобода. И эффективно присутствует вклад в тензор энергии-импульса вида

$$\tilde{T}_{\mu\nu} = (G - T)(\partial_\mu \phi)(\partial_\nu \phi), \tag{5.24}$$

который может быть интерпретирован как присутствие пылевидной материи, названной Mimetic Dark Matter, начальное распределение которой оказывается константой интегрирования [167].

Следуя нашей работе [15*], мы объясним, как простое перенесение степени свободы из одного сектора в другой повлияло на уравнения движения, а также предложим эквивалентную, но во многом более удобную, формулировку теории.

5.2.1 Структура вариационного принципа

Без потери общности, предположим, что $\tilde{g}$ унимодулярна. В таком случае детерминант g определяется четвёртой степенью множителя

$$\Omega(x) \equiv \tilde{g}^{\alpha\beta}(\partial_\alpha \phi)(\partial_\beta \phi).$$

Чтобы получить стандартные уравнения Эйнштейна, надо варьировать действие

$$-\int d^4 x R(g(\tilde{g}, \phi))$$

по отношению к метрике g, включая множитель Ω, с тем лишь ограничением (будем считать, что поверхностные слагаемые со старшими производными выкинуты или введён член Гиббонса-Хокинга), чтобы вариация обращалась в нуль на границах области интегрирования, и в частности – в начальный и конечный моменты времени. Это, однако, несовместимо с нашим определением Ω.

Рассмотрим, для простоты, пространственно однородный случай, $\phi(t) = t + \delta\phi(t)$ во фридмановской вселенной $ds^2 = dt^2 - a^2(t) d\overrightarrow{x}^2$. Тогда $\dot{\phi} = \sqrt{\Omega}$. Требуя, чтобы $\delta\phi = 0$ при $t = t_{in}$ и $t = t_{fin}$, получаем $\int\limits_{t_{in}}^{t_{fin}} dt \sqrt{\Omega} = t_{fin} - t_{in}$ для любой допустимой вариации. Соответственно, вариирование производится в ограниченном классе функций. Условию стационарности действия оказывается проще удовлетворить, и динамика допускает большую свободу.

Собственно, это совершенно общий эффект при подстановках в действии, содержащих производные. (Важное исключение – трюк Штюкельберга, поскольку он не меняет класса функций.) Положим, мы имеем действие $S = \int \dot{x}^2 dt$, и заменяем $x \equiv \dot{y}$ прямо в нём. После этого уравнения движения становятся более высокого порядка, даже в терминах исходной переменной $x(t)$, и требуется больше данных Коши для определения эволюции. Причина та же самая, что и выше. Вариирование производится в классе функций, для которых не только δx обращается в нуль на границе, но ещё требуется дополнительно, чтобы $\int\limits_{t_{in}}^{t_{fin}} \delta x(t) dt = 0$.

Отметим для полноты, что миметическую гравитацию можно понять и в более общих терминах обратимости дисформных (а не только конформных) преобразований метрики [174].

5.2.2 Эквивалентная формулировка

Предложим теперь эквивалентную формулировку модели. В качестве первого шага введём множители Лагранжа $\lambda^{\mu\nu}$ и перепишем действие как

$$S = -\int \left(R(g) + \lambda^{\mu\nu}\left(g_{\mu\nu} - \tilde{g}_{\mu\nu}\tilde{g}^{\alpha\beta}(\partial_\alpha\phi)(\partial_\beta\phi)\right)\right)\sqrt{-g}d^4x.$$

Мы игнорируем в нашем рассмотрении лагранжиан материи. Эффекты материи легко восстанавливаются во всём рассуждении с помощью вычитания $T_{\mu\nu}$ из $G_{\mu\nu}$ во всех формулах.

Вариация по отношению к λ накладывает связь (5.20). Вариация по отношению к ϕ даёт

$$\bigtriangledown_\mu\left(\lambda\partial^\mu\phi\right) = 0, \tag{5.25}$$

где $\lambda \equiv \lambda^\mu_\mu \equiv g_{\mu\nu}\lambda^{\mu\nu}$. Используя (5.20), получаем уравнения Эйнштейна

$$G_{\mu\nu} + \lambda_{\mu\nu} = 0,$$

и видим, что (5.25) эквивалентно (5.21). Наконец, вариация по отношению к $\tilde{g}$ приводит к

$$\lambda^{\mu\nu}\tilde{g}^{\alpha\beta}(\partial_\alpha\phi)(\partial_\beta\phi) - \lambda^{\rho\sigma}\tilde{g}_{\rho\sigma}\tilde{g}^{\mu\alpha}(\partial_\alpha\phi)\tilde{g}^{\nu\beta}(\partial_\beta\phi) = 0.$$

Объединяя это с (5.20) получаем

$$\lambda_{\mu\nu} = \lambda(\partial_\mu\phi)(\partial_\nu\phi). \tag{5.26}$$

Подставляя полученный результат в уравнение Эйнштейна и отделяя следовую часть, находим эффективный тензор энергии-импульса (5.24). В частности, для пространственно однородного поля ϕ имеем $\partial_i\phi = 0$, а уравнение движения (5.25) воспроизводит обычное поведение тёмной материи $\lambda \propto \frac{1}{a^3}$, что не должно быть сюрпризом, поскольку правильно соответствует виду тензора T.

Очевидно, что $\lambda_{\mu\nu}$ полностью определяется своим следом. Поэтому можно догадаться до более простого действия

$$S = -\int \left(R(g) + \lambda\left(1 - g^{\mu\nu}(\partial_\mu\phi)(\partial_\nu\phi)\right)\right)\sqrt{-g}d^4x. \tag{5.27}$$

Мы утверждаем, что это эквивалентная формулировка. В самом деле, вариация по λ сразу дает связь (5.22). Уравнения Эйнштейна принимают вид

$$G_{\mu\nu} + \lambda(\partial_\mu\phi)(\partial_\nu\phi) = 0 \tag{5.28}$$

как и раньше. Наконец, вариация по ϕ даёт уравнение движения (5.25).

* * *

Наша формулировка (5.27) стала использоваться почти во всех исследованиях по миметической гравитации. Отметим, что были предложены разные модификации миметической гравитации – $f(R)$ такого типа [175], старшие производные ($\Box\phi$ в действии) для поля ϕ [176]... Но с нашей точки зрения было бы интересно попытаться сделать полноценное динамическое поле из множителя Лагранжа.

5.3 Замечания о парадигме МОНД

Как уже обсуждалось в первой главе, одной из основ современной космологической картины мира является представление о Тёмной Материи. Её история началась в 30-е годы XX века, когда применение известной из механики теоремы вириала к скоплениям галактик привело к проблеме. Рентгеновская светимость горячего межгалактического газа, с точки зрения соответствия средней кинетической и средней потенциальной энергий, потребовала гораздо более глубокой потенциальной ямы, чем могла обеспечить только видимая материя.

Полвека спустя при изучении кривых скоростей вращения звёзд в спиральных галактиках выяснилось, что скорости периферических звёзд выходят на плато, а не следуют спадающей кривой ньютоновского типа: $v \sim \frac{1}{\sqrt{r}}$. В наши дни Тёмная Материя является важным ингридиентом для понимания процессов образования крупномасштабной структуры во Вселенной, а также необходима для правильного предсказания флуктуаций реликтового фона и барионных акустических осцилляций.

Как мы тоже уже обсуждали, современная космология, органически включая в себя гипотезу о Тёмной Материи, прекрасно описывает наблюдательные данные, по крайней мере когда речь идёт о реликтовом фоне и крупномасштабной структуре Вселенной. Вместе с тем есть целый ряд сложностей на "малых" масштабах [45, 177, 178]. Это, например, проблема малого количества карликовых спутников у больших галактик, таких как Млечный Путь или Туманность Андромеды, галактические профили плотности с гладким кором вместо сингулярного каспа...

Впрочем, эти проблемы связаны с малыми масштабами, на которых теория становится сильно нелинейной. Предсказания делаются с помощью численного моделирования, в котором не учитывается множество факторов, связанных с "барионной" физикой. Возможно, проблема просто заключается в недостаточно аккуратном моделировании. Тем более, что последние улучшенные численные модели показывают лучшее согласие [46], возможно, объясняя все расхождения.

С другой стороны, и сама гипотеза Тёмной Материи, возможно, излишне проста и должна быть дополнена самодействием в тёмном секторе и/или переходом к тёплой [179] или нечёткой (fuzzy, сверхлёгкие аксионоподобные частицы, например, из теории струн [15]) Тёмной Материи (вместо холодной).

Впрочем, существуют аргументы в пользу того, что некоторые проблемы (такие как существование дисковых систем без центрального балджа) могут быть и нечувствительны к модификациям в "барионном" секторе взаимодействия звёзд и газа [178]. А некоторые предсказываемые стандартной картиной спутники представляются слишком большими, чтобы "барионные" эффекты отдачи могли их сдуть, не разрушая всей структуры в целом (too big to fail).

С другой стороны, часто в качестве серьёзных проблем предъявляются такие проблемы (например, слишком большие пекулярные скорости, или тёмный поток, dark flow [180]), которых, по всей видимости, и вовсе в реальности не существует [181, 182].

Таким образом, стандартная ΛCDM космология прекрасно работает в тех масштабах, в которых предсказания делаются надёжно (линейный и почти линейный режимы), но по-видимому, испытывает серьёзные трудности в меньших масштабах, где, впрочем, уверенные предсказания затруднительны.

Что такое МОНД?

Модифицированная ньютоновская динамика (МОНД) была предложена как средство описания плато в кривых скоростей вращения в периферических частях галактик.

Цель достигается модификацией закона, который связывает притягивающую массу с центростремительным ускорением на круговой орбите заданного радиуса. Эта идея обсуждалась как на языке модифицированной силы притяжения, так и на языке модифицированной инерции – изменённого второго закона Ньютона, но последний вариант очевидно приводит к нарушению сохранения импульса при взаимодействии тел неравных масс.

Оказывается, что широкий круг данных можно описать в рамках чрезвычайно простой модификации. Вводится некоторая новая фундаментальная константа природы размерности ускорения

$$a_0 \approx 1.2 \cdot 10^{-8}\ \frac{\mathrm{cm}}{\mathrm{s}^2}$$

и постулируется, что если в рамках ньтоновской динамики центростремительное ускорение получается $a \gg a_0$, то это и есть правильный ответ, однако же в случае $a \ll a_0$ ускорения убывают как $\frac{1}{r}$ вместо $\frac{1}{r^2}$, что можно описать правилом $a = \sqrt{a_N a_0}$, где a_N – ускорение, получающееся по Ньютону. В промежуточном режиме применяется некоторая простая интерполирующая функция, хотя неплохих результатов можно достичь и при резком переходе от одного режима к другому при $a = a_0$.

Удивительно, но столь несложная идея хорошо работает с одним на все случаи жизни фиксированным фундаментальным ускорением. Более того, оказывается естественно объяснённым соотношение Талли-Фишера [183]. Оно применяется в астрономии для определения расстояний, и связывает полную светимость спиральной галактики со значением линейной скорости вращения на периферическом плато v_∞ простым соотношением $L \sim v_\infty^4$. В стандартной парадигме это соотношение не имеет простой интерпретации. В рамках же МОНД периферические области можно описать динамическим законом с силой притяжения, равной

$$F = m\frac{\sqrt{GMa_0}}{r}. \tag{5.29}$$

Поскольку центростремительное ускорение равно $\frac{v_\infty^2}{r}$, получаем $v_\infty = \sqrt[4]{GMa_0}$, что тривиально объясняет соотношение Талли-Фишера, ведь полная светимость пропорциональна количеству звёзд, а с ним и массе. Более того, было показано, что при замене светимости на полную видимую массу соотношение работает только лучше [184].

Разумеется, разумная теория должна быть сформулирована и для более общих ситуаций, желательно – без нарушения сохранения импульса. В нерелятивистском пределе есть достаточно общепринятая формулировка в виде нелинейной модификации уравнения Пуассона

$$\nabla\left(\mu\left(\frac{|\nabla\phi|}{a_0}\right)\cdot\nabla\phi\right)=4\pi G\rho \tag{5.30}$$

для ньютоновского потенциала гравитации ϕ, где $\mu(x)$ – гладкая интерполирующая функция между ньютоновским режимом $\mu(x)|_{x\gg1}\to 1$ и глубоким МОНДом $\mu(x)|_{x\ll1}\to x$. В сферически симметричных ситуациях имеем

$$a_N=\mu\left(\frac{a}{a_0}\right)\cdot a, \tag{5.31}$$

как и ранее. В общем случае полученное уравнение очень сложное и нелинейное, причём в режиме малых градиентов, из-за чего с ним непросто работать.

Есть также альтернативная, более простая для работы, формулировка (квази-линейный МОНД, QUMOND): следует сперва найти ньютоновский потенциал $\phi^{(N)}$ в рамках обычной теории Ньютона, а потом получить МОНДовский потенциал ϕ, решив уравнение

$$\triangle\,\phi=\nabla\left(\nu\left(\frac{|\nabla\phi^{(N)}|}{a_0}\right)\cdot\nabla\phi^{(N)}\right). \tag{5.32}$$

В сферически симметричных случаях оба предписания совпадают, но вообще говоря это разные теории.

Приняв одну из подобных формулировок, теорию уже можно применить к более сложным ситациям. И первая неудача приходит обескураживающе быстро, поскольку сила не только медленнее убывает с ростом расстояния, но и медленнее нарастает с ростом притягивающей массы, в сравнении с классической гравитацией. В результате притягивающая сила оказывается недостаточной для корректного описания скоплений галактик. Стандартное решение заключается во введении некоторого количества тёмной материи также и в рамках МОНДа [185]. Правда, требуется сравнительно небольшое количество новой материи, и обычно говорят, что это могут быть просто массивные нейтрино.

Интересно отметить, что при этом численное моделирование (N-body simulations) показывает скорее проблематично усиленный рост структур в разных версиях МОНД [186–188]. По-видимому, это можно связать с увеличением притягивающей силы на ранней стадии коллапса, в режиме очень малых ускорений и контрастов плотности, что, впрочем, строго говоря, должно потребовать полноценной релятивистской реализации парадигмы.

Понятно, что релятивистская реализация совершенно необходима для рассмотрения теории в общем космологическом контексте, и с этим связана большая трудность. Необходимость последовательной фундаментальной теории была ясна с самого начала, и даже была предложена скалярно-тензорная теория [189], которая использует дополнительное скалярное поле типа k-эссенции со специально подобранным видом кинетической функции при малых градиентах скалярного поля. Эта идея однако не приводит успеху, поскольку взаимодействие со скаляром, как присутствующее только в следе тензорных величин, не способно повлиять на отклонение лучей света (в режиме слабого линзирования), а следовательно приводит к противоречию с определением массы галактик и скоплений галактик по гравитационному линзированию.

На сегодняшний день существует несколько версий релятивистской реализации МОНД, и наиболее детально изученная и, пожалуй, наиболее консервативная (есть варианты вплоть до гравитационных диполей) версия – это $TeVeS$ (тензорно-векторно-скалярная, [190]), которая дополняет классическую скалярно-тензорную модель [189] векторным полем типа эйнштейновского эфира, специально так подобранным, чтобы обеспечить правильное линзирование.

В теории $TeVeS$ были изучены линейные космологические возмущения [191]. Нетривиальность успеха стандратной космологической модели можно оценить по достоинству, опираясь хотя бы на тот факт, что и после всех описанных ухищрений космология на основе МОНД не смогла объяснить высоты третьего пика в температурных флуктуациях реликтового фона [191]. В то же время, для флуктуаций барионной плотности был получен удовлетворительный ответ, но было аналитически показано [192], что важную роль в обеспечении гравитационной неустойчивости при этом сыграли векторные степени свободы, что по иронии судьбы возвращает нас к "тёмным" степеням свободы.

Стандартная реакция сообщества МОНД на такие неудачи – это заявление о том, что это проблемы конкретных реализаций, а не самой парадигмы. До некоторой степени это конечно же верно, но делает гипотезу МОНД практически нефальсифицируемой за пределами той узкой области, для которой она была введена, что, разумеется, совершенно не подходит для поисков фундаментальных законов природы.

Сторонники МОНД часто склонны вообще при этом игнорировать все успехи современной космологии. Некоторые авторы [177] даже строят некий "график уверенности в теории" ("theory confidence graph"), в котором каждую проблему считают за явный провал модели, причём независимый от всех остальных, и приходят к выводу о практически исчезающей степени уверенности в парадигме ΛCDM. С таким подходом мы, конечно, не можем согласиться.

О гравитационном коллапсе в МОНД

Вообще говоря, в рамках МОНД многие события могут происходить отлично от привычной теории, и наблюдательные миссии, которые будут детально изучать распределение Тёмной Материи, потенциально могут дать интересную информацию по этому вопросу.

Рассмотрим, например, коллапс однородного облака пыли. В ньютоновской механике движение каждой частицы описывается уравнением $\frac{d^2r}{dt^2} = -\frac{GM(r)}{r^2}$, которое приводит к хорошо известному результату о том, что время падения на центр $T = \sqrt{\frac{3\pi}{32G\rho}}$ не зависит от начального радиуса. В случае МОНД будем иметь $\frac{d^2r}{dt^2} = -\frac{\sqrt{GM(r)a_0}}{r}$ в режиме $a \ll a_0$, который должен выполняться в центральных областях. В силу более медленной зависимости правой части от радиуса видим, что самые внутренние части коллапсируют быстрее.

В случае более реалистичных профилей плотности, чем просто константа, режим МОНД наступает также на периферии, далеко от центрального скопления массы. Там ускорение будет убывать медленнее, чем в ньютоновской динамике, и периферические слои будут также уплотняться быстрее. Как видим, формирование структур в рамках МОНД, даже в чисто сферически симметричном рассмотрении, может сильно отличаться от ньютоновского.

Интересно, что промежуточные слои, оказывающиеся в ньютоновском режиме, не ощущают того усиленного законом МОНДа притяжения, которое имеется для самых внутренних частей. В принципе, это может приводить к появлению областей с отрицательной плотностью тёмной материи при попытке переписать решения теории МОНД на языке присутствия дополнительной массы. Это явление было обнаружено в работе [193] и может в случае обнаружения дать очень сильный аргумент в пользу модифицированной гравитации вместо тёмной материи.

Неизбежная нелинейность

Режим МОНД существенно нелинеен, что приводит к нарушению принципа суперпозиции, да иначе и быть не могло при силе, зависящей от $\sqrt{M}$. И только благодаря этой нелинейности в этой теории тоже справедливо утверждение о том, что частица, находящаяся внутри сферического слоя материи, не чувствует никакой силы (это легко получить, применяя теорему Стокса к уравнению (5.30)). Разумеется, и в МОНДе, и в стандартной гравитации это утверждение не совсем верно, если учесть возмущающее влияние частицы на сферический слой, внутрь которого она внесена. Однако можно проверить [194], что в стандартной теории этот эффект гораздо сильнее подавлен, чем в моделях МОНД. Причина в том, что уменьшение ньютоновской силы как $\frac{1}{r^2}$ прекрасно согласовано с ростом площади r^2 в заданном телесном угле, тем самым делая такую компенсацию весьма устойчивой, благодаря принципу суперпозиции. В МОНДе же несоответствие закона $\frac{1}{r}$ геометрии компенсируется лишь благодаря тонко подобранной нелинейности.

Фундаментальным вопросом будет, до каких масштабов следует считать закономерности МОНД действующими. Надо ли думать, что гравитационные взимодействия элементарных частиц тоже подчиняются этим законам? Выясним, какой массы точечная частица будет создавать на расстоянии, равном её комптновской длине волны $r_C \equiv \frac{h}{mc}$, поле, необходимое для перехода из МОНДовского в ньютоновский режим (ранее, в несколько другом контексте, этот масштаб возникал в работе [47] и ряде других):

$$m_\star = \sqrt[3]{\frac{a_0 h^2}{Gc^2}} \approx \sqrt[3]{\frac{1.2\cdot 10^{-8}\ \frac{\mathrm{cm}}{\mathrm{s}^2}\cdot\left(6.63\cdot 10^{-27}\ \frac{\mathrm{g}\cdot\mathrm{cm}^2}{\mathrm{s}^2}\right)^2}{6.67\cdot 10^{-8}\ \frac{\mathrm{cm}^3}{\mathrm{s}^2\cdot\mathrm{g}}\cdot\left(3\cdot 10^{10}\ \frac{\mathrm{cm}}{\mathrm{s}^2}\right)^2}} \approx 2\cdot 10^{-25}\ \mathrm{g}.$$

Эта масса явно больше массы электрона $m_e \approx 9.11\cdot 10^{-28}$ g, но сравнима с массой протона $m_p \approx 1.67\cdot 10^{-24}$ g. Ясно, что если протоны и электроны будут по-разному участвовать в гравитационном взаимодействии, это будет катастрофой. Можно, конечно, при достаточной доли фантазии пытаться использовать эти эффекты для решения проблемы первичных магнитных полей, но скорее это указывает на необходимость задуматься о границах применимости модели.

Другой предел, о котором следует задуматься, – это предельно слабые ускорения. Следует ли считать нелинейность и нарушение принципа суперпозиции применимыми при сколь угодно малых градиентах ньютоновского потенциала? И если да, то не будет ли это нелокальностью самого опасного типа, не позволяющего разделить Вселенную на части, чьим влиянием друг на друга можно пренебречь? Разумеется, принципу суперпозиции мешает не сам закон $\frac{1}{r}$, а пропорциональность силы величине $\sqrt{M}$ вместо M. Можно было бы ввести ещё одну фундаментальную величину размерности массы и на больших расстояниях иметь линейное взаимодействие с законом $\frac{1}{r}$, но такое дальнодействие, разумеется, заведомо привело бы к невозможности не учитывать эффектов сколь угодно далёкого распределения масс на локальные события.

* * *

Таким образом, мы видим, что прямое применение идей МОНД к космологии требует релятивистского обобщения модели, прийти к которому оказывается нетривиальной задачей, да к тому же не позволяет в рамках любых известных на сегодня методов правильно описать все имеющиеся данные. Наличие очень серьёзной нелинейности делает теорию даже в нерелятивистском пределе весьма проблематичной. Вместе с тем есть удивительный контраст этих сложностей с успехами в масштабах галактической динамики, а также с потрясающим совпадением фундаментального ускорения МОНД с естественным масштабом ускорения в расширяющейся Вселенной, $a_0 \sim cH_0$.

Однако и для Тёмной Материи есть замечательное визуальное подтверждение в виде Скопления Пули (Bullet Cluster [195]), хотя и не исключено, что его можно тоже описать в рамках МОНД [196], а величина скоростей столкновения может быть проблематична для стандартной космологии [197]. (Ныне также активно изучается другое скопление – Абель [198].)

Поразительная универсальность закона МОНД, безусловно, требует объяснения, либо в рамках действительно универсального закона гравитации, либо как некий нетривиальный коллективный эффект, приводящий к рождению порядка из хаоса. Сегодня есть попытки описать феноменологию МОНД на основе особого вида сверхтекучести [199].

С нашей точки зрения, парадигма МОНД не может претендовать на статус фундаментальной теории, но может стать неким стандартным лекалом для проверки модифицированных теорий гравитации, которые ставят перед собой целью избавиться от необходимости вводить Тёмную Материю, поскольку любой такой подход должен либо иметь эффективные тяжёлые частицы со всеми обычными проблемами стандартной космологии на малых масштабах, либо другим путём естественно описывать феноменологию галактик в ньютоновском пределе. При этом все наши рассуждения в данном разделе могут быть перенесены на соответствующие ситуации, включая необходимость самосогласованно принимать во внимание все аспекты проблемы тёмной материи в космологии.

5.4 О дополнительных потенциалах из квантовой механики со связями второго рода

В связи с моделями миров на бранах иногда высказываются идеи о том, что дополнительные квантовые потенциалы, возникающие от того, что теория квантуется на искривлённой поверхности, могут быть использованы при построении космологических моделей [168], включая возможное описание (части) тёмных секторов. При этом зачастую ссылаются на волновые уравнения для свободных частиц на поверхностях, или даже на нерелятивистскую квантовую механику систем со связями второго рода (частица на поверхности). В этом разделе мы хотим отметить, что процедуры такого квантования существенно неоднозначны, и поэтому подобные предложения не выглядят как способные решить какие-либо проблемы.

В нашей работе [17*] был представлен обзор разных методов квантования, показывающий, что абстрактная задача плохо определена. По иронии судьбы именно эта статья является одной из тех, на которые ссылаются авторы работы [168], предлагая применение квантового потенциала для космологии. Поэтому мы решили кратко рассмотреть здесь этот сюжет, который может показаться далёким по тематике от основной части Диссертации.

5.4.1 Квантование по Дираку

Разумеется, одним из мыслимых способов квантования является квантование по Дираку [200]. Если в теории есть $2N$ связей ϕ_a, $a = 1, 2, \ldots, 2N$, то они называются связями второго рода, если $\det\{\phi_a, \phi_b\} \neq 0$ даже в слабом смысле (на поверхности связей). Поскольку такие связи не образуют замкнутой алгебры относительно скобок Пуассона, их никак нельзя наложить на квантовом уровне, если скобки Пуассона заменены на коммутаторы. Возможное решение было предложено Дираком. Введём новые скобки – скобки Дирака

$$\{f,g\}_{\mathcal{D}} = \{f,g\} - \sum_{a=1}^{2N}\sum_{b=1}^{2N}\{f,\phi_a\}\Delta_{ab}\{\phi_b,g\}, \tag{5.33}$$

где Δ_{ab} матрица, обратная к $\{\phi_a,\phi_b\}$. Теперь $\{\phi_1,\phi_2\}_D = 0$, и можно провести квантование обычным способом, заменяя коммутаторами скобки Дирака. Отметим, что скобка Дирака вырожденна и поэтому не определяет симплектической структуры на исходном фазовом пространстве, но в некотором смысле она является результатом факторизации по нефизическим направлениям, см. ниже (5.39).

Например, свободное движение по сфере $\sum\limits_{i=1}^{n} x_i^2 = R^2$ в эвклидовом пространстве можно представить [201] как систему с одной парой связей второго рода ($\{\phi_1,\phi_2\} = 2\overrightarrow{x}^2 \neq 0$):

$$\phi_1 \equiv \sum_{i=1}^{n} x_i^2 - R^2 = 0, \tag{5.34}$$

$$\phi_2 \equiv \sum_{i=1}^{n} x_i p_i = 0 \tag{5.35}$$

и гамильтонианом $\frac{1}{2}p^2$, где p_i – канонические импульсы. Простое вычисление даёт скобки Дирака

$$\{x_i,x_j\}_{\mathcal{D}} = 0, \tag{5.36}$$

$$\{x_i,p_j\}_{\mathcal{D}} = \delta_{ij} - \frac{x_i x_j}{\overrightarrow{x}^2}, \tag{5.37}$$

$$\{p_i,p_j\}_{\mathcal{D}} = \frac{1}{\overrightarrow{x}^2}(p_i x_j - p_j x_i). \tag{5.38}$$

Нетрудно видеть, что полученной алгебре удовлетворяют обычные операторы координат, дающиеся умножением на переменные x_i и импульсы, из которых вычтено радиальное дифференцирование

$$-i\hbar\overrightarrow{\nabla} \longrightarrow -i\hbar\left(\overrightarrow{\nabla} - \frac{\overrightarrow{x}}{|\overrightarrow{x}|}\left(\frac{\overrightarrow{x}}{|\overrightarrow{x}|}\cdot\overrightarrow{\nabla}\right)\right) \equiv \hat{\overrightarrow{p}}. \tag{5.39}$$

Плохо, конечно, что импульсы $\hat{p}_i$ не самосопряжённые, но, пожертвовав правилом Лейбница, можно их сделать таковыми:

$$\hat{\tilde{p}}_i = \frac{1}{2}(\hat{p}_i + \hat{p}_i^\dagger) = \hat{p}_i + i\hbar\frac{n-1}{2} \cdot \frac{x_i}{\vec{x}^2}\hat{I}.$$

Первая связь (5.34) определяет тогда пространство физических состояний (имеющих дельта-функцию $\delta(x^2 - R^2)$), а вторая связь становится (5.35) тождеством $\hat{\phi}_2 = \sum_{i=1}^{n}(\hat{x}_i\hat{\tilde{p}}_i + (\hat{x}_i\hat{\tilde{p}}_i)^\dagger) \equiv 0$. Гамильтониан получается таким:

$$\hat{H}^{(\mathcal{D})} \equiv \frac{1}{2}\sum_{i=1}^{n}\hat{\tilde{p}}_i^2 = -\frac{\hbar^2}{2}\Delta_{LB} + \frac{\hbar^2(n-1)^2}{8R^2},$$

где Δ_{LB} - оператор Лапласа-Бельтрами на сфере, и он содержит типичный квантовый потенциал

$$V_q^{(\mathcal{D})} = \frac{\hbar^2(n-1)^2}{8R^2}.$$

Вообще говоря, этот результат можно было бы получить и чисто алгебраически безо всяких операторных реализаций [201].

Заметим, вместе с тем, что можно было бы отказаться от самосопряжённости (и следовательно "наблюдаемости") операторов импульса. В конце концов, они являются проекциями операторов из объемлющего пространства – нечто совсем эзотерическое для наблюдателя на сфере, а естественные операторы $i[\hat{p}_i, \hat{p}_j]$ (генераторы $SO(n)$ вращений) будут при этом самосопряжёнными. Тогда гамильтониан окажется просто оператором Бельтрами-Лапласа $-\frac{\hbar^2}{2}\Delta_{LB}$ на сфере, в соответствии с давней идеей Бориса Подольского [202] и без квантовых потенциалов.

Произвольная поверхность коразмерности 1

На самом деле, случай сферы оказывается несколько обманчивым в смысле однозначной определённости метода, поскольку у сферы есть весьма естественное уравнение, которое практически однозначно хочется использовать.

Рассмотрим поверхность $f(x) = 0$, требуя чтобы $\left|\overrightarrow{\nabla} f\right|\Big|_{f=0} \neq 0$, и проведём ту же процедуру, что и раньше. Связи принимают вид

$$\phi_1 \equiv f(x) = 0, \tag{5.40}$$

$$\phi_2 \equiv \sum_{i=1}^{n} (\partial_i f) p_i = 0, \tag{5.41}$$

а скобки Дирака оказываются равными

$$\{x_i, x_j\}_{\mathcal{D}} = 0, \tag{5.42}$$

$$\{x_i, p_j\}_{\mathcal{D}} = \delta_{ij} - \frac{(\partial_i f)(\partial_j f)}{\left(\overrightarrow{\nabla} f\right)^2}, \tag{5.43}$$

$$\{p_i, p_j\}_{\mathcal{D}} = \frac{1}{\left(\overrightarrow{\nabla} f\right)^2} \sum_{k=1}^{n} \left((\partial_j f)(\partial^2_{ik} f) - (\partial_i f)(\partial^2_{jk} f)\right) p_k. \tag{5.44}$$

Заметим, что в правой части последней скобки Дирака есть проблема упорядочения операторов.

Можно использовать явные операторы импульса по аналогии с предыдущим случаем: несамосопряжённые (но настоящие дифференцирования)

$$\hat{p}_i = -i\hbar \left(\frac{\partial}{\partial x_i} - \frac{(\partial_i f)}{\left|\overrightarrow{\nabla} f\right|} \sum_{j=1}^{n} \frac{(\partial_j f)}{\left|\overrightarrow{\nabla} f\right|} \frac{\partial}{\partial x_j} \right)$$

или самосопряжённые

$$\hat{\hat{p}}_i = \hat{p}_i + \frac{i\hbar}{2} \sum_{j=1}^{n} \left(\frac{\partial}{\partial x_j} \left(\frac{(\partial_i f)(\partial_j f)}{\left(\overrightarrow{\nabla} f\right)^2} \right) \right). \tag{5.45}$$

Проблема упорядочения операторов принимает тогда следующие возможные решения

$$[\hat{p}_i, \hat{p}_j] = \frac{i\hbar}{\left(\overrightarrow{\nabla} f\right)^2} \sum_{k=1}^{n} \left((\partial_j f)(\partial^2_{ik} f) - (\partial_i f)(\partial^2_{jk} f)\right) \hat{p}_k;$$

$$[\hat{\tilde{p}}_i, \hat{\tilde{p}}_j] = \frac{i\hbar}{2} \sum_{k=1}^{n} \left(\frac{(\partial_j f)(\partial^2_{ik} f) - (\partial_i f)(\partial^2_{jk} f)}{\left(\overrightarrow{\nabla} f\right)^2} \hat{\tilde{p}}_k + \hat{\tilde{p}}_k \frac{(\partial_j f)(\partial^2_{ik} f) - (\partial_i f)(\partial^2_{jk} f)}{\left(\overrightarrow{\nabla} f\right)^2} \right).$$

Вторая связь имеет вид тождества

$$\sum_{i=1}^{n} (\partial_i f) \hat{p}_i \equiv 0$$

или

$$\sum_{i=1}^{n} \left((\partial_i f)\hat{\tilde{p}}_i + \hat{\tilde{p}}_i (\partial_i f)\right) \equiv 0,$$

а первая связь даёт в качестве физического сектора $\Psi_{phys} = \psi(x) \cdot \delta(f(x))$.

Обозначив через $\tilde{\Delta}$ оператор Лапласа в объемлющем (эвклидовом) пространстве, а также введя вектор нормали $\overrightarrow{n} = \frac{\overrightarrow{\nabla} f}{\left|\overrightarrow{\nabla} f\right|}$, нетрудно вычислить гамильтониан. Для несамосопряжённых импульсов получается

$$\hat{H}^{(\mathcal{P})} \equiv \frac{1}{2} \sum_{i=1}^{n} \hat{p}_i^{\dagger} \hat{p}_i = -\frac{\hbar^2}{2} \left(\tilde{\Delta} - \left(\frac{\partial}{\partial \overrightarrow{n}} \right)^2 - \operatorname{div}(\overrightarrow{n}) \cdot \frac{\partial}{\partial \overrightarrow{n}} \right),$$

а для самосопряжённых импульсов гамильтониан

$$\hat{H}^{(\mathcal{D})} \equiv \frac{1}{2} \sum_{i=1}^{n} \hat{\tilde{p}}_i^2 = \hat{H}^{(\mathcal{P})} + V_q^{(\mathcal{D})}(x)$$

содержит квантовый потенциал

$$V_q^{(\mathcal{D})} = -\frac{\hbar^2}{8}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{\partial}{\partial x_j}\frac{(\partial_i f)(\partial_j f)}{\left(\overrightarrow{\bigtriangledown} f\right)^2}\right)^2 +$$
$$+\frac{\hbar^2}{4}\sum_{i=1}^{n}\left(\frac{\partial}{\partial x_i} - \sum_{k=1}^{n}\frac{(\partial_i f)(\partial_k f)}{\left(\overrightarrow{\bigtriangledown} f\right)^2}\frac{\partial}{\partial x_k}\right)\left(\sum_{j=1}^{n}\frac{\partial}{\partial x_j}\frac{(\partial_j f)(\partial_i f)}{\left(\overrightarrow{\bigtriangledown} f\right)^2}\right), \quad (5.46)$$

который можно также представить в следующем виде:

$$V_q^{(\mathcal{D})} = \frac{\hbar^2}{4}\left(\frac{1}{2}\left(\sum_i \partial_i n_i\right)^2 + \sum_{i,k}\partial_i\left(n_k\partial_k n_i\right) + \frac{1}{2}\sum_{i,k,m} n_i n_k n_m \partial^2_{km} n_i\right). \quad (5.47)$$

К сожалению, ни кинетическая ни потенциальная части не являются однозначно опеределёнными, но зависят от выбора функции f, причём даже для сфер. В самом деле, любую поверхность можно в данной точке приблизить касательным параболоидом

$$f(y) = y_n - \frac{1}{2}\sum_{\alpha=1}^{n-1} k_\alpha y_\alpha^2 + \mathcal{O}(y_\alpha^3)$$

(в декартовых координатах объемлющего пространства). Тогда из (5.46) получаем

$$V_q = \frac{\hbar^2}{8}\left(\left(\sum_{\alpha=1}^{n-1} k_\alpha\right)^2 + 2\sum_{\alpha=1}^{n-1} k_\alpha^2\right) + \mathcal{O}(y_\alpha)$$

в окрестности точки касания $\overrightarrow{y} = 0$. Для сферы все главные кривизны $k_\alpha = \frac{1}{R}$, и мы получаем

$$V_q = \frac{\hbar(n^2-1)}{8R^2},$$

что отличается от нашего предыдущего результата.

Метод можно доопределить, потребовав, чтобы функция f зависела только от расстояния до поверхности:

$$f = f(\text{distance from the surface}),$$

что, разумеется, соответствует обычному уравнению сферы. В этом случае можно убедиться, что квантовый потенциал

$$V_q^{(\mathcal{D})} = \frac{\hbar^2}{8} \left(\sum_{\alpha=1}^{n-1} k_\alpha \right)^2$$

воспроизводит предыдущий результат для сферы, а кинетическое слагаемое становится равным

$$\tilde{\Delta} - \left(\frac{\partial}{\partial \overrightarrow{n}} \right)^2 - \mathrm{div}(\overrightarrow{n}) \cdot \frac{\partial}{\partial \overrightarrow{n}} = \Delta_{LB}$$

оператору Лапласа-Бельтрами на поверхности. Это, конечно, вполне естественный рецепт, но очевидно, что абстрактная задача не определена корректно.

В принципе, все эти рассуждения можно обобщить на старшие коразмерности, но явные формулы становятся довольно громоздкими.

* * *

Из (всевдо)канонических методов можно также отметить абелеву конверсию связей, предложенную в работе [203] для случая аномалий в калибровочных симметриях, то есть для тех случаев, когда связи первого рода после квантования становятся связями второго рода, а калибровочная симметрия, соответственно, не выдерживает квантования. Очевидно, этот метод можно применить и к связям, которые исходно были второго рода. Идея состоит во введении новой пары канонических переменных, удовлетворяющих паре связей первого рода с абелевой алгеброй, то есть со скобкой Пуассона, сильно обращающейся в нуль. При этом, если жёстко положить новую пару канонических переменных равными нулю, то требуется, чтобы новые связи переходили в пару старых связей второго рода (или в связь и калибровочное условие). В работах [201, 204] таким образом был получен нулевой квантовый потенциал на сфере, но в более общем случае уравнения становятся весьма громоздкими и непрозрачными, см. нашу работу [17*].

5.4.2 Метод тонкого слоя

В методе тонкого слоя используется более физический подход к квантованию движения на поверхности. А именно, рассматривается частица, которая движется в бесконечно глубокой потенциальной яме со стенками, находящимися по обе стороны от поверхности на расстоянии $\delta \to 0$. Можно, разумеется, рассматривать и более сложные ограничивающие потенциалы. Метод в общем виде, по всей видимости, появился в работах [205, 206], см. также более современное обсуждение в статьях [207, 208], но в силу присущей ему естественности элементы его для частных задач можно проследить и в более ранних работах, например для теории химических реакций в работе [209]. Примеры использования подобного подхода можно найти и при описании движения электронов в наноструктурах [210, 211], и в физике молекул [207, 212], и даже в более общем случае квантовых графов [213]. Обсуждение математических аспектов можно найти в статьях [214, 215].

Рассмотрим $(n-1)$-мерную гиперповерхность в $\mathbb{R}^n$ с двумя бесконечно высокими потенциальными стенками на расстоянии $\delta \to 0$ по обе стороны от поверхности. Пусть свободная квантовая частица движется между этими стенками.

Введём криволинейную систему координат, такую, что $|x_n|$ равно расстоянию от поверхности, а остальные координаты задают положения на поверхностях $x_n = const$, так что метрика $\tilde{g}_{ik} = \begin{pmatrix} g_{ab} & 0 \\ 0 & 1 \end{pmatrix}$. Рассматриваем гамильтониан

$$\tilde{H} = -\frac{\hbar^2}{2}\tilde{\Delta}$$

с граничными условиями $\Psi|_{x_n=\delta} = \Psi|_{x_n=-\delta} = 0$, где оператор Лапласа равен

$$\tilde{\Delta} = \sum_{i=1}^{n}\sum_{k=1}^{n} \tilde{g}^{-1/2}\partial_i \tilde{g}^{1/2}\tilde{g}^{ik}\partial_k = \partial_n{}^2 + \left(\tilde{g}^{-1/2}\partial_n \tilde{g}^{1/2}\right)\partial_n + \Delta_{LB},$$

а Δ_{LB} – оператор Лапласа-Бельтрами на поверхности $x_n = const$.

Рассмотрим касательный параболоид к нашей поверхности

$$y_n = \frac{1}{2}\sum_{a=1}^{n-1} k_a y_a^2 + \mathcal{O}(y_a^3),$$

где k_a – её главные кривизны. Единичная нормаль равна

$$n_a = \frac{k_a y_a}{\sqrt{1 + \sum\limits_{a=1}^{n-1} k_a^2 y_a^2}} + \mathcal{O}(y_a^2) = k_a y_a + \mathcal{O}(y_a^2),$$

$n_n = -1 + \mathcal{O}(y_a^2)$, а близлежащая поверхность $x_n = \epsilon$ получается как $\overrightarrow{y} \longrightarrow \overrightarrow{y}' = \overrightarrow{y} + \epsilon\overrightarrow{n}$, и

$$dy'_a = dy_a\left(1 + \epsilon k_a + \mathcal{O}(y_a)\right).$$

Отсюда

$$\frac{dS'}{dS} = \frac{\prod\limits_{a=1}^{n-1}\left(1 + \mathcal{O}(y_a'^2)\right) dy'_a}{\prod\limits_{a=1}^{n-1}\left(1 + \mathcal{O}(y_a^2)\right) dy_a} = \prod_{a=1}^{n-1}(1 + \epsilon k_a) + \mathcal{O}(y_a).$$

В частности, на линии $y_a = 0 \quad \forall a = 1, \ldots, n-1$ получаем

$$\frac{dS'}{dS} = 1 + \epsilon\sum_{a=1}^{n-1} k_a + \frac{1}{2}\epsilon^2\left(\left(\sum_{a=1}^{n-1} k_a\right)^2 - \sum_{a=1}^{n-1} k_a^2\right) + \mathcal{O}(\epsilon^3). \qquad (5.48)$$

Конечно же, соотношение (5.48) верно в любой точке поверхности, но главные кривизны, вообще говоря, зависят от точки.

Следуя работам [205, 206], введём новую волновую функцию

$$\chi(x) = \Psi(x)\sqrt{\frac{dS'}{dS}},$$

физический смысл которой ясен из соотношения

$$\int\limits_{|x_n|\le\delta} dV|\Psi(x)|^2 = \int\limits_{-\delta}^{\delta} dx_n \int dS|\chi(x)|^2.$$

На низшем уровне энергии по отношению к нормальному движению $\chi(x_1,\dots,x_n) = f(x_1,\dots,x_{n-1})\cos\frac{\pi x_n}{2\delta}$ имеем

$$\tilde\Delta\Psi(x) = \Delta_{LB}\chi(x) + \partial_n^2\chi(x) + \left(\frac{1}{2}\sum_{a=1}^{n-1}k_a^2 - \frac{1}{4}\left(\sum_{a=1}^{n-1}k_a\right)^2\right)\chi(x) + \mathcal{O}(x_n).$$

Беря предел $\delta\to 0$ и вычитая бесконечную постоянную энергию (пропорциональную $1/\delta^2$), получаем гамильтониан

$$\hat H = -\frac{\hbar^2}{2}\Delta_{LB} + \frac{\hbar^2}{8}\left(\left(\sum_{a=1}^{n-1}k_a\right)^2 - 2\sum_{a=1}^{n-1}k_a^2\right) \tag{5.49}$$

с квантовым потенциалом

$$V_q = \frac{\hbar^2}{8}\left(\left(\sum_{a=1}^{n-1}k_a\right)^2 - 2\sum_{a=1}^{n-1}k_a^2\right).$$

Для двумерной поверхности в $\mathbb{R}^3$ воспроизводим тем самым результат да Косты [206], $V_q = -\frac{\hbar^2}{8}(k_1-k_2)^2$, а для сфер в произвольной размерности ($k_a = \frac{1}{R}$) потенциал равен $V_q = \frac{\hbar^2(n-1)(n-3)}{8R^2}$. Таким образом, метод тонкого слоя приводит к совершенно другим квантовым потенциалам в сравнении с методом Дирака.

Разумеется, если бы толщина тонкого слоя менялась от точки к точке, получились бы дополнительные эффективные силы [207].

О старших коразмерностях

Отметим, что ситуация несколько усложняется в старших коразмерностях. Рассмотрим для начала случай кривой (одномерная поверхность). Задав кривую в виде $y_2 = \frac{1}{2}ky_1^2 + \mathcal{O}(y_1^3)$; $y_3,\dots,y_n = \mathcal{O}(y_1^3)$, легко получить ответ с квантовым потенциалом $V_q = -\frac{\hbar^2}{8}k^2$, где k – внешняя кривизна кривой [206]. Однако, легко понять, что если кривая обладает кручением (в смысле внешней геометрии кривых), то репер Френе вращается вокруг кривой, и если мы хотим получить описание в невращающейся системе координат, то появятся слагаемые в гамильтониане, связанные с вращением вокруг кривой [216]. Если поперечное сечение тонкого слоя инвариантно относительно вращений вокруг кривой, то локально картина будет эквивалентной, но могут появляться глобальные фазы типа фазы Берри.

В случае поверхностей, у которых и размерность и коразмерность больше единицы, подобное описание вообще не всегда возможно. Препятствием является кривизна нормального расслоения. Это проявляется в том, что не удаётся избавиться в гамильтониане от смешанных (нормальных и тангенциальных) вторых производных [217]. Однако можно проверить, что они собираются в операторы, описывающие вращение в поперечных сечениях тонкого слоя [207, 208, 218, 219], как и в случае кривой. Если по каким-либо причинам рассматриваемый энергетический уровень поперечного движения вырожден, это может приводить к появлению калибровочных структур [207, 208, 218–221].

* * *

Таким образом, абстрактная задача квантования систем со связями вряд ли может служить средством введения и обоснования тех или иных квантовых потенциалов для космологии на бране. Здесь проявляется общая неоднозначность квантования, не позволяющая, конечно же, восстановить теорию по ее $\hbar \to 0$ пределу. Мы привыкли к гораздо большей определённости в квантовой теории поля, где для придания разумного смысла всем величинам приходится действовать вполне определённым образом, в частности переходя к нормальному упорядочению операторов. Возможно, обсуждение на языке эффекта Казимира было бы более осмысленным [222].

Но, строго говоря, для обсуждения подобных вопросов может потребоваться лучшее понимание квантовой гравитации, о которой мы пока знаем мало, и которая может сильно изменить наше понимание основ квантовой теории. Как показал опыт квантования бозонной струны методами петлевой квантовой гравитации [223, 224], представление о том, что теория в слабой связи должна допускать описание на языке (квази)частиц и пространства Фока, – это уже довольно сильный принцип, отказ от которого может радикально изменить общую форму теории.

5.5 Замечания о Чёрных Дырах и информационном парадоксе

По всей видимости, у нас слишком мало данных для обсуждения проблем квантовой гравитации за пределами чисто абстрактной постановки задачи. Однако есть один класс систем, которые позволяют получать интересные результаты на основе самых общих принципов квантовой механики и теории гравитации. Это Чёрные Дыры и процессы их квантового испарения.

Применение квантовой механики к Чёрным Дырам сразу же привело к удивительному заключению о том, что Чёрные Дыры должны излучать [225]. Излучение получается с хорошей точностью тепловым, и предполагая, что в конечном счёте Чёрная Дыра испарится полностью, приходим к парадоксальному выводу о том, что практически вся информация, проглоченная Чёрной Дырой за время её жизни (и при образовании тоже) полностью исчезла вопреки унитарному характеру квантовомеханической эволюции.

Стандартным решением стало предположение о существовании растянутого горизонта (stretched horizon) планковской толщины, который может поглощать, термализовывать и переизлучать информацию [226]. Разумеется, в силу принципа эквивалентности, падающий наблюдатель не должен вообще ничего особенного заметить при пересечении горизонта, поскольку кривизна горизонта Чёрной Дыры макроскопической массы очень мала. Однако, обнаружив сей замечательный факт отсутствия физической мембраны, падающий наблюдатель уже не будет иметь шанса передать эту информацию удалённому коллеге, и таким образом мы можем избежать противоречия, приняв точку зрения принципа дополнительности [226] и отказавшись от глобального описания всего пространства-времени в рамках одной эффективной теории привычного нам вида.

Нетрудно установить характерные для процесса испарения масштабы времён. В самом деле, Чёрная Дыра массы M имеет температуру Хокинга

$$\mathrm{k}_B T = \frac{\hbar c^3}{8\pi G M},$$

зависящую от значения постоянной Планка $\hbar \equiv \frac{h}{2\pi}$, скорости света c, гравитационной постоянной G, и в температурных единицах также, конечно, содержащую постоянную Больцмана k_B. По закону Штефана-Больцмана поверхностная яркость определяется величиной $T^4 \propto \frac{1}{M^4}$, и при поверхности горизонта $\propto M^2$ имеем $\dot{M} \propto \frac{1}{M^2}$ для скорости потерь энергии. Соответственно, характерное время определяется кубом массы M^3, или явно вводя планковские единицы: $t \sim \left(\frac{M}{m_{Pl}}\right)^3 t_{Pl}$.

Отмеченная недавно проблема с принципом комплементарности (дополнительности) проявляется [227] (см. также [228, 229]), когда Чёрная Дыра излучает половину своей энтропии. Возраст Чёрной Дыры на этот момент времени мы, по сути, оценили выше

$$t_{\mathrm{P}} \sim \left(\frac{M}{m_{Pl}}\right)^3 \cdot t_{Pl},$$

он называется временем Пейджа [230]. Можно показать из общих принципов [231], что на этот момент практически вся информация, которая была поглочена Чёрной Дырой, должна содержаться в излучении, покинувшем Чёрную Дыру. В таком случае, поскольку новой информации взяться неоткуда, всё излучение, уже удалившееся от Чёрной Дыры на значительное расстояние (раннее излучение) должно быть в максимально квантовомеханически запутанном состоянии с тем излучением (позднее излучение), которое сейчас покидает Чёрную Дыру и находится снаружи горизонта, но в непосредственной близости от него ("зона").

С другой стороны, для падающего наблюдателя окрестности горизонта весьма мало отличаются от пространства Минковского, и в силу принципа эквивалентности, он должен наблюдать обычный вакуум в пространстве Минковского. Но вакуумное состояние – весьма упорядоченно, и это проявляется в том, что две половины почти пространства Минковского (внутренность Чёрной Дыры и зона) должны быть максимально запутанны друг с другом. Однако же невозможно для одной системы (зоны) быть максимально запутанной с двумя различными системами (моногамность запутывания). Таким образом, мы получили противоречие. Это и есть парадокс АМПС [227]. Нашими предположениями были: принцип эквивалентности в общей теории относительности (пересечение горизонта как пустого пространства), постулаты квантовой механики для вселенной как целого, включая Чёрную Дыру (свойства запутывания и унитарность эволюции), приближённая справедливость локальной квантовой теории поля в областях малой кривизны (свойства вакуума).

Проблема заведомо возникает для старой Чёрной Дыры (начиная с пейджовского возраста). Отметим, что также есть аргументы [227] в пользу того, что проблемы начинаются раньше – начиная со времени, за которое Чёрная Дыра успевает перерабатывать поступающую информацию в хаотическое состояние (забывание начальных условий при огрублении; тут есть некоторые тонкие отличия от времени термализации). Существуют оценки, показывающие, что это время "переваривания" (scrambling time) весьма мало [232],

$$t_{scr} \sim \left(\frac{M}{m_{Pl}} \log \frac{M}{m_{Pl}} \right) \cdot t_{Pl},$$

но реальный статус этого временного масштаба гораздо менее ясен [233].

Однако, неважно, начиная с какого времени, но проблема заведомо возникает, и решением, предложенным в работе [227], была "стена огня" (firewall) сразу (планковская глубина) под горизонтом. Таким образом, в жертву приносится принцип эквивалентности. Максимальное запутывание между ранним и поздним излучением приводит к тому, что две стороны горизонта оказываются в незапутанном состоянии, и это превращается в утверждение о том, что в окрестности горизонта существуют реальные кванты всех возможных энергий (firewall).

Отметим, что другая сразу приходящая в голову возможность – полностью отождествить внутренность Чёрной Дыры и раннее излучение, объявив их одной и той же системой, – не только чрезвычайно нелокальна, но и приводит к проблеме замороженного вакуума [234]: падающий наблюдатель, при всём своем желании, не может возбудить квантополевые состояния около горизонта, что нарушает принцип эквивалентности ни чуть не меньше, чем полная противоположность, стена огня.

За последние несколько лет на эту тему написано огромное множество работ с разными точками зрения и предложениями. В частности, можно пытаться встать на точку зрения сильной дополнительности, утверждающей, что это неважно, если два наблюдателя наблюдают абсолютно разную физическую реальность в отношении одной и той же физической системы, если у них нет даже принципиальной возможности сообщить о своих результатах друг другу. Похоже, этот вариант не проходит (по крайней мере, не без проблем), поскольку падающий наблюдатель может провести очень точные измерения раннего излучения перед непосредственным попаданием в зону (для выяснения вопроса о запутанности), или оттого, что он может поменять решение и повернуть назад, уже находясь в зоне, но перед пересечением горизонта [235]. Однако, можно возразить, что проблемные измерения и выводы оказываются нереальными с точки зрения вычислительной сложности [236], требовать экстремальной точности сродни демону Максвелла [237], полностью менять состояние самой Чёрной Дыры, или практически соответствовать наблюдению квантовых суперпозиций макроскопических систем [238].

Вопрос остаётся спорным, и новые идеи могут радикально изменить наши представления о квантовой механике и гравитации. Так, одна из возможностей заключается в том [239], что именно сам факт попытки измерения со стороны удалённого наблюдателя и создаёт ту стену огня, что убивает его путешествующего коллегу. Предлагаемая интерпретация [240] связывает квантовую запутанность с существованием моста Эйнштейна-Розена, и выражается ярким слоганом $EPR = ER$.

* * *

Нам представляется, что возникшая проблема может скорее указывать на ограниченную применимость методов локальной квантовой теории поля, в том смысле, что по ходу распространения хокинговского изучения от Чёрной Дыры неизбежно должно возникать запутывание его квантового состояния со степенями свободы квантовой пространственно-временной пены, так что позднее излучение само по себе уже нельзя будет считать чистой подсистемой, а квантовогравитационные флуктуации могут обладать и свойствами нелокальности.

Основная идея заключается в том, что эффективное квантовополевое описание может быть достаточно аккуратным во всех областях не очень большой кривизны, включая и горизонт и области вне его, но при этом ошибки могут накопиться и стать решающими при попытке применять это описание на протяжении огромных пространственных расстояний и времён. Здесь может быть проведена аналогия с отличием Григорианского и Юлианского календарей, в которых продолжительность года разная, но отличие меньше одиннадцати минут, что может поначалу показаться не имеющим никакого значения, но за четыреста лет накапливается ошибка в три полных дня. Случайные флуктуации растут, конечно, медленнее, чем направленный тренд, но в конце концов и они накапливаются.

Мы хотим привести некоторые оценки, показывающие, как в рамках этих идей могут возникать характерные для задачи времена.

Рассмотрим сперва очень простую постановку задачи. Примем во внимание эффекты квантовой гравитации, предположив, что типичная длина волны фотона хокинговского излучения

$$\lambda \sim \frac{M}{m_{Pl}} \cdot l_{Pl}$$

имеет неопределённость порядка планковской длины, которую будем рассматривать как неизбежную минимальную флуктуацию

$$\delta\lambda \sim l_{Pl}.$$

При распространении на много (N) длин волн оценим статистическую неопределённость длины пройденного пути $L = N\lambda$ как

$$\delta L \sim \sqrt{N} \cdot l_{Pl},$$

откуда видим, что неопределённость пути δL достигает полной длины волны λ на расстоянии

$$L \sim \frac{\lambda^3}{l_{Pl}^2} \sim \left(\frac{M}{m_{Pl}}\right)^3 \cdot l_{Pl} \sim ct_{\mathrm{P}}.$$

Тем самым в задаче возникает характерный масштаб пейджевского времени $t_{\mathrm{P}} \propto M^3$.

К моменту времени t_{P} информация об относительных фазах фотонов полностью потеряна за счёт того, что мы не можем рассматривать пространство-время в виде лишь бессловесной арены для разыгрывания квантовополевых драм. Квантовое состояние фотонов запутывается с квантовыми флуктуациями геометрии, усреднение по которым делает размытым тот образ, что несли с собой эти фотоны, и всей информации уже нельзя восстановить без обращения к (вполне возможно, нелокальным) состояниям квантовой геометрии. Таким образом, при рассмотрении огромных пространственных и временных масштабов, чрезвычайно малые эффекты квантовой гравитации могут становиться весьма существенными.

Отметим, что есть наблюдательные возражения против моделирования пространственно-временной пены на основе случайного блуждания [241], но они, с нашей точки зрения, несколько сомнительны. К тому же, мы не ставим своей целью построить феноменологически приемлемую модель для квантовогравитационной пены, но скорее показываем, как характерные временные масштабы в принципе могут возникнуть.

Время переваривания информации $t_{scr} \propto M \log M$ получить намного сложнее, да оно и само имеет более спекулятивный характер. Тем не менее, не стремясь к большей точности, чем заслуживает само понятие времени t_{scr}, попробуем привести аргументы на основе более тонкого рассмотрения гравитационно индуцированной декогеренции. Естественно полагать, что малые квантовогравитационные поправки могут описываться заменой чистого пучка фотонов на открытую квантовую систему, описываемую уравнением Линдблада [242, 243]

$$\dot{\rho} = \hat{\mathfrak{L}}\rho,$$

где ρ – матрица плотности, а генератор динамической полугруппы состоит из двух частей: коммутатора с гамильтонианом (стандартная шрёдингеровская эволюция) и оператора Линдблада, моделирующего открытость квантовой системы (взаимодействие с внешними степенями свободы).

Отметим, что такой подход уже обсуждался ранее в применении к моделированию эффектов квантовой гравитации, например в работе [244] указаны возможные эффекты в осцилляциях нейтральных каонов.

Порядок величины коэффициентов в матрице линбладовского генератора зависит от принятого уровня огрубления (coarse graining). Нам интересно продвинуть эффективную теорию поля до самых дальних мыслимых границ её применимости, поэтому естественно считать, что обрезание (cutoff) находится на планковском масштабе. В соответствии с этим будет считать, что за счёт планковского (квантовогравитационного) шума соседние огрублённые квантовые состояния фотонов начинают перепутываться за характерное время порядка $\tau \sim \frac{\lambda}{c}$, а характерный масштаб коэффициентов в уравнении Линдблада естественно оценивать как $\frac{c}{\lambda} \sim \frac{m_{Pl}}{M t_{Pl}}$.

Заметим, что мы заинтересованы не в простом превращении одного состояния в другое посредством внедиагональных элементов оператора Линдблада, поскольку само по себе это ещё не означает декогеренции, хотя и может производить интересные эффекты, такие как CPTнарушение [244]. Скорее мы хотим видеть независимый рост вероятностей других состояний, превращающий почти чистое состояние в статистическую смесь. В качестве начальных данных естественно считать, что данный фотон был в данном чистом состоянии с вероятностью p_1, практически равной 1. Но, благодаря хотя бы квантовым эффектам, другие состояния не могли полностью отсутствовать. Их начальные вероятности $p_j(0)$ можно оценить как $\sim \frac{m_{Pl}}{M}$, или какой-нибудь разумной степенью этой дроби. В соответствии с нашими предположениями, ясно, что величины $\log p_j$ будут типично расти как $\frac{m_{Pl}}{Mt_{Pl}} \cdot t$. Смело экстраполируя это соотношение за всякие мыслимые пределы, приходим к выводу, что безусловно смешанное состояние формируется, когда

$$\log \frac{M}{m_{Pl}} \sim \frac{m_{Pl}}{Mt_{Pl}} \cdot t,$$

или ко времени переваривания.

* * *

Конечно, всегда было ясно, что так или иначе отменив предположение о запутывании между поздним и ранним хокинговским излучением, парадокс заведомо можно разрешить [245]. Более того, есть конкретная реализация в терминах многомировой интерпретации квантовой механики [246]. Дело в том, что Чёрные Дыры случайно испускают огромное количество очень мягких фотонов, и можно оценить, что для старой Чёрной Дыры результирующая неопределённость положения за счёт эффектов отдачи чудовищна [230]. Соответственно, настоящее чистое состояние представляет собой суперпозицию Чёрной Дыры в совершенно различных положениях и с совершенно разным составом поглощённого или непоглощённнного вещества. (Интересно было бы, кстати, сравнить эти суперпозиции с таковыми из работы [238].) Безусловно, есть возможность того, что при выборе конкретной ветви многомирового состояния унитарность будет нарушена, несмотря на справедливость её в полной картине многих миров. Это довольно радикальная идея, и следует по-думать, не может ли она привести к наблюдаемости макроскопических суперпозиций с помощью Чёрных Дыр.

Наше предложение другое. Случайная квантовая среда (флуктуирующая геометрия) – это, конечно, не просто излучённые гравитоны, и предполагает более глубокие аспекты квантовой гравитации, но при этом совершенно независима, по крайней мере, от сложных проблем из оснований квантовой механики. Конечно, для более глубокого описания этих процессов необходимо лучшее понимание квантовой теории гравитации, которой у нас (пока) нет. Однако те квантовогравитационные эффекты, включая возможную нелокальность, которые мы рассматриваем здесь, не вынуждают нас радикально менять физику в областях малой кривизны, оставаясь, по сути, эффектами планковской физики.

Мы далеки, конечно, от окончательных ответов. Но не исключено, что физика Чёрных Дыр – это единственная на сегодняшний день область, в которой мы можем обсуждать проблемы квантовой гравитации за пределами чисто абстрактных и математических формулировок. Чёрные Дыры – замечательные объекты теории (и, по всей видимости, отнюдь не только теории), которые не перестают нас удивлять с тех самых пор как были открыты законы их термодинамики [247], весьма неожиданное событие для столь простых и, на первый взгляд, бесструктурных объектов. Возможно, они пытаются рассказать нам что-то о более фундаментальной физике.

Заключение

Подводя итоги, в данной Диссертации представлено исследование целого ряда моделей модифицированной гравитации, которые представляют несомненный интерес как с точки зрения фундаментальных задач описания гравитационного взаимодействия, так и с точки зрения проблем современной космологии. Мы надеемся, что проделанная работа станет весомым вкладом на пути к лучшему пониманию наиболее фундаментальных законов, управляющих развитием Вселенной. Конкретные результаты, можно сформулировать в следующем виде.

1. Во второй главе построены модели векторной инфляции, изучены проблемы их неустойчивости, проанализированы возможные нарушения гиперболичности в теориях с неканоническими векторными полями.

2. В третьей главе рассмотрены избранные вопросы теорий массивной гравитации де Рам - Габададзе - Толли. Предложен новый способ доказательства отсутствия духа Боулвара-Дезера. Подробно описана проблема извлечения квадратных корней из матриц. Дана новая формулировка теории возмущений, с более широкой областью применимости, чем стандартная. Обнаружено наличие духа Боулвара-Дезера в массивной гравитации с расширенным квазидилатоном.

3. В четвёртой главе разработан гамильтонов формализм для анализа биметрических теорий со связностью, порождаемой вспомогательной метрикой, а также предложено описание моделей Амендолы - Энквиста - Койвисто на языке любой из двух метрик, включая необходимость доопределения, а также потенциальную связь с нелокальными теориями гравитации.

4. В пятой главе рассмотрено несколько различных вопросов. Наиболее важными являются проблемы локальной лоренц-инвариантности в телепараллельных теориях гравитации и скалярно-тензорная формулировка миметической гравитации. В других разделах обсуждаются идеи МОНД, квантование систем со связями второго рода, а также информационный парадокс в физике чёрных дыр.

Более подробно основные результаты сформулированы во Введении в качестве положений, выносимых на защиту.

Список основных публикаций автора по теме диссертации

1. A. Golovnev, V. Mukhanov, V. Vanchurin. Vector inflation. Journal of Cosmology and Astroparticle Physics, JCAP06(2008)009 (2008).

2. A. Golovnev, V. Mukhanov, V. Vanchurin. Gravitational waves in vector inflation. Journal of Cosmology and Astroparticle Physics, JCAP11(2008)018 (2008).

3. A. Golovnev, V. Vanchurin. Cosmological perturbations from vector inflation. Physical Review D 79, 103524 (2009).

4. A. Golovnev. Linear perturbations in vector inflation and stability issues. Physical Review D 81, 023514 (2010).

5. A. Golovnev. On cosmic inflation in vector field theories. Classical and Quantum Gravity, 28, 245018 (2011).

6. A. Golovnev, A. Klementev. On hyperbolicity violations in cosmological models with vector fields. Journal of Cosmology and Astroparticle Physics, JCAP02(2014)033 (2014).

7. A. Golovnev. On the Hamiltonian analysis of non-linear massive gravity. Physics Letters B 707, 404-408 (2012).

8. A. Golovnev, F. Smirnov. Dealing with ghost-free massive gravity without explicit square roots of matrices. Physics Letters B 707, 209 (2017).

9. A. Golovnev, F. Smirnov. Unusual square roots in the ghost-free theory of massive gravity. Journal of High Energy Physics, JHEP06(2017)130 (2017).

10. A. Golovnev, A. Trukhin. Ghosts in extended quasidilaton theories. Physical Review D 96, 104032 (2017).

11. J. Beltran Jimenez, A. Golovnev, M. Karciauskas, T. Koivisto. The bimetric variational principle for General Relativity. Physical Review D 86, 084024 (2012).

12. A. Golovnev, M. Karciauskas, H.J. Nyrhinen. ADM Analysis of Bimetric Variational Principle. Journal of Cosmology and Astroparticle Physics, JCAP05(2015)021 (2015).

13. A. Golovnev, T. Koivisto, M. Sandstad. Effectively nonlocal metric-affine gravity. Physical Review D 93, 064081 (2016).

14. A. Golovnev, T. Koivisto, M. Sandstad. On the covariance of teleparallel gravity theories. Classical and Quantum Gravity 34, 145013 (2017).

15. A. Golovnev. On the recently proposed Mimetic Dark Matter. Physics Letters B 728, 39 (2014).

16. A. Golovnev, N. Masalaeva. Modified gravitational collapse, or the wonders of the MOND. General Relativity and Gravitation 46, 1754 (2014).

17. A. Golovnev. Canonical quantization of motion on submanifolds. Reports on Mathematical Physics 64, 59-77 (2009).

18. A. Golovnev. Smooth horizons and quantum ripples. European Physical Journal C 75, 185 (2015).

Литература

[1] Ю.Г. Борисович, Н.М. Близняков, Я.А. Израилевич, Т.Н. Фоменко. Введение в топологию. Москва, "Наука" 1995.

[2] Ш. Кобаяси, К. Номидзу. Основы дифференциальной геометрии. В 2 томах. Новокузнецк, НФМИ 1999.

[3] Ч. Мизнер, К. Торн, Дж. Уилер. Гравитация. В 3 томах. Москва, "Мир" 1977.

[4] R. Arnowitt, S. Deser, C.W. Misner, глава в сборнике *Gravitation: an introduction to current research*, под ред. L. Witten (Wiley 1962), стр. 227; также доступно в архиве препринтов arxiv.org/gr-qc/0405109

[5] Д.С. Горбунов, В.А. Рубаков. Введение в теорию ранней Вселенной: Теория горячего Большого взрыва. Москва, "ЛЕНАРД" 2016.

[6] Д.С. Горбунов, В.А. Рубаков. Введение в теорию ранней Вселенной: Космологические возмущения. Инфляционная теория. Москва, URSS 2010.

[7] W. Hu, J. Silk. Thermalization constraints and spectral distortions for massive unstable relic particles. Physical Review Letters 70 (1993), 2661.

[8] J. Chluba, Y. Ali-Haïmoud. COSMOSPEC: fast and detailed computation of the cosmological recombination radiation from hydrogen and helium. Monthly Notices of the Royal Astronomical Society 456 (2016), 3494. См. также препринт arxiv.org/abs/1510.03877

[9] T.P. Walker, G. Steigman, H. Kang, D.M. Schramm, K.A. Olive. Primordial nucleosynthesis redux. Astrophysical Journal 376 (1991), 51.

[10] A.J. Korn, F. Grundahl, O. Richard, P.S. Barklem, L. Mashonkina, R. Collet, N. Piskunov, B. Gustafsson. A probable stellar solution to the cosmological lithium discrepancy. Nature 442 (2006), 657. См. также препринт arxiv.org/abs/astro-ph/0608201

[11] K. Kajantie, M. Laine, K. Rummukainen, M. Shaposhnikov. The electroweak phase transition: a non-perturbative analysis. Nuclear Physics B 466 (1996), 189. См. также препринт arxiv.org/abs/hep-lat/9510020

[12] Z. Fodor, S.D. Katz. Critical point of QCD at finite T and μ, lattice results for physical quark masses. Journal of High Energy Physics 2004, JHEP04(2004)050. См. также препринт arxiv.org/abs/hep-lat/0402006

[13] M. Fukugita, T. Yanagida. Barygenesis without grand unification. Physics Letters B 174 (1986), 45.

[14] I. Affleck, M. Dine. A new mechanism for baryogenesis. Nuclear Physics B 249 (1985), 361.

[15] L. Hui, J.P. Ostriker, S. Tremaine, E. Witten. Ultralight scalars as cosmological dark matter. Physical Review D 95 (2017), 043541. См. также препринт arxiv.org/abs/1610.08297

[16] E.R. Harrison. Fluctuations at the Threshold of Classical Cosmology. Physical Review D 1 (1970), 2726.

[17] Ya.B. Zeldovich. A hypothesis, unifying the structure and the entropy of the Universe. Monthly Notices of the Royal Astronomical Society 160 (1972), 1P.

[18] V.F. Mukhanov, H.A. Feldman, R.H. Brandenberger. Theory of cosmological perturbations. Physics Reports 215 (1992), 203.

[19] R.K. Sachs, A.M. Wolfe. Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. Astrophysical Journal 147 (1967), 73.

[20] U. Seljak, M. Zaldarriaga. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. Astrophysical Journal 469 (1996), 437. См. также препринт astro-ph/9603033

[21] A.G. Riess, L. Macri, S. Casertano, H. Lampeitl, H.C. Ferguson, A.V. Filippenko, S.W. Jha, W. Li, R. Chornock. A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3. The Astrophysical Journal 730 (2011), 119. См. также препринт arxiv.org/abs/1103.2976

[22] A.G. Riess, L.M. Macri, S.L. Hoffmann, D. Scolnic, S. Casertano, A.V. Filippenko, B.E. Tucker, M.J. Reid, D.O. Jones, J.M. Silverman, R. Chornock, P. Challis, W. Yuan, P.J. Brown, R.J. Foley. A 2.4% Determination of the Local Value of the Hubble Constant. The Astrophysical Journal 826 (2016), 56. См. также препринт arxiv.org/abs/1604.01424

[23] Planck Collaboration. Planck 2015 results XIII. Cosmological parameters. Astronomy & Astrophysics 594 (2016), A13. См. также препринт arxiv.org/abs/1502.01589

[24] D.J. Schwarz, C.J. Copi, D. Huterer, G.D. Starkman. CMB anomalies after Planck. Classical and Quantum Gravity 33 (2016), 184001. См. также препринт arxiv.org/abs/1510.07929

[25] Planck Collaboration. Planck 2015 results XI. CMB power spectra, likelihoods, and robustness of parameters. Astronomy & Astrophysics 594 (2016), A11. См. также препринт arxiv.org/abs/1507.02704

[26] D. Spergel, R. Flauger, R. Hložek. Planck Data Reconsidered. Physical Review D 91 (2015), 023518. См. также препринт arxiv.org/abs/1312.3313

[27] G.E. Addison, Y. Huang, D.J. Watts, C.L. Bennett, M. Halpern, G. Hinshaw, J.L. Weiland. Quantifying discordance in the 2015 Planck CMB spectrum. The Astrophysical Journal 818 (2016), 132. См. также препринт arxiv.org/abs/1511.00055

[28] The COrE Collaboration. COrE (Cosmic Origins Explorer): A White Paper. arxiv.org/abs/1102.2181

[29] The Euclid Study Team. Euclid Definition Study Report ESA/SRE(2011)12. arxiv.org/abs/1110.3193

[30] R. Mandelbaum, P. McDonald, U. Seljak, R. Cen. Precision Cosmology from the Lyman-α Forest: Power Spectrum and Bispectrum. Monthly Notices of the Royal Astronomical Society 344 (2003), 776. См. также препринт arxiv.org/abs/astro-ph/0302112

[31] E. Armengaud, N. Palanque-Delabrouille, C. Yèche, D.J.E. Marsh, J. Baur. Constraining the mass of light bosonic dark matter using SDSS Lyman-α forest. Monthly Notices of the Royal Astronomical Society 471 (2017), 4606. См. также препринт arxiv.org/abs/1703.09126

[32] N. Palanque-Delabrouille, C. Yèche, J. Baur, C. Magneville, G. Rossi, J. Lesgourgues, A. Borde, E. Burtin, J.-M. LeGoff, J. Rich, M. Viel, D. Weinberg. Neutrino masses and cosmology with Lyman-alpha forest power spectrum. Journal of Cosmology and Astroparticle Physics 2015, JCAP11(2015)011. См. также препринт arxiv.org/abs/1506.05976.

[33] Planck Collaboration. Planck 2015 results XVII. Constraints on primordial non-Gaussianity. Astronomy & Astrophysics 594 (2016), A17. См. также препринт arxiv.org/abs/1502.01592.

[34] J. Chluba. Which spectral distortions does ΛCDM actually predict? Monthly Notices of the Royal Astronomical Society 460 (2016), 227. См. также препринт arxiv.org/abs/1603.02496.

[35] J.P. Gardner, J.C. Mather, M. Clampin, R. Doyon, M.A. Greenhouse, H.B. Hammel, J.B. Hutchings, P. Jakobsen, S.J. Lilly, K.S. Long,

J.I. Lunine, M.J. McCaughrean, M. Mountain, J. Nella, G.H. Rieke, M.J. Rieke, H.-W. Rix, E.P. Smith, G. Sonneborn, M. Stiavelli, H.S. Stockman, R.A. Windhorst, G.S. Wright. The James Webb Space Telescope. Space Science Reviews 123 (2006), 485. См. также препринт arxiv.org/abs/astro-ph/0606175

[36] R. Maartens, F.B. Abdalla, M. Jarvis, M.G. Santos. Cosmology with the SKA – overview. Proceedings of Science 2014, PoS(AASKA14)016. См. также arxiv.org/abs/1501.04076

[37] J.R. Pritchard, A. Loeb. 21-cm cosmology. Reports on Progress in Physics 75 (2012), 086901. См. также препринт arxiv.org/abs/1109.6012

[38] The LIGO Scientific Collaboration, the Virgo Collaboration. Observation of Gravitational Waves from a Binary Black Hole Merger. Physical Review Letters 116 (2016), 061102. См. также препринт arxiv.org/abs/1602.03837

[39] The LIGO Scientific Collaboration, the Virgo Collaboration. GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. Physical Review Letters 116 (2016), 241103. См. также препринт arxiv.org/abs/1606.04855.

[40] The LIGO Scientific Collaboration, the Virgo Collaboration. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. Physical Review Letters 118 (2017), 221101. См. также препринт arxiv.org/abs/1706.01812

[41] The LIGO Scientific Collaboration, the Virgo Collaboration. GW170814: A three-detector observation of gravitational waves from a binary black hole coalescence. Physical Review Letters 119 (2017), 141101. См. также препринт https://arxiv.org/abs/1709.09660

[42] The LIGO Scientific Collaboration, the Virgo Collaboration. GW170817: Observation of Gravitational Waves from a Binary Neutron

Star Inspiral GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. Physical Review Letters 119 (2017), 161101. См. также препринт arxiv.org/abs/1710.05832

[43] J.J.M. Carrasco, M.P. Hertzberg, L. Senatore. The Effective Field Theory of Cosmological Large Scale Structures. Journal of High Energy Physics 2012, JHEP09(2012)082. См. также препринт arxiv.org/abs/1206.2926

[44] J.B. Mertens, J.T. Giblin Jr, G.D. Starkman. Integration of inhomogeneous cosmological spacetimes in the BSSN formalism. Physical Review D 93 (2016), 124059. См. также препринт arxiv.org/abs/1511.01106

[45] J.S. Bullock, M. Boylan-Kolchin. Small-Scale Challenges to the ΛCDM Paradigm. Annual Review of Astronomy and Astrophysics 55 (2017), 343. См. также препринт arxiv.org/abs/1707.04256

[46] T. Sawala, C.S. Frenk, A. Fattahi, J.F. Navarro, R.G. Bower, R.A. Crain, C. Dalla Vecchia, M. Furlong, J.C. Helly, A. Jenkins, K.A. Oman, M. Schaller, J. Schaye, T. Theuns, J. Trayford, S.D.M. White. The APOSTLE simulations: solutions to the Local Group's cosmic puzzles. Monthly Notices of the Royal Astronomical Society 457 (2016), 1931. См. также препринт arxiv.org/abs/1511.01098

[47] M. Milgrom. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. The Astrophysical Journal 270 (1983), 365.

[48] J.L. Anderson, D. Finkelstein. Cosmological Constant and Fundamental Length. American Journal of Physics 39 (1971), 901.

[49] M. Henneaux, C. Teitelboim. The cosmological constant and general covariance. Physics Letters B 222 (1989), 195.

[50] S. Weinberg. The Cosmological Constant Problem. Reviews of Modern Physics 61 (1989), 1.

[51] G. Dvali, S. Hofmann, J. Khoury. Degravitation of the Cosmological Constant and Graviton Width. Physical Review D 76 (2007), 084006. См. также препринт arxiv.org/abs/hep-th/0703027

[52] G. Dvali, G. Gabadadze, M. Porrati. Metastable Gravitons and Infinite Volume Extra Dimensions. Physics Letters B 484 (2000), 112. См. также препринт arxiv.org/abs/hep-th/0002190

[53] S.L.Dubovsky, V.A.Rubakov. Brane-induced gravity in more than one extra dimensions: violation of equivalence principle and ghost. Physical Review D 67 (2003), 104014. См. также препринт arxiv.org/abs/hep-th/0212222

[54] F. Berkhahn, S. Hofmann, F. Niedermann. Brane Induced Gravity: From a No-Go to a No-Ghost Theorem. Physical Review D 86 (2012), 124022. См. также препринт arxiv.org/abs/1205.6801

[55] C. de Rham, G. Dvali, S. Hofmann, J. Khoury, O. Pujolas, M. Redi, A.J. Tolley. Cascading DGP. Physical Review Letters 100 (2008), 251603. См. также препринт arxiv.org/abs/0711.2072

[56] F.L. Bezrukov, M.E. Shaposhnikov. The Standard Model Higgs boson as the inflaton. Physics Letters B 659 (2008), 703. См. также препринт arxiv.org/abs/0710.3755

[57] A. Ijjas, P.J. Steinhardt, A. Loeb. Inflationary paradigm in trouble after Planck2013. Physics Letters B 723 (2013), 261. См. также препринт arxiv.org/abs/1304.2785

[58] Sergei Winitzki. Eternal Inflation. Singapore, World Scientific 2009.

[59] C. Deffayet, G. Esposito-Farèse, A. Vikman. Covariant Galileon. Physical Review D 79 (2009), 084003. См. также препринт arxiv.org/abs/0901.1314

[60] M.S. Turner, L.M. Widrow. Inflation-produced, large-scale magnetic fields. Physical Review D 37 (1988), 2743.

[61] D. Grasso, H.R. Rubinstein. Magnetic fields in the early Universe. Physics Reports 348 (2001), 163; См. также препринт arxiv.org/abs/astro-ph/0009061

[62] V. Demozzi, V. Mukhanov, H. Rubinstein. Magnetic fields from inflation? Journal of Cosmology and Astroparticle Physics 2009, JCAP08(2009)025; См. также препринт arxiv.org/abs/0907.1030

[63] L.H. Ford. Inflation driven by a vector field. Physical Review D 40 (1989), 967.

[64] C. Armendariz-Picon. Could dark energy be vector-like? Journal of Cosmology and Astroparticle Physics 2004, JCAP07(2004)007. См. также препринт arxiv.org/abs/astro-ph/0405267

[65] Y. Hosotani. Exact solutions to the Einstein-Yang-Mills equation. Physics Letters B 147 (1984), 44.

[66] D.V. Galt'sov, M.S. Volkov. Yang-Mills cosmology. Cold matter for a hot universe. Physics Letters B 256 (1991), 17.

[67] V.V. Kiselev. Vector field as a quintessence partner. Classical and Quantum Gravity 21 (2004), 3323. См. также препринт arxiv.org/abs/gr-qc/0402095

[68] S.M. Carroll, E.A. Lim. Lorentz-violating vector fields slow the universe down. Physical Review D 70 (2004), 123525. См. также препринт arxiv.org/abs/hep-th/0407149

[69] T.S. Koivisto, D.F. Mota. Vector Field Models of Inflation and Dark Energy. Journal of Cosmology and Astroparticle Physics 2008, JCAP08(2008)021. См. также препринт arxiv.org/abs/0805.4229

[70] B. Himmetoglu, C. Contaldi, M. Peloso. Instability of anisotropic cosmological solutions supported by vector fields. Physical

Review Letters 102 (2009), 111301. См. также препринт arxiv.org/abs/0809.2779

[71] B. Himmetoglu, C. Contaldi, M. Peloso. Instability of the ACW model, and problems with massive vectors during inflation. Physical Review D 79 (2009), 063517. См. также препринт arxiv.org/abs/0812.1231

[72] B. Himmetoglu, C. Contaldi, M. Peloso. Ghost instabilities of cosmological models with vector fields nonminimally coupled to the curvature. Physical Review D 80 (2009), 123530. См. также препринт arxiv.org/abs/0909.3524

[73] K. Dimopoulos, M. Karciauskas, D. Lyth, Y. Rodriguez. Statistical anisotropy of the curvature perturbation from vector field perturbations. Journal of Cosmology and Astroparticle Physics 2009, JCAP05(2009)013. См. также препринт arxiv.org/abs/0809.1055

[74] S. Dimopoulos, S. Kachru, J. McGreevy, J.G. Wacker. N-flation. Journal of Cosmology and Astroparticle Physics 2008, JCAP08(2008)003. См. также препринт arxiv.org/abs/hep-th/0507205

[75] M. Karciauskas, K. Dimopoulos, D.H. Lyth. Anisotropic non-Gaussianity from vector field perturbations. Physical Review D 80 (2009), 023509. См. также препринт arxiv.org/abs/0812.0264

[76] C.A. Valenzuela-Toledo, Y. Rodriguez, D.H. Lyth. Non-gaussianity at tree and one-loop levels from vector field perturbations. Physical Review D 80 (2009), 103519. См. также препринт arxiv.org/abs/0909.4064

[77] M. Karciauskas, D.H. Lyth. On the health of a vector field with $RA^2/6$ coupling to gravity. Journal of Cosmology and Astroparticle Physics 2010, JCAP11(2010)023. См. также препринт arxiv.org/abs/1007.1426

[78] D. Wands, K.A. Malik, D.H. Lyth, A.R. Liddle. A new approach to the evolution of cosmological perturbations on large scales. Physical Review D 62 (2000), 043527. См. также препринт arxiv.org/abs/astro-ph/0003278

[79] M. Novello, S.E. Perez Bergliaffa, J. Salim. Non-linear electrodynamics and the acceleration of the universe. Physical Review D 69 (2004), 127301. См. также препринт arxiv.org/abs/astro-ph/0312093

[80] G. Esposito-Farèse, C. Pitrou, J.-P. Uzan. Vector theories in cosmology. Physical Review D 81 (2010), 063519. См. также препринт arxiv.org/abs/0912.0481

[81] C.G. Böhmer. Dark spinor inflation – theory primer and dynamics. Physical Review D 77 (2008), 123535. См. также препринт arxiv.org/abs/0804.0616

[82] D. Gredat, S. Shankaranarayanan. Modified scalar and tensor spectra in spinor driven inflation. Journal of Cosmology and Astroparticle Physics 2010, JCAP01(2010)008. См. также препринт arxiv.org/abs/0807.3336

[83] R.W. Wald. Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant. Physical Review D 28 (1983), 2118(R).

[84] C. Germani, A. Kehagias. P-nflation: generating cosmic Inflation with p-forms. Journal of Cosmology and Astroparticle Physics 2009, JCAP03(2009)028. См. также препринт arxiv.org/abs/0902.3667

[85] T.S. Koivisto, D.F. Mota, C. Pitrou. Inflation from N-Forms and its stability. Journal of High Energy Physics 2009, JHEP09(2009)092. См. также препринт arxiv.org/abs/0903.4158

[86] C. Germani, A. Kehagias. Scalar perturbations in p-nflation: the 3-form case. Journal of Cosmology and Astroparticle Physics 2009, JCAP11(2009)005. См. также препринт arxiv.org/abs/0908.0001

[87] T. Kobayashi, S. Yokoyama. Gravitational waves from p-form inflation. Journal of Cosmology and Astroparticle Physics 2009, JCAP05(2009)004. См. также препринт arxiv.org/abs/0903.2769

[88] T.S. Koivisto, N.J. Nunes. Three-form cosmology. Physics Letters B 685 (2010), 105. См. также препринт arxiv.org/abs/0907.3883

[89] T.S. Koivisto, N.J. Nunes. Inflation and dark energy from three-forms. Physical Review D 80 (2009), 103509. См. также препринт arxiv.org/abs/0908.0920

[90] M. Watanabe, S. Kanno, J. Soda. Inflationary Universe with Anisotropic Hair. Physical Review Letters 102 (2009), 191302. См. также препринт arxiv.org/abs/0902.2833

[91] S. Kanno, J. Soda, M. Watanabe. Anisotropic Power-law Inflation. Journal of Cosmology and Astroparticle Physics 2010, JCAP12(2010)024. См. также препринт arxiv.org/abs/1010.5307

[92] S. Hervik, D.F. Mota, M. Thorsrud. Inflation with stable anisotropic hair: is it cosmologically viable? Journal of High Energy Physics 2011, JHEP11(2011)146. См. также препринт arxiv.org/abs/1109.3456

[93] M. Karčiauskas. Dynamical Analysis of Anisotropic Inflation. Modern Physics Letters A 31 (2016), 1640002. См. также препринт arxiv.org/abs/1604.00269

[94] A. Maleknejad, M.M. Sheikh-Jabbari. Gauge-flation: Inflation From Non-Abelian Gauge Fields. Physics Letters B 723 (2013), 224. См. также препринт arxiv.org/abs/1102.1513

[95] A. Maleknejad, M.M. Sheikh-Jabbari. Non-Abelian Gauge Field Inflation. Physical Review D 84 (2011), 043515. См. также препринт arxiv.org/abs/1102.1932

[96] A. Maleknejad, M.M. Sheikh-Jabbari, J. Soda. Gauge Fields and Inflation. Physics Reports 528 (2013), 161. См. также препринт arxiv.org/abs/1212.2921

[97] P. Adshead, E.I. Sfakianakis. Higgsed Gauge-flation. Journal of High Energy Physics 2017, JHEP08(2017)130. См. также препринт arxiv.org/abs/1705.03024

[98] J. Beltrán Jiménez, R. Durrer, L. Heisenberg, M. Thorsrud. Stability of Horndeski vector-tensor interactions. Journal of Cosmology and Astroparticle Physics, JCAP10(2013)064. См. также препринт arxiv.org/abs/1308.1867

[99] K. Hinterbichler. Theoretical Aspects of Massive Gravity. Reviews of Modern Physics 84 (2012), 671. См. также препринт arxiv.org/abs/1105.3735

[100] C. de Rham. Massive Gravity. Living Reviews in Relativity 17 (2014), 7. См. также препринт arxiv.org/abs/1401.4173

[101] A. Schmidt-May, M. von Strauss. Recent developments in bimetric theory. Journal of Physics A: Mathematical and Theoretical 49 (2016), 183001 (topical review). См. также препринт arxiv.org/abs/1512.00021

[102] K. Hinterbichler, R.A. Rosen. Interacting Spin-2 Fields. Journal of High Energy Physics 2012, JHEP07(2012)047. См. также препринт arxiv.org/abs/1203.5783

[103] M. Fierz, W. Pauli. On relativistic wave equations for particles of arbitrary spin in an electromagnetic field. Proceedings of the Royal Society A 173 (1939), 211.

[104] H. van Dam, M.J.G. Veltman. Massive and mass-less Yang-Mills and gravitational fields. Nuclear Physics B 22 (1970), 397.

[105] В.И. Захаров. Линеаризованная теория гравитации и масса гравитона. Письма в ЖЭТФ 12 (1970), 447.

[106] A.I. Vainshtein. To the problem of nonvanishing gravitation mass. Physics Letters B 39 (1972), 393.

[107] J. Khoury, A. Weltman. Chameleon Fields: Awaiting Surprises for Tests of Gravity in Space. Physical Review Letters 93 (2004), 171104. См. также препринт arxiv.org/abs/astro-ph/0309300

[108] K. Hinterbichler, J. Khoury. Symmetron Fields: Screening Long-Range Forces Through Local Symmetry Restoration. Physical Review Letters 104 (2010), 231301. См. также препринт arxiv.org/abs/1001.4525

[109] E. Babichev, C. Deffayet. An introduction to the Vainshtein mechanism. Classical and Quantum Gravity 30 (2013), 184001 (2013). См. также препринт arxiv.org/abs/1304.7240

[110] D.G. Boulware, S. Deser. Can Gravitation Have a Finite Range? Physical Review D 6 (1972), 3368.

[111] C. de Rham, G. Gabadadze. Generalization of the Fierz-Pauli Action. Physical Review D 82 (2010) 044020. См. также препринт arxiv.org/abs/1007.0443

[112] N. Arkani-Hamed, H. Georgi, M.D. Schwartz. Effective Field Theory for Massive Gravitons and Gravity in Theory Space. Annals of Physics 305 (2003), 96. См. также препринт arxiv.org/abs/hep-th/0210184

[113] C. de Rham, G. Gabadadze, A.J. Tolley. Resummation of Massive Gravity. Physical Review Letters 106 (2011), 231101. См. также препринт arxiv.org/abs/1011.1232

[114] C. de Rham, G. Gabadadze, A. Tolley. Ghost free Massive Gravity in the Stückelberg language. Physics Letters B 711 (2012), 190. См. также препринт arxiv.org/abs/1107.3820

[115] S.F. Hassan, R.A. Rosen. On Non-Linear Actions for Massive Gravity. Journal of High Energy Physics 2011, JHEP07(2011)009. См. также препринт arxiv.org/abs/1103.6055

[116] S.F. Hassan, R.A. Rosen. Resolving the Ghost Problem in non-Linear Massive Gravity. Physical Review Letters 108 (2012) 041101. См. также препринт arxiv.org/abs/1106.3344

[117] S.F. Hassan, R.A. Rosen A. Schmidt-May. Ghost-free Massive Gravity with a General Reference Metric. Journal of High Energy Physics 2012, JHEP02(2012)026. См. также препринт arxiv.org/abs/1109.3230

[118] S.F. Hassan, R.A. Rosen. Bimetric Gravity from Ghost-free Massive Gravity. Journal of High Energy Physics 2012, JHEP02(2012)126. См. также препринт arxiv.org/abs/1109.3515

[119] S.F. Hassan, R.A. Rosen. Confirmation of the Secondary Constraint and Absence of Ghost in Massive Gravity and Bimetric Gravity. Journal of High Energy Physics 2012, JHEP04(2012)123. См. также препринт arxiv.org/abs/1111.2070

[120] S. F. Hassan, A. Schmidt-May, and M. von Strauss. Proof of Consistency of Nonlinear Massive Gravity in the Stuckelberg Formulation. Physics Letters B 715 (2012), 335. См. также препринт arxiv.org/abs/1203.5283

[121] C. Deffayet, J. Mourad and G. Zahariade. Covariant constraints in ghost free massive gravity. Journal of Cosmology and Astroparticle Physics 2013, JCAP01(2013)032. См. также препринт arxiv.org/abs/1207.6338

[122] C. Deffayet, J. Mourad. G. Zahariade. A note on "symmetric"vielbeins in bimetric, massive, perturbative and non perturbative gravities. Journal of High Energy Physics 2013, JHEP03(2013)086. См. также препринтarxiv.org/abs/1208.4493

[123] P. Gratia, W. Hu, M. Wyman. Self-accelerating Massive Gravity: How Zweibeins Walk through Determinant Singularities. Classical and Quantum Gravity 30 (2013), 184007. См. также препринт arxiv.org/abs/1305.2916

[124] P. Gratia, W. Hu, M. Wyman. Self-accelerating Massive Gravity: Bimetric Determinant Singularities. Physical Review D 89 (2014), 027502. См. также препринт arxiv.org/abs/1309.5947

[125] D. Comelli, M. Crisostomi, K. Koyama, L. Pilo, G. Tasinato. New Branches of Massive Gravity. Physical Review D 91 (2015), 121502 (2015); См. также препринт arxiv.org/abs/1505.00632

[126] N.A. Ondo, A.J. Tolley. Complete Decoupling Limit of Ghost-free Massive Gravity. Journal of High Energy Physics 2013, JHEP11(2013)059. См. также препринт arxiv.org/abs/1307.4769

[127] C. de Rham, A.J. Tolley, S.-Y. Zhou. The Λ_2 limit of massive gravity. Journal of High Energy Physics 2016, JHEP04(2016)188. См. также препринт arxiv.org/abs/1602.03721

[128] C. de Rham, L. Heisenberg, R.H. Ribeiro. On couplings to matter in massive (bi-)gravity. Classical and Quantum Gravity 32 (2015), 035022. См. также препринт arxiv.org/abs/1408.1678

[129] C. de Rham, L. Heisenberg, R.H. Ribeiro. Ghosts and matter couplings in massive gravity, bigravity and multigravity. Physical Review D 90 (2014), 124042. См. также препринт arxiv.org/abs/1409.3834

[130] S.F. Hassan, M. Kocic. On the local structure of spacetime in ghost-free bimetric theory and massive gravity. Препринт arxiv.org/abs/1706.07806

[131] S.F. Hassan, A. Schmidt-May, M. von Strauss. On Partially Massless Bimetric Gravity. Physics Letters B 726 (2013), 834. См. также препринт arxiv.org/abs/1208.1797

[132] L. Apolo, S.F. Hassan, A. Lundkvist. Gauge and global symmetries of the candidate partially massless bimetric gravity. Physical Review D 94 (2016), 124055. См. также препринт arxiv.org/abs/1609.09515

[133] Ф.Р. Гантмахер. Теория Матриц. Москва, "Наука" 1988.

[134] L. Bernard, C. Deffayet, A. Schmidt-May, M. von Strauss. Linear spin-2 fields in most general backgrounds. Physical Review D 93 (2016), 084020. См. также препринт arxiv.org/abs/1512.03620

[135] S. Deser, K. Izumi, Y.C. Ong, A. Waldron. Problems of Massive Gravities. Modern Physics Letters A 30 (2015) 1540006. См. также препринт arxiv.org/abs/1410.2289

[136] С.В. Красников. Некоторые вопросы причинности в ОТО: "машины времени" и "сверхсветовые перемещения". Москва, "ЛЕНАНД" 2015.

[137] C. Burrage, C. de Rham, L. Heisenberg, A.J. Tolley. Chronology Protection in Galileon Models and Massive Gravity. Journal of Cosmology and Astroparticle Physics 2012, JCAP07(2012)004. См. также препринт arxiv.org/abs/1111.5549

[138] C. de Rham, L. Heisenberg, R.H. Ribeiro. Quantum Corrections in Massive Gravity. Physical Review D 88 (2013), 084058. См. также препринт arxiv.org/abs/1307.7169

[139] G. D'Amico, C. de Rham, S. Dubovsky, G. Gabadadze, D. Pirtskhalava, A.J. Tolley. Massive Cosmologies. Physical Review D 84 (2011), 124046 (2011). См. также препринт arxiv.org/abs/1108.5231

[140] A.E. Gümrükçüoğlu, C. Lin, S. Mukohyama. Open FRW universes and self-acceleration from nonlinear massive gravity. Journal of Cosmology and Astroparticle Physics 2011, JCAP11(2011)030. См. также препринт arxiv.org/abs/1109.3845

[141] A. De Felice, A.E. Gümrükçüoğlu, C. Lin, S. Mukohyama. Nonlinear stability of cosmological solutions in massive gravity. Journal of Cosmology and Astroparticle Physics 2013, JCAP05(2013)035. См. также препринт arxiv.org/abs/1303.4154

[142] M. von Strauss, A. Schmidt-May, J. Enander, E. Mortsell, S.F. Hassan. Cosmological Solutions in Bimetric Gravity and their Observational Tests. Journal of Cosmology and Astroparticle Physics 2012, JCAP03(2012)042. См. также препринт arxiv.org/abs/1111.1655

[143] Y. Akrami, S.F. Hassan, F. Könnig, A. Schmidt-May, A.R. Solomon. Bimetric gravity is cosmologically viable. Physics Letters B 748 (2015), 37. См. также препринт arxiv.org/abs/1503.07521

[144] E. Babichev, L. Marzola, M. Raidal, A. Schmidt-May, F. Urban, H. Veermäe, M. von Strauss. Heavy spin-2 Dark Matter. Journal of Cosmology and Astroparticle Physics 2016, JCAP09(2016)016. См. также препринт arxiv.org/abs/1607.03497

[145] Q.-G. Huang, Y.-S. Piao, S.-Y. Zhou. Mass-Varying Massive Gravity. Physical Review D 86 (2012), 124014. См. также препринт arxiv.org/abs/1206.5678

[146] Y.-F. Cai, F. Duplessis, and E.N. Saridakis. $F(R)$ nonlinear massive theories of gravity and their cosmological implications. Physical Review D 90 (2014), 064051. См. также препринт arxiv.org/abs/1307.7150

[147] G. D'Amico, G. Gabadadze, L. Hui, D. Pirtskhalava. Quasi-Dilaton: Theory and Cosmology. Physical Review D 87 (2013), 064037; arxiv.org/abs/1206.4253

[148] A.E. Gümrükçüoğlu, K. Hinterbichler, C. Lin, S. Mukohyama, M. Trodden. Cosmological Perturbations in Extended Massive Gravity. Physical Review D 88 (2013), 024023; См. также препринт arxiv.org/abs/1304.0449

[149] A. De Felice, S. Mukohyama. Towards consistent extension of quasidilaton massive gravity. Physics Letters B 728 (2014), 622. См. также препринт arxiv.org/abs/1306.5502

[150] L. Heisenberg. Revisiting perturbations in extended quasidilaton massive gravity. Journal of Cosmology and Astroparticle Physics 2015, JCAP04(2015)010. См. также препринт arxiv.org/abs/1501.07796

[151] S. Mukohyama. Extended quasidilaton massive gravity is ghost free. Препринт arxiv.org/abs/1309.2146v1

[152] J. Kluson. Note About Consistent Extension of Quasidilaton Massive Gravity. Journal of Gravity 2014, 413835. См. также препринт arxiv.org/abs/1309.0956

[153] S. Anselmi, S. Kumar, D. López Nacir, G.D. Starkman. Failures of homogeneous and isotropic cosmologies in Extended Quasi-Dilaton Massive Gravity. Препринт arxiv.org/abs/1706.01872

[154] S. Mukohyama. Boulware-Deser ghost in extended quasidilaton massive gravity. Препринт arxiv.org/abs/1309.2146v2 (вторая версия работы [151]). Опубликовано в Physical Review D 96 (2017), 044029.

[155] A.E. Gümrükçüoğlu, K. Koyama, S. Mukohyama.Stable cosmology in ghost-free quasidilaton theory. Physical Review D 96 (2017), 044041. См. также препринт

[156] C. de Rham, A. Matas, A.J. Tolley. New Kinetic Interactions for Massive Gravity? Classical and Quantum Gravity 31 (2014), 165004. См. также препринт arxiv.org/abs/1707.02004 arxiv.org/abs/1311.6485

[157] C. de Rham, A. Matas, A.J. Tolley. New Kinetic Terms for Massive Gravity and Multi-gravity: A No-Go in Vielbein Form. Classical and Quantum Gravity 32 (2015), 215027. См. также препринт arxiv.org/abs/1505.00831

[158] M. Milgrom. Bimetric MOND gravity. Physical Review D 80 (2009),123536. См. также препринт arxiv.org/abs/0912.0790

[159] L. Blanchet, L. Heisenberg. Dipolar dark matter with massive bigravity. Journal of Cosmology and Astroparticle Physics 2015, JCAP12(2015)026. См. также препринт arxiv.org/abs/1505.05146

[160] T.S. Koivisto. On new variational principles as alternatives to the Palatini method. Physical Review D 83 (2011),101501. См. также препринт arxiv.org/abs/1103.2743

[161] L. Amendola, K. Enqvist, and T. Koivisto. Unifying Einstein and Palatini gravities. Physical Review D 83 (2011), 044016. См. также препринт arxiv.org/abs/1010.4776

[162] J. Beltrán Jiménez, T.S. Koivisto. Spacetimes with vector distortion: Inflation from generalised Weyl geometry. Physics Letters B 756 (2016), 400. См. также препринт arxiv.org/abs/1509.02476.

[163] M. Sandstad, T.S. Koivisto, D.F. Mota. Non-locality of the C- and D-theories. Classical and Quantum Gravity 30 (2013), 155005. См. также препринт arxiv.org/abs/1305.0695

[164] T.S. Koivisto. The post-Newtonian limit in C-theories of gravitation. Physical Review D 84 (2011), 121502(R). См. также препринт arxiv.org/abs/1109.4585

[165] T. Biswas, E. Gerwick, T. Koivisto, A. Mazumdar. Towards singularity and ghost free theories of gravity. Physical Review Letters 108 (2012), 031101. См. также препринт arxiv.org/abs/1110.5249

[166] I. Dimitrijevic, B. Dragovich, J. Grujic, Z. Rakic. Some Cosmological Solutions of a Nonlocal Modified Gravity. Filomat 29 (2015), 619. См. также препринт arxiv.org/abs/1508.05583

[167] A.H. Chamseddine, V. Mukhanov. Mimetic Dark Matter. Journal of High Energy Physics 2013, JHEP11(2013)135. См. также препринт arxiv.org/abs/1308.5410

[168] E. Langmann, M. Sundin. Extrinsic curvature effects in brane-world scenarios. Препринт arxiv.org/abs/1103.3230

[169] S.L. Cherkas, V.L. Kalashnikov. An inhomogeneous toy-model of the quantum gravity with explicitly evolvable observables. General Relativity and Gravitation 44 (2012), 3081. См. также препринт arxiv.org/abs/1107.2224

[170] R. Aldrovandi, J.G. Pereira. Teleparallel gravity. An introduction. Springer 2013.

[171] B. Li, T.P. Sotiriou, J.D. Barrow. $f(T)$ Gravity and local Lorentz invariance. Physical Review D 83 (2011), 064035. См. также препринт arxiv.org/abs/1010.1041

[172] M. Krššák, E.N. Saridakis. The covariant formulation of $f(T)$ gravity. Classical and Quantum Gravity 33 (2016), 115009. См. также препринт arxiv.org/abs/1510.08432

[173] Y.-F. Cai, S. Capozziello, M. de Laurentis, E.N. Saridakis. $f(T)$ teleparallel gravity and cosmology. Reports on Progress in Physics 79 (2016), 106901. См. также препринт arxiv.org/abs/1511.07586

[174] N. Deruelle, J. Rua. Disformal Transformations, Veiled General Relativity and Mimetic Gravity. Journal of Cosmology and Astroparticle Physics 2014, JCAP09(2014)002. См. также препринт arxiv.org/abs/1407.0825

[175] S. Nojiri, S.D. Odintsov. Mimetic $F(R)$ gravity: inflation, dark energy and bounce. Modern Physics Letters A 29 (2014), 1450211. См. также препринт arxiv.org/abs/1408.3561

[176] K. Hammer, A. Vikman. Many Faces of Mimetic Gravity. Препринт arxiv.org/abs/1512.09118

[177] P. Kroupa. The Dark Matter Crisis: Falsification of the Current Standard Model of Cosmology. Publications of the Astronomical Society of Australia 29 (2012), 395. См. также препринт arxiv.org/abs/1204.2546

[178] P.J.E. Peebles, A. Nusser. Clues from nearby galaxies to a better theory of cosmic evolution. Nature 465 (2010), 565. См. также препринт arxiv.org/abs/1001.1484

[179] N. Menci, F. Fiore, A. Lamastra. Galaxy Formation in WDM Cosmology. Monthly Notices of the Royal Astronomical Society 421 (2012), 2384. См. также препринт arxiv.org/abs/1201.1617

[180] A. Kashlinsky, F. Atrio-Barandela, D. Kocevski, H. Ebeling. A measurement of large-scale peculiar velocities of clusters of galaxies: results and cosmological implications. The Astrophysical Journal Letters 686 (2008), L49. См. также препринт arxiv.org/abs/0809.3734

[181] S.J. Osborne, D.S.Y. Mak, S.E. Church, E. Pierpaoli. Measuring the Galaxy Cluster Bulk Flow from WMAP data. The Astrophysical Journal 737 (2011), 98. См. также препринт arxiv.org/abs/1011.2781

[182] Planck Collaboration. Planck intermediate results XIII. Constraints on peculiar velocities. Astronomy & Astrophysics 561 (2014), A97. См. также препринт arxiv.org/abs/1303.5090

[183] B.R. Tully, J.R. Fisher. A new method of determining distances to galaxies. Astronomy & Astrophysics 54 (1977), 661.

[184] S.S. McGaugh, J.M. Schombert, G.D. Bothun, W.J.G. de Blok. The Baryonic Tully-Fisher Relation. The Astrophysical Journal Letters 533 (2000), L99. См. также препринт astro-ph/0003001

[185] A. Aguirre, J. Schaye, E. Quataert. Problems for MOND in Clusters and the Ly-α Forest. The Astrophysical Journal 561 (2001), 550. См. также препринт astro-ph/0105184

[186] A. Nusser. Modified Newtonian Dynamics of Large Scale Structure. Monthly Notices of the Royal Astronomical Society 331 (2002), 909. См. также препринт astro-ph/0109016

[187] C. Llinares, A. Knebe, H.S. Zhao. Cosmological Structure Formation under MOND: a new numerical solver for Poisson's equation. Monthly Notices of the Royal Astronomical Society 391 (2008), 1778. См. также препринт arxiv.org/abs/0809.2899

[188] G.W. Angus, A. Diaferio. The abundance of galaxy clusters in MOND: Cosmological simulations with massive neutrinos. Monthly Notices of the Royal Astronomical Society 417 (2011), 941. См. также препринт arxiv.org/abs/1104.5040

[189] J. Bekenstein, M. Milgrom. Does the missing mass problem signal the breakdown of Newtonian gravity? The Astrophysical Journal 286 (1984), 7.

[190] J. Bekenstein. Relativistic gravitation theory for the MOND paradigm. Physical Review D 70 (2004), 083509. См. также препринт arxiv.org/abs/astro-ph/0403694

[191] C. Skordis, D.F. Mota, P.G. Ferreira, C. Bœhm. Large Scale Structure in Bekenstein's theory of relativistic Modified Newtonian Dynamics. Physical Review Letters 96 (2006), 011301. См. также препринт arxiv.org/abs/astro-ph/0505519

[192] S. Dodelson, M. Liguori. Can Cosmic Structure form without Dark Matter? Physical Review Letters 97 (2006), 231301. См. также препринт arxiv.org/abs/astro-ph/0608602

[193] M. Milgrom. Can the hidden mass be negative? The Astrophysical Journal 306 (1986), 9.

[194] D. Dai, R. Matsuo, G. Starkman. Limited utility of Birkhoff's theorem in modified Newtonian dynamics: Nonzero accelerations inside a shell. Physical Review D 81 (2010), 024041. См. также препринт arxiv.org/abs/0811.1565

[195] D. Clowe, M. Bradač, A.H. Gonzalez, M. Markevitch, S.W. Randall, Ch. Jones, D. Zaritsky. A direct empirical proof of the existence of dark matter. The Astrophysical Journal Letters 648 (2006), 109. См. также препринт arxiv.org/abs/astro-ph/0608407

[196] G.W. Angus, B. Famaey, H.S. Zhao. Can MOND take a bullet? Analytical comparisons of three versions of MOND beyond spherical symmetry. Monthly Notices of the Royal Astronomical Society 371 (2006), 138. См. также препринт astro-ph/0606216

[197] J. Lee, E. Komatsu. Bullet Cluster: A Challenge to ΛCDM Cosmology. The Astrophysical Journal 718 (2010), 60. См. также препринт arxiv.org/abs/1003.0939

[198] D. Clowe, M. Markevitch, M. Bradač, A.H. Gonzalez, S.M. Chung, R. Massey, D. Zaritsky. On Dark Peaks and Missing Mass: A Weak Lensing Mass Reconstruction of the Merging Cluster System Abell 520. The Astrophysical Journal 758 (2012), 128. См. также препринт https://arxiv.org/abs/1209.2143

[199] J. Khoury. An Alternative to Particle Dark Matter. Physical Review D 91 (2015), 024022. См. также препринт arxiv.org/abs/1409.0012

[200] П.А.М. Дирак. Лекции по квантовой механике. Ижевск, R&C Dynamics 1998.

[201] H. Kleinert, S.V. Shabanov. Proper Dirac Quantization of Free Particle on D-Dimensional Sphere. Physics Letters A 232, 327 (1997). См. также препринт quant-ph/9702006

[202] B. Podolsky. Quantum-Mechanically Correct Form of Hamiltonian Function for Conservative Systems. Physical Review 32 (1928), 812.

[203] L.D. Faddeev, S.L. Shatashvili. Realization of the Schwinger term in the Gauss law and the possibility of correct quantization of a theory with anomalies. Physics Letters B 167 (1986), 225.

[204] J.R. Klauder, S.V. Shabanov. Coordinate-free quantization of second-class constraints. Nuclear Physics B 511 (1998), 713. См. также препринт arxiv.org/abs/hep-th/9702102

[205] H. Jensen, H. Koppe. Quantum mechanics with constraints. Annals of Physics 63 (1971), 586.

[206] R.C.T. da Costa. Quantum mechanics of a constrained particle. Physical Review A 23 (1981), 1982.

[207] K.A. Mitchell. Geometric Phase, Curvature, and Extrapotentials in Constrained Quantum Systems. Physical Review A 63 (2001), 042112. См. также препринт arxiv.org/abs/quant-ph/0001059

[208] P.C. Schuster, R.L. Jaffe. Quantum Mechanics on Manifolds Embedded in Euclidean Space. Annals of Physics 307 (2003), 132. См. также препринт arxiv.org/abs/hep-th/0302216

[209] R.A. Marcus. On the Analytical Mechanics of Chemical Reactions. Quantum Mechanics of Linear Collisions. The Journal of Chemical Physics 45 (1966), 4493.

[210] M. Encinosa, B. Etemadi. Energy shifts resulting from surface curvature of quantum nanostructures. Physical Review A 58 (1998), 77.

[211] M. Encinosa, L. Mott, B. Etemadi. Wave functions in the neighborhood of a toroidal surface; hard vs. soft constraint. Physica Scripta 72 (2005), 13. См. также препринт quant-ph/0409141

[212] P. Maraner. Monopole Gauge Fields and Quantum Potentials Induced by the Geometry in Simple Dynamical Systems. Annals of Physics 246 (1996), 325. См. также препринт arxiv.org/abs/hep-th/9406004

[213] G. Dell'Antonio, L. Tenuta. Quantum graphs as holonomic constraints. Journal of Mathematical Physics 47 (2006), 072102. См. также препринт arxiv.org/abs/math-ph/0603044

[214] R. Froese, I. Herbst. Realizing holonomic constraints in classical and quantum mechanics. Communications in Mathematical Physics 220 (2001), 489. См. также препринт arxiv.org/abs/math-ph/9909007

[215] J. Wachsmuth, S. Teufel. Effective Hamiltonians for Constrained Quantum Systems. Memoirs of the American Mathematical Society 230 (2014), MEMO/1083. См. также препринт arxiv.org/abs/0907.0351

[216] S. Takagi, T. Tanzawa. Quantum Mechanics of a Particle Confined to a Twisted Ring. Progress of Theoretical Physics 87 (1992), 561.

[217] R.C.T. da Costa. Constraints in quantum mechanics. Physical Review A 25 (1982), 2893.

[218] P. Maraner, C. Destri. Geometry-induced Yang-Mills fields in constrained quantum mechanics. Modern Physics Leters A 8 (1993), 861.

[219] P. Maraner. A Complete Perturbative Expansion for Constrained Quantum Dynamics. Journal of Physics A 28 (1995), 2939. См. также препринт arxiv.org/abs/hep-th/9409080

[220] K. Fujii, N. Ogawa. Generalization of Geometry-Induced Gauge Structure to Any Dimensional Manifold. Progress of Theoretical Physics 89 (1993), 575.

[221] K. Fujii, N. Ogawa, S. Uchiyama, N.M. Chepilko. Geometrically Induced Gauge Structure on Manifolds Embedded in a Higher-Dimensional Space. International Journal of Modern Physics A 12 (1997), 5235. См. также препринт arxiv.org/abs/hep-th/9702191

[222] B.R. Greene, J. Levin. Dark Energy and Stabilization of Extra Dimensions. Journal of High Energy Physics 2007, JHEP11(2007)096. См. также препринт arxiv.org/abs/0707.1062

[223] T. Thiemann. The LQG – String: Loop Quantum Gravity Quantization of String Theory I. Flat Target Space. Classical and Quantum Gravity 23 (2006), 1923. См. также препринт arxiv.org/abs/hep-th/0401172

[224] R.C. Helling. Lessons from the LQG String. Препринт arxiv.org/abs/hep-th/0610193

[225] S.W. Hawking. Particle creation by black holes. Communications in Mathematical Physics 43 (1975), 199.

[226] L. Susskind, L. Thorlacius, J. Uglum. The Stretched Horizon and Black Hole Complementarity. Physical Review D 48 (1993), 3743. См. также препринт arxiv.org/abs/hep-th/9306069

[227] A. Almheiri, D. Marolf, J. Polchinski, J. Sully. Black Holes: Complementarity or Firewalls? Journal of High Energy Physics 2013, JHEP02(2013)062. См. также препринт arxiv.org/abs/1207.3123

[228] S.L. Braunstein, S. Pirandola, K. Zyczkowski. Better Late than Never: Information Retrieval from Black Holes. Physical Review Letters 110 (2013), 101301. См. также препринт arxiv.org/abs/0907.1190

[229] S.L. Braunstein, S. Pirandola. Evaporating black holes have leaky horizons or exotic atmospheres. Препринт arxiv.org/abs/1311.1326

[230] D.N. Page. Is Black-Hole Evaporation Predictable? Physical Review Letters 44 (1980), 301.

[231] D.N. Page. Average Entropy of a Subsystem. Physical Review Letters 71 (1993), 1291. См. также препринт arxiv.org/abs/gr-qc/9305007

[232] Y. Sekino, L. Susskind. Fast Scramblers. Journal of High Energy Physics 2008, JHEP10(2008)065. См. также препринт arxiv.org/abs/0808.2096

[233] L. Susskind. Singularities, Firewalls, and Complementarity. Препринт arxiv.org/abs/1208.3445

[234] R. Bousso. Frozen Vacuum. Physical Review Letters 112 (2014), 041102. См. также препринт arxiv.org/abs/1308.3697

[235] R. Bousso. Complementarity Is Not Enough. Physical Review D 87 (2013), 124023. См. также препринт arxiv.org/abs/1207.5192v2 (Обратим внимание, что первая версия arxiv.org/abs/1207.5192v1 этого препринта содержит противоположный вывод: Observer Complementarity Upholds the Equivalence Principle.)

[236] D. Harlow, P. Hayden. Quantum Computation vs. Firewalls. Journal of High Energy Physics 2013, JHEP06(2013)085. См. также препринт arxiv.org/abs/1301.4504

[237] L. Susskind. Butterflies on the Stretched Horizon. Препринт arxiv.org/abs/1311.7379

[238] Y. Nomura, J. Varela, S.J. Weinberg. Complementarity Endures: No Firewall for an Infalling Observer. Journal of High Energy Physics 2013, JHEP03(2013)059. См. также препринт arxiv.org/abs/1207.6626

[239] L. Susskind. New Concepts for Old Black Holes. Препринт arxiv.org/abs/1311.3335

[240] J. Maldacena, L. Susskind. Cool horizons for entangled black holes. Fortschritte der Physik 61 (2013), 781. См. также препринт arxiv.org/abs/1306.0533

[241] W.A. Christiansen, D.J.E. Floyd, Y.J. Ng, E.S. Perlman. Limits on Spacetime Foam. Physical Review D 83 (2011), 084003. См. также препринт arxiv.org/abs/0912.0535

[242] G. Lindblad. On the generators of quantum dynamical semigroups. Communications in Mathematical Physics 48 (1976), 119.

[243] V. Gorini, A. Kossakowski, E.C.G. Sudarshan. Completely positive dynamical semigroups of N-level systems. Journal of Mathematical Physics 17 (1976), 821.

[244] A.A. Andrianov, R. Tarrach, J. Taron. Neutral Kaons in Medium: Decoherence Effects. Physics Letters B 507 (2001), 200. См. также препринт arxiv.org/abs/hep-ph/0010276

[245] S. Hossenfelder. Comment on the black hole firewall. Препринт arxiv.org/abs/1210.5317

[246] S.D.H. Hsu. Macroscopic superpositions and black hole unitarity. Препринт arxiv.org/abs/1302.0451

[247] J.D. Bekenstein. Generalized second law of thermodynamics in black-hole physics. Physical Review D 9 (1974), 3292.