University of Gdańsk
Institute of Theoretical Physics and Astrophysics

The Gravitating Universe

Dissertation submitted for the degree of

Doctor of Philosophy

by

Mehraveh Nikjoo

Supervisor: Dr hab. Thomas Zlosnik

Gdańsk

June 16, 2026

Declaration

I hereby affirm that no part of this thesis has been submitted for the attainment of any other degree or professional qualification at this University or at any other institution. This thesis has been authored by me, and the research and findings contained herein are entirely my own, except where explicit acknowledgment has been made.

Mehraveh Nikjoo

To Mom and Dad

دل گرچه درین بادیه بسیار شتافت
یک موی ندانست و بسی موی شکافت
گرچه ز دلم هزار خورشید بتافت
آخر به کمال ذره‌ای راه نیافت

Ibn Sina —

Acknowledgments

In undertaking an odyssey of any kind, there comes a moment of dénouement when travelers find themselves compelled to reflect on their journey, even though, in scholarly matters, a final resolution remains inconceivably far from reached. The path that brought me to my doctoral studies was perhaps, at times, non-linear, yet it instilled in me a deep appreciation for all intellectual pursuits, and I consider myself incredibly fortunate to have encountered many individuals along the way who have enkindled my curiosity time and again. The preceding four years of my Ph.D. have continued this tradition, becoming one of the select exceptions to my usual desire for change. This consistency was anchored by the invaluable counsel of my supervisor, Dr hab. Thomas Zlosnik. True mentorship, I have found, extends beyond deciphering the academic difficulties of a field, and perhaps even more so in inspiring me towards broader horizons. Nor does my gratitude end there. I also wish to extend my unfeigned thanks to Prof. dr hab. Paweł Horodecki, whose support during the early stages of my doctoral studies provided a wellspring of encouragement. Equally, I would like to impart my warmest thanks to the entire community of the Theoretical Physics and Astrophysics Institute (IFTIA), which has been incredibly welcoming. Moreover, I am deeply appreciative to Mgr Małgorzata Szczekocka, Gosia, whose warmth and kindness made the foreign feel familiar. I cannot thank you enough for your boundless help and for being a pillar of positivity.

Beyond my home institute, my academic path led me to unexplored territories, and I am immensely grateful to those who made my visit to their domain possible. Firstly, I am sincerely thankful to Prof. Pedro Ferreira and Prof. David Alonso for hosting my visit to the University of Oxford and for the stimulating project they proposed. Secondly, I greatly appreciate the insights of Prof. Joe Zuntz, who made my visit to the University of Edinburgh possible and crafted an engaging and rewarding project. Finally, I wish to convey my genuine gratitude to Prof. Niayesh Afshordi, who hosted my visit to Perimeter Institute.

Furthermore, I offer my heartfelt thanks to two individuals for their invaluable advice and support throughout my scholarly endeavors: my master's supervisors, Prof. Roya Mohayaee and Prof. Jacques Colin.

Past the frontiers of the university, there stand those dear to my heart who have been a steadfast presence through all my days. To my fantastically strange and unusual family, who have been a catalyst for my creativity; to mom, whose enchanted gaze has always revealed extraordinary hidden in the mundane, to paint my world with magic; and to dad, my forever debate partner who always emboldened me to keep up the challenge, I can never say how grateful I am.

Unquestionably, we meet a small infinity of souls who shape our path. To those dear friends who have journeyed with me and to those I have yet to meet: I might write their story another time, but I am no less grateful.

Abstract

This thesis investigates gravitational theory and cosmological observables across theoretical, phenomenological, and empirical regimes.

The first part develops the Hamiltonian formulation of a class of gravitational theories that may be interpreted as spontaneously broken gauge theories of the complexified Lorentz group. In this framework, the gravitational field is described by a gauge field in the Lie algebra of $SO(1,3)_{\mathbb{C}}$ and by a Higgs-like field in its fundamental representation. The analysis shows that the number of propagating degrees of freedom depends crucially on the parameter controlling the relative weighting of the self-dual and anti-self-dual sectors. For nonzero values of this parameter, the theory propagates three complex degrees of freedom on general backgrounds, while the purely imaginary values $\pm i$ recover an extension of General Relativity in a symmetry-broken regime. By contrast, the null value yields no local degrees of freedom. Thus, in this class of models, the existence of local gravitational degrees of freedom is tied to chiral asymmetry in the gravitational sector.

An inflationary model motivated by left-right symmetric extensions of the matter sector and by the first-order formulation of gravity is then considered. The enlarged Higgs sector permits non-minimal couplings to gravity, giving rise to effective inflationary dynamics after reduction to a homogeneous and isotropic cosmological background. The background dynamics are implemented numerically, allowing the viable parameter space to be explored and the associated inflationary observables to be analyzed. Particular attention is given to the role of the Barbero–Immirzi parameter, which may take either real or imaginary values in this framework. The exploration of the model indicates that, within the viable region, variations of the effective Barbero–Immirzi parameter have no statistically significant direct impact on the scalar observables.

The empirical chapters first study weak gravitational lensing by photometric density ridges. The analysis reports a high-significance detection of this effect in Dark Energy Survey Year 3 data, although not independently of galaxy–galaxy lensing. It further develops improvements to the subspace-constrained mean-shift (SCMS) algorithm used to locate ridges efficiently at survey scale and investigates the dependence of the lensing signal on cosmological and algorithmic parameters in simulations. The signal is found to depend primarily on $S_8 = \sigma_8(\Omega_m/0.3)^{1/2}$, while further methodological improvements are required before precision parameter estimation can be performed.

Finally, an examination of tomographic constraints on the astrophysical stochastic gravitational-wave background through cross-correlations between gravitational-wave intensity maps and galaxy surveys is performed. Using large-scale SGWB maps constructed from the first three LIGO–Virgo observing runs and several tomographic galaxy samples, the analysis searches for correlations tracing the astrophysical production of unresolved gravitational waves across cosmic time. No statistically significant galaxy–SGWB cross-correlation is detected with current data. Nevertheless, the tomographic structure of the analysis allows upper bounds to be placed on the bias-weighted production rate density of gravitational waves by astrophysical sources, commonly parametrized by $\langle b\dot{\Omega}_{\mathrm{GW}}\rangle$.

Abstrakt

Niniejsza rozprawa bada teorię grawitacji oraz obserwable kosmologiczne w reżimach teoretycznym, fenomenologicznym i empirycznym.

Pierwsza część rozwija sformułowanie hamiltonowskie klasy teorii grawitacji, które można interpretować jako teorie cechowania zespolonej grupy Lorentza ze spontanicznym łamaniem symetrii. W tym formalizmie pole grawitacyjne opisane jest przez pole cechowania przyjmujące wartości w algebrze Liego grupy $SO(1,3)_{\mathbb{C}}$ oraz przez pole typu Higgsa przyjmujące wartości w jej reprezentacji fundamentalnej. Analiza wykazuje, że liczba propagujących się stopni swobody zależy zasadniczo od parametru kontrolującego względne ważenie sektorów samo-dualnego i anty-samo-dualnego. Dla niezerowych wartości tego parametru teoria propaguje trzy zespolone stopnie swobody na ogólnych tłach, natomiast czysto urojone wartości $\pm i$ odtwarzają rozszerzenie Ogólnej Teorii Względności w reżimie złamanej symetrii. Dla wartości zerowej teoria nie posiada lokalnych stopni swobody. Zatem w tej klasie modeli istnienie lokalnych grawitacyjnych stopni swobody jest powiązane z chiralną asymetrią w sektorze grawitacyjnym.

Następnie rozważany jest model inflacyjny motywowany lewoprawosymetrycznymi rozszerzeniami sektora materii oraz pierwszorzędowym sformułowaniem grawitacji. Rozszerzony sektor Higgsa dopuszcza nieminimalne sprzężenia z grawitacją, prowadząc do efektywnej dynamiki inflacyjnej po redukcji do jednorodnego i izotropowego tła kosmologicznego. Ewolucja tła jest implementowana numerycznie, co pozwala na eksplorację dopuszczalnej przestrzeni parametrów oraz analizę związanych z nią obserwabli inflacyjnych. Szczególną uwagę poświęcono roli parametru Barbero–Immirziego, który w tym formalizmie może przyjmować zarówno wartości rzeczywiste, jak i urojone. Eksploracja modelu wykazuje, że wartość tego parametru nie wpływa w żaden statystycznie istotny sposób na rozkład a posteriori.

Rozdziały empiryczne poświęcone są najpierw słabemu soczewkowaniu grawitacyjnemu przez fotometryczne grzbiety gęstości. Analiza raportuje detekcję tego efektu o wysokiej istotności statystycznej w danych Dark Energy Survey Year 3, choć nie niezależnie od soczewkowania galaktyka–galaktyka. Rozwijane są ponadto udoskonalenia algorytmu przesunięcia średniej z ograniczeniem podprzestrzenią, stosowanego do wyznaczania grzbieni w skali przeglądu, a zależność sygnału soczewkującego od parametrów kosmologicznych i algorytmicznych badana jest w symulacjach. Stwierdzono, że sygnał zależy

przede wszystkim od $S_8 = \sigma_8(\Omega_m/0{,}3)^{1/2}$, zaś przed przeprowadzeniem precyzyjnego szacowania parametrów konieczne są dalsze ulepszenia metodologiczne.

Na koniec przeprowadzona zostaje analiza tomograficznych ograniczeń na astrofizyczne stochastyczne tło fal grawitacyjnych poprzez krzyżowe korelacje map intensywności fal grawitacyjnych z przeglądami galaktyk. Korzystając z wielkoskalowych map SGWB skonstruowanych na podstawie trzech pierwszych obserwacyjnych serii LIGO–Virgo oraz kilku tomograficznych próbek galaktyk, poszukujemy korelacji śladów astrofizycznej produkcji nierozwiązanych fal grawitacyjnych w czasie kosmicznym. Przy użyciu bieżących danych nie wykryto statystycznie istotnej korelacji krzyżowej galaktyki–SGWB. Niemniej tomograficzna natura analizy pozwala na wyznaczenie górnych granic na ważoną biasem gęstość tempa produkcji fal grawitacyjnych przez źródła astrofizyczne, powszechnie parametryzowaną przez $\langle b\dot{\Omega}_{\mathrm{GW}}\rangle$.

Contents

[1]The beginnings of all things are small —Cicero

List of Figures

List of Tables

Introduction

> "Before there was any earth or sea, before the canopy of heaven stretched overhead, Nature presented the same aspect the world over, that to which men have given the name of Chaos. This was a shapeless, uncoordinated mass, nothing but a weight of lifeless matter, whose ill-assorted elements were indiscriminately heaped together in one place. There was no sun, in those days, to provide the world with light, no crescent moon ever filling out her horns: the earth was not poised in the enveloping air, balanced there by its own weight, nor did the sea stretch out its arms along the margins of the shores. Although the elements of land and air and sea were there, the earth had no firmness, the water no fluidity, there was no brightness in the sky. Nothing had any lasting shape, but everything got in the way of everything else; for, within that one body, cold warred with hot, moist with dry, soft with hard, and light with heavy."
>
> — Ovid, *Metamorphoses*

Some of the greatest stories begin with a moment of unexpected change, and some with the anticipation of one. The Universe, it turned out, was a great storyteller. Throughout its many epochs, it has experienced numerous metamorphoses and, perhaps somewhat mysteriously, has remained rather kind to some of our earliest conceptions of primordial chaos.

Millennia have since elapsed, and we now stand in an era of unprecedented observational acuity. Our triumphant ability to capture some of the faintest emanations from the abyssal frontiers of space illuminates the path ahead while vindicating the claim that many of our foundational inquiries remain unresolved.

Among the most elusive of these unresolved elements is gravity. Although its cumulative effect is written most visibly across the largest scales of our Universe, at the level of fundamental description, its formulation remains among the least reconciled with the rest of modern physics. Within this domain of uncertainty emerges the towering achievement of General Relativity and its empirical success. However, its standard formulation makes gravity difficult to place alongside the other fundamental interactions, whose description is built on the language of gauge symmetry and internal degrees of freedom.

The work developed here approaches this concern from several connected directions. We begin by building the language required to reformulate gravity in terms of gauge symmetries [1], which reveal how different mathematical configurations may represent the same physical state. Broken symmetries, in turn, explain how distinct phases can emerge from a more unified underlying structure. Indeed, for gravity, the relevant symmetry structure is especially rich. At each spacetime point, we may introduce a local inertial frame. The Lorentz group governs the transformations between such frames, and its representation theory determines how matter and geometry can be coupled. Moreover, upon complexification, a richer chiral structure is uncovered. These sectors correspond, in a mathematical sense, to the left- and right-handed components familiar from spinor representations. Then, the Einstein-–Cartan formulation [2, 3] provides a natural bridge between metric gravity and gauge gravity, opening a gateway to a reformulation of General Relativity where the gravitational phase space can be described using connection variables closely related to those of Yang–Mills theory [4].

The question then arises as to whether the relative treatment of these chiral sectors changes the dynamical content of the theory. The Hamiltonian formulation provides a stringent means of addressing such questions [5]. Once spacetime is decomposed into spatial hypersurfaces evolving in time, gravitational dynamics acquire the form of a constrained system in which variables may represent physical degrees of freedom or gauge redundancies. This perspective will be treated in Chapter 2, where gravity is regarded as a spontaneously broken gauge theory of the complexified Lorentz group.

The same ideas are naturally extended toward cosmology. The early universe offers a landscape in which gravitational theory, high-energy physics, and symmetry breaking are most tightly interwoven. Herein, the principal focus lies upon inflation [6, 7]. Within this paradigm, a transitory span of precipitous expansion elucidates the homogeneity of the cosmos, whilst concurrently engendering fluctuations that shall pave the way for structure formation. Within this cosmological setting, Chapter 4 studies a left-right symmetric model where the enlarged Higgs sector permits non-minimal gravitational couplings to matter that generate effective inflationary dynamics.

The concluding portion of this work turns its gaze towards the observational exploration of gravitational phenomena. Two distinct empirical windows are considered in

Chapter 5, namely, weak gravitational lensing and the stochastic gravitational-wave background (SGWB) [8]. Weak lensing probes the projected distribution of matter through the coherent distortion of background galaxy shapes [9]. It is therefore directly sensitive to the geometry and growth of cosmic structure. Particular attention is given to weak lensing by photometric density ridges, which trace filamentary structures in the cosmic web. The stochastic gravitational-wave background, then, offers a different but equally important observational channel. Here, rather than resolving individual gravitational-wave events, we study the accumulated background produced by many unresolved sources across cosmic history. Cross-correlating this background with galaxy surveys would then serve to test whether gravitational-wave anisotropies trace the same large-scale structure as luminous matter.

As any dutiful student of the natural sciences, I would be remiss not to clarify my precise contribution to these endeavors.

In the first instance, my contribution to the Hamiltonian analysis project [10] consisted of the explicit calculation of the constraint structures and the analysis of their time evolution. I also considered general values of the coupling pair denoted (g_+, g_-) as a generalization of [11], which made it possible to examine how the constraint structure and number of degrees of freedom depend on the relative weighting of the self-dual and anti-self-dual sectors. More particularly, this led to the realization that the special model $g_+ = g_-$ contains extra symmetries.

In the inflationary model, my work focused on the implementation of the background dynamics, together with the numerical analysis and production of the associated phenomenological parameter space. Additionally, I deduced the background and tensor perturbation equations of motion. These findings are to be integrated into a forthcoming preprint.

Concerning the weak-lensing project [12], I performed the numerical construction, including the identification and labeling of density ridges, the computation of the corresponding shear profiles, calibrating relevant parameters, and examining their impact on the shear signal. Finally, my contribution to the stochastic gravitational-wave background project [13] was more limited, consisting mainly of checking the behavior of the measured galaxy–SGWB cross-spectra for different spectral indices and for different choices of SGWB and galaxy maps.

Chapter 1

Foundations of Gravitational Theory

"The beginning is perhaps more difficult than anything else, but keep heart, it will turn out all right."

— Vincent Van Gogh

Gravity asserts itself uniquely amongst fundamental theories. The patient reader may therefore regard the construct of the theory as standing as a monument to theoretical inscrutability, notwithstanding the author's best effort to present a coherent analytical framework.

The idea is this: while our perception of reality might be distorted, mathematical knowledge has often provided an anchor through its axiomatic structure, but paradoxical as it may sound, the essence of mathematics lies also in its freedom"[1]. As a simple exercise, were we to survey a common property of the majority of objects perceived throughout a day, we might conclude that their prevailing characteristic is their symmetry. While the breaking of symmetries, crucial for phenomena such as the Higgs mechanism, might be the reason we exist, their upholding ensures the stability of our Universe. This follows from laws of physics being the same in space and in time, and thus, exhibiting symmetries that we will later on identify as translations and rotations. Viewed through the lens of symmetry, unexpected patterns have guided some of the key discoveries, especially in particle physics [2]. This inspires the hope that the same stroke of insight may await gravity.

[1]Perhaps fittingly, George Cantor, the mind behind the theory of infinity, championed this.

[2]The discovery of quarks, for instance, by Murray Gell-Mann. Interestingly, the name quark was borrowed from James Joyce's Finnegan's Wake, whose perusal might be considered on par with deciphering gravity!

1.1 Lorentz Group Complexification and Representations

The core ideas of this work rely heavily on a gauge theory description of gravity. To formulate the underlying theory, we examine the Lorentz group $SO(1,3)$ containing matrices of unit determinant that preserve the Minkowski metric. This group governs the transformation between local inertial frames, and it so happens that its action on the metric, connection, and curvature forms underlies approaches to gravitational dynamics.

Moreover, upon complexification, the Lorentz group's Lie algebra is enriched as it becomes isomorphic to the direct sum of two $sl(2,\mathbb{C})$ algebras, which enables a natural decomposition of fields into self-dual and anti-self-dual components. This chiral splitting plays an important role in the construction of the Ashtekar variables [14], which in turn are preliminary to potential parity-sensitive gravitational actions.

For instance, models studied later in this thesis include complexified gauge gravity with Higgs-like fields and left-right symmetric inflationary scenarios, where different representations introduce different available degrees of freedom.

This section, therefore, lays the essential mathematical foundation by developing a representation of the Lorentz group and the tools required for the symmetry-breaking mechanism that follows.

Additionally, fermionic couplings are to be allowed. We will then see that it will become necessary to build an understanding of spinors. In fact, we can think of the latter as the fundamental particle-like building blocks because other objects can be constructed from tensor products of their spins.

Spinors are examples of curious objects that came about independently within physics and mathematics. From the physics side, they were inspired by the behavior of particles such as electrons, for which the degrees of freedom seemed to double in the presence of an electromagnetic field, most famously revealed by the Stern–Gerlach experiment [15]. More generally, it seemed that many fundamental fermionic particles were naturally represented by pairs of complex numbers. From a standard mathematical perspective, many familiar geometric objects transform tensorially under changes of frame or coordinates, as in the case of vectors, the metric tensor, or curvature tensors. Now, if we investigate rotations in the most general context, one answer points toward a vector representation, but there are also other objects represented by a pair of complex numbers that can be consistently rotated but cannot be built from vectors.

A spinor, therefore, is not a tensor in the usual sense. Spatial vectors transform under ordinary linear representations of special orthogonal groups such as $SO(3)$, whereas spinors transform linearly under their double cover. With this view in mind, perhaps the most natural starting point would be the 2×2 special unitary group $SU(2)$. We shall then build the isomorphism that provides the link with the special orthogonal groups and later on, the Lorentz group.

1.1.1 The group $SU(2)$

The special unitary group of degree two is defined as

$$SU(2) = \{\, U \in GL(2, \mathbb{C}) \mid U^\dagger U = \mathbb{I},\ \det U = 1 \,\}. \tag{1.1}$$

It consists of complex 2×2 matrices that are unitary and unimodular.

Every element of $SU(2)$ can be written in the form

$$U = a_0\, \mathbb{I} + i\, a_i\, \sigma^i, \tag{1.2}$$

where $a_0, a_i \in \mathbb{R}$ satisfy

$$a_0^2 + a_1^2 + a_2^2 + a_3^2 = 1. \tag{1.3}$$

This identifies $SU(2)$ topologically with the three-sphere S^3.

We then introduce the Pauli spinor, one of the main objects of interest,

$$\psi = \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix}. \tag{1.4}$$

which transforms under $SU(2)$ as

$$\psi \mapsto U\psi. \tag{1.5}$$

This representation captures the behavior of spin-$\frac{1}{2}$ particles under spatial rotations. The generators of $SU(2)$ may be taken to be the Pauli matrices.

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.6}$$

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{1.7}$$

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.8}$$

They satisfy the following properties

1. Hermiticity:

$$\sigma_i^\dagger = \sigma_i \tag{1.9}$$

2. Tracelessness:
$$\mathrm{tr}(\sigma_i) = 0 \tag{1.10}$$

3. Anticommutation relations:
$$\{\sigma_i, \sigma_j\} = 2\delta_{ij}\mathbb{I} \tag{1.11}$$

4. Commutation relations:
$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \tag{1.12}$$

5. Squaring to identity:
$$\sigma_i^2 = \mathbb{I} \tag{1.13}$$

We may build a bridge to the real vectors by noting that any three-dimensional real vector v^i can be represented as a Hermitian 2×2 matrix.

$$v = v^i \sigma_i \tag{1.14}$$

This establishes a bijection between $\mathbb{R}^3$ and the space of traceless Hermitian matrices by introducing the Infeld–van der Waerden symbols [16], $\sigma^i_{AA'}$, providing the following isomorphism

$$\mathbb{R}^3 \longrightarrow \mathbb{C}^2 \otimes \overline{\mathbb{C}}^2 \tag{1.15}$$

Thus, the real vector can be formulated as

$$v^{AA'} = v^i \sigma_i^{AA'} \tag{1.16}$$

expressing a spatial vector as a bispinor. We conclude that any vector representation of $SO(3)$, denoting the rotation group in $\mathbb{R}^3$, can then be constructed from the tensor product of two fundamental representations of $SU(2)$.

1.1.2 The group $SU(2)$ as the double cover of the group $SO(3)$

The special orthogonal group in three dimensions is defined as

$$SO(3) = \{\, R \in GL(3, \mathbb{R}) \mid R^T R = \mathbb{I},\ \det R = 1 \,\}. \tag{1.17}$$

This group consists of all real 3×3 matrices - acting on spatial vectors - that are orthogonal and have determinant equal to one. It describes orientation-preserving rotations in three-dimensional Euclidean space. The generators of $SO(3)$ satisfy the commutation relations
$$[J_i, J_j] = \epsilon_{ijk} J_k \tag{1.18}$$

Following the line of reasoning from the previous section, we consider a vector written in matrix form.

$$v = v^i \sigma_i \tag{1.19}$$

Under an $SU(2)$ transformation, we have

$$v' = U v U^\dagger \tag{1.20}$$

Because U is unitary and unimodular, the transformed matrix remains Hermitian and traceless. We can show explicitly that this transformation corresponds to

$$v'^i = R^i{}_j \, v^j \tag{1.21}$$

where $R \in SO(3)$. However, both U and $-U$ induce the same rotation R. Thus, the mapping

$$SU(2) \longrightarrow SO(3) \tag{1.22}$$

is a two-to-one homomorphism. This property is commonly expressed as $SU(2)$ being a double cover of $SO(3)$.

The topological inspiration of this statement is grounded in the shape of these spaces. Using the parametrization for $U \in SU(2)$ encountered in Equation 1.2 and 1.3 we see the characteristic embedding of S^3 in $\mathbb{R}^4$. The rotation group $SO(3)$, by contrast, is obtained by identifying the antipodal points U and $-U$ and is given by the real projective 3-space. Hence

$$SO(3) \simeq SU(2)/\{\pm\mathbb{I}\} \simeq \mathbb{RP}^3. \tag{1.23}$$

This behavior is shown in Figure 1.1.

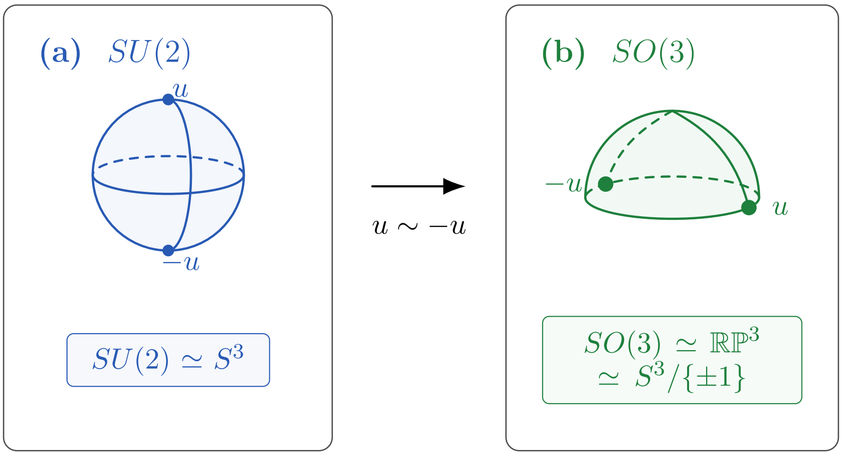


Figure 1.1: Schematic topological interpretation of the double cover $SU(2) \to SO(3)$. The group $SU(2)$ is topologically S^3, while $SO(3)$ is obtained by identifying antipodal points, so that $SO(3) \simeq \mathbb{RP}^3 \simeq S^3/\{\pm 1\}$.

This can also explain the doubling of the angle relating physical states and quantum states, where we get the characteristic spinorial behavior, i.e., a 2π rotation changes the sign of a spinor, while a 4π rotation restores the original state.

Up to this point, spinors have been introduced concretely as two-component complex column vectors acted upon by $SU(2)$ matrices. While operationally sufficient, this description does not explain what a spinor is in intrinsic algebraic terms. The mathematical language appropriate for this discussion requires us to define rings, projectors, and ideals. An associative ring $\mathcal{R}$ is a set equipped with addition and multiplication satisfying the following properties.

(i) $(\mathcal{R}, +)$ is an abelian group.

(ii) Multiplication is associative.

(iii) Multiplication distributes over addition.

The algebra of complex 2×2 matrices forms such a ring, containing the identity element and being closed under multiplication.

A subset $I \subset \mathcal{R}$ is called a left ideal if

$$ri \in I \quad \text{for all } r \in \mathcal{R},\ i \in I \tag{1.24}$$

Expressing that a left ideal is preserved under left multiplication by any element of the ring. The mechanism by which such ideals are constructed is provided by projection operators.

An element $P \in \mathcal{R}$ is called a projector if

$$P^2 = P \tag{1.25}$$

A projector is said to be primitive if it cannot be written as a sum of two nonzero orthogonal projectors.

Given any projector P, the set

$$\mathcal{R}P = \{\, rP \mid r \in \mathcal{R} \,\} \tag{1.26}$$

forms a left ideal. Indeed, for any $s \in \mathcal{R}$ and $rP \in RP$,

$$s(rP) = (sr)P \in \mathcal{R}P \tag{1.27}$$

so the subset is stable under left multiplication.

If P is primitive, then $\mathcal{R}P$ is minimal. In this way, primitive projectors carve out algebraically irreducible subspaces of the ring.

We may now return to the physical motivation. Within the matrix algebra generated by the Pauli matrices, primitive projectors isolate minimal left ideals [17]. These minimal ideals are two-dimensional complex linear spaces embedded inside the full matrix algebra. These minimal left ideals are associated with spinors.

1.1.3 Lie groups and Lie algebras

From the standpoint of physics, the continuous nature of symmetries imposes a differential description, and, therefore, we rely on Lie groups to describe infinitesimal changes.

A Lie group $\mathcal{G}$ is a set equipped with both a group and a smooth manifold structure, such that the group operations are smooth maps.

$$m : \mathcal{G} \times \mathcal{G} \longrightarrow \mathcal{G}, \qquad m(g,h) = gh \tag{1.28}$$

$$\iota : \mathcal{G} \longrightarrow \mathcal{G}, \qquad \iota(g) = g^{-1}. \tag{1.29}$$

We may now define the Lie algebra associated with the groups $SU(2)$ and $SO(3)$. The Lie algebra $\mathfrak{su}(2)$ is defined as the set of 2×2 complex matrices X satisfying

$$X^\dagger = -X, \qquad \mathrm{tr}(X) = 0 \tag{1.30}$$

A standard basis is then given by

$$T_i := -\frac{i}{2}\,\sigma_i \qquad (i = 1, 2, 3) \tag{1.31}$$

Satisfying the commutation relation

$$[\sigma_i, \sigma_j] = 2i\,\epsilon_{ijk}\,\sigma_k \tag{1.32}$$

Similarly, the Lie algebra of the rotation group is

$$\mathfrak{so}(3) = \{\, A \in M_3(\mathbb{R}) \mid A^T = -A \,\} \tag{1.33}$$

Where the generators $\{J_i\}$, satisfy the commutation relation $[J_i, J_j] = \epsilon_{ijk}\,J_k$ mentioned previously. Here we expectably note the similarity of the commutation expressions between $\mathfrak{su}(2)$ and $\mathfrak{so}(3)$, which reflects the double cover property of $SU(2)$.

Next, we aim to understand how abstract group elements, as linear transformations, act on vector spaces by introducing the concept of group representations.

Let G be a Lie group. A representation of G on an n-dimensional complex vector space is a homomorphism

$$\rho : G \longrightarrow GL(n, \mathbb{C}). \tag{1.34}$$

It satisfies the property

$$\rho(g_1 g_2) = \rho(g_1)\, \rho(g_2), \qquad g_1, g_2 \in G. \tag{1.35}$$

where we obtain

$$\rho(e) = \mathbb{I}_n, \tag{1.36}$$

since the identity element of the group can be rewritten as

$$\rho(e) = \rho(gg^{-1}) = \rho(g)\, \rho(g^{-1}). \tag{1.37}$$

It follows that

$$\rho(g^{-1}) = \rho(g)^{-1}. \tag{1.38}$$

Thus, a representation preserves the group multiplication structure. Additionally, we will define a Lie algebra representation. Given a Lie group representation ρ, differentiation at the identity produces a corresponding representation of the Lie algebra.

Let $\mathfrak{g} = T_e G$ be the Lie algebra of G. A Lie algebra representation is a linear map

$$\pi : \mathfrak{g} \longrightarrow \mathrm{Mat}(n, \mathbb{C}) \tag{1.39}$$

such that

$$\pi([X, Y]) = [\, \pi(X),\, \pi(Y)\,], \tag{1.40}$$

for all $X, Y \in \mathfrak{g}$, ensuring that the algebraic structure of infinitesimal generators is preserved. We may now turn our attention to $\mathfrak{su}(2)$, whose generators are denoted by

$$\mathcal{Y}_x, \qquad \mathcal{Y}_y, \qquad \mathcal{Y}_z.$$

Physically, these correspond to infinitesimal rotations about the x, y, and z axes, respectively, and they satisfy, in quantum-mechanical convention, the commutation relations

$$[\mathcal{Y}_x, \mathcal{Y}_y] = i\mathcal{Y}_z, \qquad [\mathcal{Y}_y, \mathcal{Y}_z] = i\mathcal{Y}_x, \qquad [\mathcal{Y}_z, \mathcal{Y}_x] = i\mathcal{Y}_y. \tag{1.41}$$

These relations encode the non-abelian structure of spatial rotations at the infinitesimal level.

The considerations in subsequent sections prompt us to introduce the combinations

$$\mathcal{Y}_{\pm} = i\,\mathcal{Y}_x \mp \mathcal{Y}_y. \tag{1.42}$$

These are referred to as ladder operators, where we verify that

$$[\mathcal{Y}_z,\, \mathcal{Y}_{\pm}] = \pm\mathcal{Y}_{\pm}. \tag{1.43}$$

The operators $\mathcal{Y}_{\pm}$ raise or lower eigenvalues of $\mathcal{Y}_z$. This structure allows an orderly construction of states in any representation.

Additionally, we can define the quadratic Casimir operator

$$\mathcal{Y}^2 = \mathcal{Y}_x^2 + \mathcal{Y}_y^2 + \mathcal{Y}_z^2. \tag{1.44}$$

Although the Lie algebra itself is defined only through commutators, and more specifically, $\mathcal{Y}^2$ is not an element of the Lie algebra, this quadratic combination has a distinguished property, namely, it commutes with all generators.

$$[\mathcal{Y}^2,\, \mathcal{Y}_i] = 0, \qquad i = x, y, z. \tag{1.45}$$

It is important to emphasize that in physical applications, the generators are taken to be Hermitian:

$$\mathcal{Y}_i^\dagger = \mathcal{Y}_i. \tag{1.46}$$

which in turn ensures their eigenvalues are real, and eigenvectors corresponding to distinct eigenvalues are orthogonal.

Using the ladder operators, we find the following identities

$$\mathcal{Y}_+\mathcal{Y}_- = \mathcal{Y}^2 - \mathcal{Y}_z^2 - \mathcal{Y}_z, \tag{1.47}$$

$$\mathcal{Y}_-\mathcal{Y}_+ = \mathcal{Y}^2 - \mathcal{Y}_z^2 + \mathcal{Y}_z. \tag{1.48}$$

Let $|j, m\rangle$ denote simultaneous eigenvectors of $\mathcal{Y}^2$ and $\mathcal{Y}_z$:

$$\mathcal{Y}^2|j, m\rangle = j(j+1)\,|j, m\rangle, \tag{1.49}$$

$$\mathcal{Y}_z|j, m\rangle = m\,|j, m\rangle. \tag{1.50}$$

Then the ladder operators act as

$$\mathcal{Y}_+|j, m\rangle = \sqrt{j(j+1) - m(m+1)}\;\;|j, m+1\rangle, \tag{1.51}$$

$$\mathcal{Y}_-|j, m\rangle = \sqrt{j(j+1) - m(m-1)}\;\;|j, m-1\rangle. \tag{1.52}$$

The existence of a highest weight state implies that the ladder must end. Consequently, we obtain

$$m = -j,\, -j+1,\, \dots,\, j. \tag{1.53}$$

Therefore the dimension of the representation is given by

$$n = 2j + 1. \tag{1.54}$$

1.1.4 Lorentz group and transformations

To extend the preceding discussion beyond purely spatial rotations, we must incorporate the relevant kinematical structure, i.e., Minkowski spacetime $\mathbb{R}^{1,3}$, equipped with a non-degenerate bilinear form of signature $(-,\, +,\, +,\, +)$.

We consider Minkowski spacetime endowed with the metric

$$\eta_{\mu\nu} = \mathrm{diag}(-1,\, +1,\, +1,\, +1). \tag{1.55}$$

For a four-vector $x^\mu = (ct,\, x,\, y,\, z)$ with t, x, y, z inertial coordinates we define the invariant interval as

$$s^2 = \eta_{\mu\nu}\, x^\mu x^\nu = -c^2 t^2 + x^2 + y^2 + z^2. \tag{1.56}$$

In this context, a Lorentz transformation is a linear transformation

$$x^\mu \;\mapsto\; x'^\mu = \Lambda^\mu{}_\nu\, x^\nu, \tag{1.57}$$

such that the interval, s, remains invariant:

$$s'^2 = s^2. \tag{1.58}$$

In matrix notation, the interval can be represented as

$$s^2 = x^T \eta\, x \tag{1.59}$$

satisfying the following transformation

$$s'^2 = x'^T \eta\, x' = x^T \Lambda^T \eta\, \Lambda\, x. \tag{1.60}$$

where requiring invariance for all x implies

$$\Lambda^T \eta\, \Lambda = \eta. \tag{1.61}$$

This condition defines the group $O(1,3)$.

We may now restrict the study to the connected component with

$$\det \Lambda = 1, \qquad \Lambda^0{}_0 > 0, \tag{1.62}$$

yielding the proper orthochronous Lorentz group $SO(1,3)$. Once again, we follow the same logical structure as the previous sections. We proceed to the Lie algebra by considering a one-parameter family

$$\Lambda(\epsilon) = \mathbb{I} + \epsilon M + O(\epsilon^2). \tag{1.63}$$

substituting into the defining condition,

$$(\mathbb{I} + \epsilon M^T)\, \eta\, (\mathbb{I} + \epsilon M) = \eta, \tag{1.64}$$

and keeping only first-order terms gives

$$M^T \eta + \eta M = 0. \tag{1.65}$$

This provides the defining condition for the Lorentz algebra:

$$M^T = -\, \eta\, M\, \eta^{-1}. \tag{1.66}$$

In index notation, we emphasize the antisymmetric property of the matrix M

$$M_{\mu\nu} = -\, M_{\nu\mu}. \tag{1.67}$$

Given the antisymmetric property of the generators in their spacetime indices, and since $\mu, \nu = 0, 1, 2, 3$, it follows that there are six independent components, corresponding to the six independent antisymmetric index pairs.

To interpret these generators physically, it is useful to separate time and space components. Writing spatial indices as $i, j = 1, 2, 3$, the generators decompose into two distinct classes: Firstly, the purely spatial components M_{ij}. Secondly, the mixed components M_{0i}.

The components M_{ij} generate infinitesimal transformations that leave the time coordinate invariant while rotating the spatial coordinates among themselves. Explicitly, an infinitesimal transformation generated by M_{ij} acts only within the spatial 3×3 block of the vector representation. In that block, these generators are antisymmetric and traceless, and they satisfy the same commutation relations as the generators of $\mathfrak{so}(3)$. They therefore correspond to spatial rotations.

The components, M_{0i}, by contrast, generate transformations that mix one temporal and one spatial coordinate. In the vector representation, these generators occupy the off-diagonal blocks connecting the time direction with spatial directions. Their structure is such that they preserve the Minkowski metric but do not preserve spatial distances independently. These generators correspond to Lorentz boosts.

Consequently, spatial rotations act within spacelike hypersurfaces and preserve Euclidean spatial geometry. Boosts, on the other hand, mix temporal and spatial directions in a manner consistent with the invariance of the spacetime interval.

It is important to emphasize that although the rotation generators form a closed subalgebra isomorphic to $\mathfrak{so}(3)$, the boost generators do not form a subalgebra on their own. Their commutator yields a rotation generator, reflecting the intrinsically non-Euclidean structure of Lorentzian geometry. We will re-encounter this property in our discussion on complexification and chiral decomposition of the Lorentz algebra.

Having introduced the Lorentz group in its standard vector form, it is now useful to reformulate its action in a language more naturally adapted to spinorial objects. The essential observation is that a Minkowski four-vector may be represented equivalently by a 2×2 Hermitian matrix. It will then follow that the group $SL(2, \mathbb{C})$ acts as the double cover of the group $SO(1, 3)$ in a manner analogous to the previous sections.

Let $x^\mu = (ct,\, x,\, y,\, z)$ be a real four-vector in Minkowski spacetime. To x^μ, we associate the following Weyl vector

$$W := x^\mu \sigma_\mu = \begin{pmatrix} ct + z & x - iy \\ x + iy & ct - z \end{pmatrix}. \tag{1.68}$$

where we have

$$\sigma_\mu = (\mathbb{I},\, \sigma_i). \tag{1.69}$$

Since the coefficients x^μ are real, W is Hermitian. We note that any Hermitian 2×2 matrix can be written uniquely in this form. Thus, Minkowski vectors are in one-to-one correspondence with Hermitian 2×2 matrices. Moreover, we have the following results concerning the determinant of the Weyl vector

$$\det W = (ct + z)(ct - z) - (x - iy)(x + iy) = c^2t^2 - x^2 - y^2 - z^2, \tag{1.70}$$

written in terms of the metric

$$\det W = -\eta_{\mu\nu}\, x^\mu x^\nu. \tag{1.71}$$

We see that the Minkowski norm is represented by $-\det W$.

We shall first restrict our attention to the manner in which W transforms under the

$SU(2)$. Considering $U \in SU(2)$, then

$$W \;\mapsto\; UWU^{\dagger}. \tag{1.72}$$

Because U is unitary,

$$UU^{\dagger} = \mathbb{I}, \tag{1.73}$$

we obtain

$$UWU^{\dagger} = U\Big(ct\,\mathbb{I} + x^{i}\sigma_{i}\Big)\,U^{\dagger} = ct\,\mathbb{I} + x^{i}\,U\sigma_{i}U^{\dagger}. \tag{1.74}$$

We note here that the coefficient of the identity matrix is unchanged, whereas the traceless part transforms among itself. In this way, the temporal component behaves as a scalar under spatial rotations, while the spatial part transforms as a three-vector.

The general Lorentzian situation, however, is slightly different. Let now $L \in SL(2,\mathbb{C})$ with

$$SL(2,\mathbb{C}) = \{\, L \in \mathrm{Mat}(2,\mathbb{C}) \mid \det L = 1 \,\}, \tag{1.75}$$

The defining property of $SL(2,\mathbb{C})$ is unimodularity, but we do not have the unitarity property in this case. This means, in general, we cannot write,

$$LL^{\dagger} \neq \mathbb{I}. \tag{1.76}$$

Thus, in the case of the transformation

$$W \;\mapsto\; W' = LWL^{\dagger}, \tag{1.77}$$

as W is Hermitian, W' is also Hermitian. Furthermore,

$$\det W' = \det\Big(LWL^{\dagger}\Big) = \det(L)\;\det(W)\;\det\Big(L^{\dagger}\Big) = \det(W), \tag{1.78}$$

because $\det L = 1$. Hence, the Minkowski norm is preserved. This shows that L induces a Lorentz transformation on the associated four-vector. Additionally, the absence of unitarity yields

$$W' = L\Big(ct\,\mathbb{I} + x^{i}\sigma_{i}\Big)\,L^{\dagger} = ct\,LL^{\dagger} + x^{i}L\sigma_{i}L^{\dagger}. \tag{1.79}$$

Since in general $LL^{\dagger} \neq \mathbb{I}$, the coefficient of the identity matrix, i.e., the temporal component, cannot remain isolated. The temporal and spatial parts, therefore, mix, and the Lorentz transformation does not preserve the separation between time and space components.

In order to solve the mixing problem, we will need to change perspective and treat the 4-vector as a single object rather than decomposing it. Let

$$L = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \qquad \alpha\delta - \beta\gamma = 1. \tag{1.80}$$

generate the transformation of the Weyl vector, $\widetilde{W} = LWL^\dagger$. This yields four real linear relations between $(\tilde{t},\, \tilde{x},\, \tilde{y},\, \tilde{z})$ and $(t,\, x,\, y,\, z)$, thereby defining a real 4×4 matrix $\Lambda(L)$

$$\begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & i & -i & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} t+z \\ x-iy \\ x+iy \\ t-z \end{pmatrix}. \tag{1.81}$$

satisfying

$$\tilde{x}^\mu = \Lambda^\mu{}_\nu(L)\, x^\nu \tag{1.82}$$

As we mentioned before, the determinant of W is the Minkowski norm. Using $\det L = 1$, we find that

$$\det \widetilde{W} = \det\left(LWL^\dagger\right) = \det L\, \det W\, \det L^\dagger = \det W. \tag{1.83}$$

Therefore, the quadratic form $\eta_{\mu\nu}x^\mu x^\nu$ is preserved, and the induced matrix $\Lambda(L)$ satisfies

$$\Lambda(L)^T\, \eta\, \Lambda(L) = \eta. \tag{1.84}$$

Each $L \in SL(2,\mathbb{C})$ determines a Lorentz transformation. Since $SL(2,\mathbb{C})$ is connected and contains the identity, the image lies in the proper orthochronous Lorentz group $SO(1,3)$.

We therefore obtain a group homomorphism

$$\rho : SL(2,\mathbb{C}) \to SO(1,3), \qquad \rho(L) = \Lambda(L). \tag{1.85}$$

We note that the above mapping is not injective. Indeed,

$$(-L)\, W\, (-L)^\dagger = LWL^\dagger, \tag{1.86}$$

so L and $-L$ induce the same Lorentz transformation yielding the kernel

$$\ker \rho = \{\pm \mathbb{I}\}. \tag{1.87}$$

It follows that the map is two-to-one, and, therefore, $SL(2,\mathbb{C})$ is the double cover of

$SO(1,3)$

$$SO(1,3) \simeq SL(2,\mathbb{C})/\{\pm\mathbb{I}\}. \tag{1.88}$$

We may now clarify the structure of the generators. A one-parameter subgroup may be written as

$$L(\theta) = e^{\theta M}. \tag{1.89}$$

Using the identity $\det\left(e^{A}\right) = e^{\mathrm{tr}(A)}$, the condition $\det L(\theta) = 1$ implies

$$\mathrm{tr}(M) = 0. \tag{1.90}$$

Thus, the generators of $SL(2,\mathbb{C})$ are traceless. However, because there is no unitarity condition, they are not required to be anti-Hermitian, underlining the difference from $SU(2)$, which contributes to the description of boosts alongside rotations.

Furthermore, another crucial difference concerns the spin-$\frac{1}{2}$ representations. For $SU(2)$, there is a single nontrivial two-dimensional complex representation relevant to spin-$\frac{1}{2}$ rotations. For $SL(2,\mathbb{C})$, by contrast, there are two inequivalent two-dimensional spinor transformations. This is because the rotation group $SO(3)$ does not possess a faithful two-dimensional representation, but its double cover $SU(2)$ does. This group acts on two-component spinors, providing a unique nontrivial irreducible two-dimensional complex representation. If $U \in SU(2)$, a spinor $\psi \in \mathbb{C}^2$ transforms as

$$\psi \;\mapsto\; U\psi. \tag{1.91}$$

All spin-$\frac{1}{2}$ states transform according to this representation, and other representations may be obtained from the tensor products.

For the full Lorentz group, the situation proves different. Lorentz transformations can be represented through matrices $L \in SL(2,\mathbb{C})$ acting on a Hermitian matrix associated with a four-vector $W = x^{\mu}\sigma_{\mu}$. This construction encodes two distinct actions of $SL(2,\mathbb{C})$: one acting on the left and one acting on the right. These two actions lead to two inequivalent ways in which a two-component spinor may transform.

A left-handed Weyl spinor transforming according to the fundamental representation,

$$\psi_L \;\mapsto\; L\,\psi_L. \tag{1.92}$$

while the right-handed Weyl spinor, which transforms according to the conjugate representation,

$$\psi_R \;\mapsto\; (L^{\dagger})^{-1}\psi_R. \tag{1.93}$$

These two representations would naturally coincide under the spatial rotations. When

$L \in SU(2)$, we have $L^\dagger = L^{-1}$, and therefore the two transformation laws become identical. For a general Lorentz transformation $L \in SL(2, \mathbb{C})$, however, the matrices L and $(L^\dagger)^{-1}$ are distinct, and the corresponding spinor transformations are inequivalent. This leads to the so-called chiral structure of the Lorentz group. Particularly, in the language of representations, we obtain two inequivalent two-dimensional spinor representations,

$$\left(\tfrac{1}{2}, 0\right) \qquad \text{and} \qquad \left(0, \tfrac{1}{2}\right),$$

corresponding respectively to left-handed and right-handed Weyl spinors. It is important to note that these two sectors are not related by any Lorentz transformations, unless one includes parity.

Parity reverses spatial coordinates,

$$(t,\, \mathbf{x}) \;\mapsto\; (t,\, -\mathbf{x}), \tag{1.94}$$

which leaves rotations unchanged but reverses the direction of boosts. This operation exchanges the two chiral sectors,

$$\left(\tfrac{1}{2}, 0\right) \;\longleftrightarrow\; \left(0, \tfrac{1}{2}\right). \tag{1.95}$$

For many physical applications, it is convenient to combine the two Weyl spinors into a single object called the Dirac spinor,

$$\Psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}. \tag{1.96}$$

which transforms block-diagonally under Lorentz transformation.

$$\Psi \;\mapsto\; \begin{pmatrix} L & 0 \\ 0 & (L^\dagger)^{-1} \end{pmatrix} \Psi. \tag{1.97}$$

The left- and right-chiral components may then be extracted using projection operators constructed from the matrix γ_5 taking the following block-diagonal form in the chiral, or

Weyl, representations [3]

$$\gamma^5 = \begin{pmatrix} -\mathbb{I}_2 & 0 \\ 0 & \mathbb{I}_2 \end{pmatrix}. \tag{1.98}$$

The left-handed and right-handed Weyl spinors are distinguished as the eigenspaces of γ^5 with eigenvalues -1 and $+1$, respectively. This motivates the definition of the chiral projection operators

$$P_L = \frac{1}{2}\left(1 - \gamma^5\right), \qquad P_R = \frac{1}{2}\left(1 + \gamma^5\right). \tag{1.99}$$

They obey

$$P_L^2 = P_L, \qquad P_R^2 = P_R, \qquad P_L P_R = 0, \qquad P_L + P_R = 1, \tag{1.100}$$

and therefore project onto complementary subspaces of the Dirac spinor. Acting on Ψ, we obtain

$$\Psi_L = P_L\Psi = \begin{pmatrix} \psi_L \\ 0 \end{pmatrix}, \qquad \Psi_R = P_R\Psi = \begin{pmatrix} 0 \\ \psi_R \end{pmatrix}. \tag{1.101}$$

Hence, the Dirac spinor decomposes into its left- and right-chiral components as

$$\Psi = \Psi_L + \Psi_R. \tag{1.102}$$

The Lie algebra $\mathfrak{sl}(2,\mathbb{C})$ admits the following decomposition

$$\mathfrak{sl}(2,\mathbb{C}) \;=\; \mathfrak{su}(2) \;\oplus\; i\,\mathfrak{su}(2) \tag{1.103}$$

where a qualitatively different structure emerges upon complexification. The complexification of $\mathfrak{su}(2)$ is defined by

$$\mathfrak{su}(2)_{\mathbb{C}} = \mathfrak{su}(2) \otimes_{\mathbb{R}} \mathbb{C}, \tag{1.104}$$

and is isomorphic to $\mathfrak{sl}(2,\mathbb{C})$ as a complex Lie algebra. Applying this procedure to the Lorentz algebra, we conclude that the complexified Lorentz algebra admits the direct sum decomposition

$$\mathfrak{so}(1,3)_{\mathbb{C}} \;\cong\; \mathfrak{su}(2)_{\mathbb{C}} \;\oplus\; \mathfrak{su}(2)_{\mathbb{C}}. \tag{1.105}$$

[3] For the interested reader, the matrices γ^μ are understood as a representation of the Clifford algebra,

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}\mathbb{I}_4,$$

and γ_5 is defined by

$$\gamma_5 := i\gamma^0\gamma^1\gamma^2\gamma^3.$$

Its explicit matrix form depends on the chosen representation of the Clifford algebra; in the chiral, or Weyl, representation it is diagonal and distinguishes the left- and right-handed Weyl subspaces.

By analogy with the ladder operators introduced in Equation 1.42, we may now introduce the combinations

$$J_i^{\pm} = \frac{1}{2}\Big(J_i \pm iK_i\Big), \tag{1.106}$$

where J_i are the rotation generators and K_i the boost generators. We find that they satisfy

$$[J_i^+,\, J_j^+] = \epsilon_{ijk}\, J_k^+, \qquad [J_i^-,\, J_j^-] = \epsilon_{ijk}\, J_k^-, \tag{1.107}$$

while

$$[J_i^+,\, J_j^-] = 0. \tag{1.108}$$

Thus, the generators J_i^+ and J_i^- define two independent copies of the $\mathfrak{su}(2)$ algebra that provides the chiral structure introduced previously. Representations of the Lorentz algebra are therefore labeled by a pair of spins $(j_+,\, j_-)$, corresponding to the two independent $\mathfrak{su}(2)$ factors.

The language we have used so far is more commonly used to describe objects in quantum mechanics. However, we may shift perspective to establish the gap between this formalism and objects of interest in the language of gravitational theories.

1.1.5 Hodge duality and the self-dual / anti-self-dual decomposition

The chiral structure of the Lorentz group admits an equivalent and highly useful tensorial formulation in terms of antisymmetric rank-two Lorentz tensors. This formulation will be of central importance later, when the spin connection and curvature are decomposed into chiral sectors. The appropriate starting point is the Hodge dual operator acting on antisymmetric pairs of internal Lorentz indices.

Let $F^{IJ} = -F^{JI}$ be an antisymmetric Lorentz tensor, where $I, J = 0, 1, 2, 3$ are internal indices raised and lowered with the Minkowski metric η_{IJ}. Since F^{IJ} is antisymmetric, it contains six independent components and may be regarded as an element of the space of Lorentz bivectors. On this space, we define the internal Hodge dual by

$$(\star F)^{IJ} := \frac{1}{2}\, \epsilon^{IJ}{}_{KL}\, F^{KL}, \tag{1.109}$$

where ϵ_{IJKL} is the totally antisymmetric Levi-Civita symbol, with convention $\epsilon_{0123} = +1$.

The Hodge dual maps antisymmetric tensors to antisymmetric tensors. In Lorentzian signature, applying it twice yields negative the identity on bivectors:

$$\star(\star F) = -\, F. \tag{1.110}$$

Equivalently,

$$\frac{1}{2}\,\epsilon^{IJ}{}_{KL}\,(\star F)^{KL} = -\,F^{IJ}. \tag{1.111}$$

Thus the operator $\star$ has eigenvalues $\pm i$ on the complexified space of bivectors. This shows that a real antisymmetric Lorentz tensor cannot be an eigenvector of $\star$, and that its decomposition arises only after complexification. Accordingly, we decompose F^{IJ} into self-dual and anti-self-dual parts,

$$F^{IJ} = F^{+\,IJ} + F^{-\,IJ}. \tag{1.112}$$

where

$$F^{\pm IJ} := \frac{1}{2}\Big(F^{IJ} \mp i\,(\star F)^{IJ}\Big) = \frac{1}{2}\Big(F^{IJ} \mp \frac{i}{2}\,\epsilon^{IJ}{}_{KL}\,F^{KL}\Big)\,. \tag{1.113}$$

By construction, these satisfy the eigenvalue equations

$$\star F^{\pm} = \pm i\,F^{\pm}, \tag{1.114}$$

or, in index notation,

$$\epsilon^{IJ}{}_{KL}\,F^{\pm KL} = \pm 2i\,F^{\pm IJ}. \tag{1.115}$$

This decomposition is analogous to the one encountered earlier at the level of the Lorentz algebra, where the complexified algebra split into two commuting sectors generated by Equation 1.106.

Here, the same chiral splitting is realized tensorially through the eigenspaces of the Hodge dual operator. The two descriptions are equivalent: the self-dual and anti-self-dual bivectors furnish the two chiral sectors of the complexified Lorentz algebra.

In representation theory terms, the fundamental Lorentz spinors correspond to the two inequivalent chiral sectors

$$\left(\tfrac{1}{2},0\right), \quad \left(0,\tfrac{1}{2}\right). \tag{1.116}$$

The tensor product of two left-handed spinors yields

$$\left(\tfrac{1}{2},0\right)\otimes\left(\tfrac{1}{2},0\right) = (0,0)\ \oplus\ (1,0). \tag{1.117}$$

while the tensor product of two right-handed spinors yields

$$(0,\tfrac{1}{2})\otimes(0,\tfrac{1}{2}) = (0,0)\oplus(0,1). \tag{1.118}$$

The symmetric traceless parts correspond to $(1,0)$ and $(0,1)$. They are the self-dual and anti-self-dual bivector sectors. Likewise, the mixed product

$$(\tfrac{1}{2},0)\otimes(0,\tfrac{1}{2}) = (\tfrac{1}{2},\tfrac{1}{2}) \tag{1.119}$$

(m, n)	Representation	Dimension	Interpretation
$(0, 0)$	Scalar	1	Invariant field
$\left(\frac{1}{2}, 0\right)$	Left Weyl spinor	2	Left-chiral spinor
$\left(0, \frac{1}{2}\right)$	Right Weyl spinor	2	Right-chiral spinor
$\left(\frac{1}{2}, \frac{1}{2}\right)$	Vector	4	4-vector x^μ
$(1, 0)$	Self-dual bivector	3	F^+_{IJ}
$(0, 1)$	Anti-self-dual bivector	3	F^-_{IJ}
$(1, 1)$	Traceless symmetric tensor	9	Rank-2 symmetric tensor

Table 1.1: Representations of the Lorentz group labeled by (m, n). The decomposition $(1, 0) \oplus (0, 1)$ corresponds to the splitting of antisymmetric tensors into self-dual and anti-self-dual parts.

corresponds to the Lorentz four-vector representation. We may gather the previous results in Table 1.1

1.2 Gauge Theory and Foundations of Einstein–Cartan Gravity

Having established the structure of the Lorentz group representation and its chiral decomposition, we may now proceed to a more geometrically oriented formulation.

In particular, the Einstein–Cartan description of gravity extends general relativity by incorporating objects denoted the tetrad field and the spin connection acting as independent variables. The tetrad relates spacetime directions to local Lorentz frames and reconstructs the metric, while the spin connection defines covariant differentiation and parallel transport within those frames.

This first-order formulation, based on the Palatini variational principle [18], incorporates torsion. This perspective places gravity on equal footing with Yang–Mills theories—but also offers technical advantages for canonical quantization and the study of matter coupling, particularly to fermions with intrinsic spin.

We begin this section by introducing the geometrical structure underlying Einstein–Cartan gravity.

1.2.1 Differential forms

Let x^μ denote a system of general local coordinates on a smooth manifold. The basic one-forms dx^μ provide a dual basis to the coordinate vector fields ∂_μ. A general one-form may therefore be written as

$$A = A_\mu \, dx^\mu. \tag{1.120}$$

This covector field plays the role of a gauge potential. It assigns to each tangent vector a scalar, and it transforms covariantly under changes of coordinates.

Next, we introduce the exterior derivative acting on differential forms as the map

$$d : \Omega^p(M) \longrightarrow \Omega^{p+1}(M), \tag{1.121}$$

which sends a p-form to a $(p+1)$-form. For a function f, regarded as a zero-form, it reduces to the ordinary differential,

$$df = \partial_\mu f \, dx^\mu. \tag{1.122}$$

For a general p-form

$$A = \frac{1}{p!} A_{\mu_1 \cdots \mu_p} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_p}, \tag{1.123}$$

the exterior derivative is

$$dA = \frac{1}{p!} \partial_\nu A_{\mu_1 \cdots \mu_p} dx^\nu \wedge dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_p}. \tag{1.124}$$

Or equivalently,

$$dA = \frac{1}{(p+1)!} (dA)_{\nu \mu_1 \cdots \mu_p} dx^\nu \wedge dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_p}, \tag{1.125}$$

with

$$(dA)_{\mu_0 \mu_1 \cdots \mu_p} = (p+1) \, \partial_{[\mu_0} A_{\mu_1 \cdots \mu_p]}. \tag{1.126}$$

where the operator d is pseudo Leibnizian, i.e., if $A \in \Omega^p(M)$ and $B \in \Omega^q(M)$, then

$$d(A \wedge B) = dA \wedge B + (-1)^p A \wedge dB. \tag{1.127}$$

The sign appears because d must pass through the p-form A before acting on B. Finally, the exterior derivative is nilpotent:

$$d^2 = 0. \tag{1.128}$$

This follows from the commutativity of partial derivatives, together with the antisymmetry of the wedge product. Applied to the one-form A,

$$A = A_\mu \, dx^\mu \tag{1.129}$$

with exterior derivative

$$dA = \partial_\nu A_\mu \, dx^\nu \wedge dx^\mu, \tag{1.130}$$

it yields the two-form

$$F = dA. \tag{1.131}$$

In coordinates, this becomes

$$F = d(A_\mu dx^\mu) = \partial_\nu A_\mu \, dx^\nu \wedge dx^\mu = \frac{1}{2}\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right) dx^\mu \wedge dx^\nu. \tag{1.132}$$

It is easy to see that the exterior derivative extracts the antisymmetric part of the first derivatives of the potential. Indeed, the antisymmetry of F is encoded in the wedge product of two linearly independent one-forms, which in turn defines an oriented area element, or bivector.

This point may be illustrated explicitly in two dimensions. Let

$$\mathbf{a} = a_1 e_1 + a_2 e_2, \qquad \mathbf{b} = b_1 e_1 + b_2 e_2, \tag{1.133}$$

where $\{e_1, e_2\}$ is a basis. Their wedge product is

$$\mathbf{a} \wedge \mathbf{b} = (a_1 b_2 - a_2 b_1)\, e_1 \wedge e_2. \tag{1.134}$$

The coefficient

$$a_1 b_2 - a_2 b_1 \tag{1.135}$$

is the determinant of the matrix whose columns are the components of $\mathbf{a}$ and $\mathbf{b}$,

$$\det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}, \tag{1.136}$$

thus, measuring the oriented area spanned by two vectors. More generally, in higher dimensions, antisymmetric products of basis one-forms encode oriented volume elements, and the determinant appears as the factor by which such oriented volumes scale under linear transformations.

Anticipating future analysis, we may define an operation that maps antisymmetric p-forms to antisymmetric $(n-p)$-forms.

$$\star : \Omega^p(M) \longrightarrow \Omega^{n-p}(M). \tag{1.137}$$

This is done using the Hodge dual operator defined above.

Let A be a p-form given by Equation 1.123 on an n-dimensional manifold, its Hodge dual $\star A$ is the $(n-p)$-form defined by

$$\star A = \frac{1}{(n-p)!} (\star A)_{\mu_{p+1}\cdots\mu_n} dx^{\mu_{p+1}} \wedge \cdots \wedge dx^{\mu_n}, \tag{1.138}$$

with components

$$(\star A)_{\mu_{p+1}\cdots\mu_n} = \frac{1}{p!}\epsilon_{\mu_1\cdots\mu_p\mu_{p+1}\cdots\mu_n} A^{\mu_1\cdots\mu_p}. \tag{1.139}$$

where the completely antisymmetric Levi–Civita symbol is defined as

$$\epsilon_{\mu_1\cdots\mu_n} = \begin{cases} +1 & \text{if } (\mu_1,\ldots,\mu_n) \text{ is an even permutation of } (1,\ldots,n), \\ -1 & \text{if } (\mu_1,\ldots,\mu_n) \text{ is an odd permutation of } (1,\ldots,n), \\ 0 & \text{otherwise.} \end{cases} \tag{1.140}$$

and the indices on A are raised using the inverse metric.

In subsequent sections, this property features in four-dimensional gravity, where curvature two-forms may be dualized into two-forms, introducing the self-dual and anti-self-dual decomposition for curvature and spin-connection forms.

The above observations are also instrumental for integration on curved manifolds, where, according to Stokes' theorem, for any region Ω with boundary $\partial\Omega$,

$$\int_\Omega dA = \int_{\partial\Omega} A. \tag{1.141}$$

Applied to the field-strength two-form, this gives [19]

$$\int_\Omega F = \int_\Omega dA = \oint_{\partial\Omega} A. \tag{1.142}$$

Thus, the flux of the field strength through a surface is equal to the circulation of the gauge potential around its boundary. We will see next that the metric proves a valid candidate as an integration measure.

On an n-dimensional manifold endowed with a metric $g_{\mu\nu}$, the invariant volume form is

$$\mathrm{vol} = \sqrt{|\det g_{\mu\nu}|}\, dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n. \tag{1.143}$$

The factor $\sqrt{|\det g_{\mu\nu}|}$ ensures that this object transforms appropriately under coordinate changes. Indeed, if

$$dx'^\mu = T^\mu{}_\nu\, dx^\nu, \tag{1.144}$$

then the wedge product of all coordinate differentials transforms by the determinant of the Jacobian matrix,

$$dx'^1 \wedge \cdots \wedge dx'^n = \det(T)\, dx^1 \wedge \cdots \wedge dx^n. \tag{1.145}$$

At the same time, the determinant of the metric transforms with the inverse square of the same Jacobian factor, so that the combination

$$\sqrt{|\det g_{\mu\nu}|}\, dx^1 \wedge \cdots \wedge dx^n \tag{1.146}$$

remains invariant.

Notably, in terms of the Levi–Civita symbol we have

$$\epsilon_{\mu_1 \cdots \mu_n} = \sqrt{|g|}\, \tilde{\epsilon}_{\mu_1 \cdots \mu_n}, \tag{1.147}$$

where $g = \det(g_{\mu\nu})$.

Analogously to Equation 1.145 if $T^{\mu}{}_{\nu}$ is a linear transformation, it may be expressed in terms of the Levi–Civita symbol as

$$\tilde{\epsilon}_{\mu_1 \cdots \mu_n} T^{\mu_1}{}_{\nu_1} \cdots T^{\mu_n}{}_{\nu_n} = (\det T)\, \tilde{\epsilon}_{\nu_1 \cdots \nu_n}. \tag{1.148}$$

The presence of the determinant factor shows that $\tilde{\epsilon}_{\mu_1 \cdots \mu_n}$ does not transform tensorially. Objects of this kind are called tensor densities. A tensor density transforms like a tensor up to an additional power of the determinant of the Jacobian, i.e., the density weight.

Given that the metric volume form is built by combining the coordinate volume element with the square root of the metric determinant, it transforms correctly under changes of coordinates and therefore provides the appropriate integration measure. In this way, the metric factor compensates for the density character of the coordinate volume element and gives an invariant notion of volume.

Accordingly, the Levi–Civita symbol behaves as a tensor density of weight -1 or $+1$, depending on convention and on whether covariant or contravariant indices are used. We will encounter densitized variables further ahead when discussing Ashtekar-type formulations.

To summarize briefly, a gauge potential is represented by a one-form, while its field strength is represented by the exterior derivative of that one-form, namely a two-form. We then discussed the antisymmetric structure of the wedge product, useful to define the field strengths and oriented integration. Finally, the volume form provides the invariant measure with respect to which physical action functionals are defined. In the present framework, the volume element is built from the co-tetrads, the gauge potential will be replaced by the spin connection, and the field strength by the curvature two-form.

1.2.2 The Einstein–Cartan formulation of gravity

The gauge-theory formulation of gravity hinges on the fact that the tangent space at each point of spacetime carries an action of the Lorentz group $SO(1,3)$. This reflects a freedom of choice in the local Lorentz frame, corresponding to transformations that leave the physical content unchanged. This results from decomposing the metric $g_{\mu\nu}$ in terms of co-tetrad (or vierbein) field, $e^I{}_\mu$, rather than positing it as a fundamental field from the outset. As a reminder, the Greek indices $\mu, \nu, \ldots = 0, 1, 2, 3$ refer to spacetime coordinates and capital Latin indices $I, J, \ldots = 0, 1, 2, 3$ refer to an internal Lorentz frame equipped with the Minkowski metric $\eta_{IJ} = \text{diag}(-1, +1, +1, +1)$.

More specifically, as seen in the previous section, the Weyl spinors of the Standard Model transform in the fundamental representations of the group $SL(2,\mathbb{C})$. Consequently, invariance of the spinorial Lagrangian under global $SL(2,\mathbb{C})$ transformations requires the introduction of the tetrad $e^I{}_\mu$, together with the identification of $\Lambda^I{}_J$ as the $SO(1,3)$ transformation corresponding to the given $SL(2,\mathbb{C})$ element.

The tetrad defines a local isomorphism between the spacetime tangent space and an internal Minkowski space, and the spacetime metric is reconstructed from it according to

$$g_{\mu\nu} = \eta_{IJ} e^I{}_\mu e^J{}_\nu. \tag{1.149}$$

Here, the invariance of the metric under local Lorentz rotations of the internal frame is shown with

$$e^I{}_\mu \longrightarrow \Lambda^I{}_J(x)\, e^J{}_\mu, \tag{1.150}$$

with $\Lambda^I{}_J(x) \in SO(1,3)$, $g_{\mu\nu}$ remains unchanged because Λ preserves η_{IJ}. This has been previously demonstrated in the section concerning the Lorentz group. Moreover, the expression in Equation 1.150 remains invariant when the associated Lorentz transformation Λ^{IJ} is allowed to depend on the spacetime point.

In order for spinorial actions to remain invariant under the corresponding local $SL(2,\mathbb{C})$ transformations, however, it becomes necessary to introduce an additional compensating field, namely the spin connection $\omega_\mu{}^{IJ}$, where $\omega_\mu{}^{IJ} = -\omega_\mu{}^{JI}$. The antisymmetry in I, J reflects the fact that $\omega_\mu{}^{IJ}$ takes values in the Lie algebra $\mathfrak{so}(1,3)$.

In differential-form notation,

$$\omega^{IJ} = \omega_\mu{}^{IJ}\, dx^\mu. \tag{1.151}$$

Its role is analogous to that of a Yang–Mills gauge field connection associated with local Lorentz symmetry and permits the construction of a covariant derivative acting on spinors. For a Lorentz vector V^I, we have

$$D_\mu V^I = \partial_\mu V^I + \omega_\mu{}^I{}_J\, V^J. \tag{1.152}$$

The spin connection transforms inhomogeneously under local Lorentz transformations,

$$\omega_\mu{}^I{}_J \;\longrightarrow\; \omega'_\mu{}^I{}_J = \Lambda^I{}_K\, \omega_\mu{}^K{}_L\, (\Lambda^{-1})^L{}_J - (\partial_\mu \Lambda^I{}_K)(\Lambda^{-1})^K{}_J. \tag{1.153}$$

Once the spin connection has been introduced, its corresponding field strength is the curvature two-form

$$R^{IJ} = d\omega^{IJ} + \omega^I{}_K \wedge \omega^{KJ}. \tag{1.154}$$

In components this reads

$$R_{\mu\nu}{}^{IJ} = \partial_\mu \omega_\nu{}^{IJ} - \partial_\nu \omega_\mu{}^{IJ} + \omega_\mu{}^I{}_K\, \omega_\nu{}^{KJ} - \omega_\nu{}^I{}_K\, \omega_\mu{}^{KJ}. \tag{1.155}$$

Geometrically, R^{IJ} measures the failure of parallel transport to close around an infinitesimal loop. The pair $(e^I{}_\mu, \omega_\mu{}^{IJ})$ therefore describes spacetime geometry: the tetrad identifies local inertial frames, while the spin connection governs how these frames are compared from point to point.

It should be emphasized that in the standard metric formulation of general relativity, we normally impose from the outset that the connection is torsion-free and fully determined by the tetrad. This special connection, denoted $\bar{\omega}_\mu{}^{IJ}(e, \partial e)$, is defined implicitly by the condition

$$\partial_{[\mu} e^I{}_{\nu]} + \bar{\omega}_{[\mu}{}^I{}_J\, e^J{}_{\nu]} = 0. \tag{1.156}$$

In that case, the spin connection is not an independent variable; it is fixed once the tetrad is known. This reproduces ordinary general relativity in tetrad language.

The Einstein–Cartan formulation departs from this by promoting the spin connection to an independent dynamical field. We then work with the pair

$$\left(e^I{}_\mu, \omega_\mu{}^{IJ}\right) \tag{1.157}$$

as fundamental variables, rather than eliminating $\omega_\mu{}^{IJ}$ in favor of the tetrad. The geometric quantity that diagnoses the difference between the two cases is, therefore, the torsion two-form,

$$T^I := De^I = de^I + \omega^I{}_J \wedge e^J. \tag{1.158}$$

In components,

$$T_{\mu\nu}{}^I = D_\mu e_\nu{}^I - D_\nu e_\mu{}^I, \tag{1.159}$$

and, after contraction with the inverse tetrad,

$$T^\lambda{}_{\mu\nu} = e^\lambda{}_I\, T_{\mu\nu}{}^I. \tag{1.160}$$

In this extension, the intrinsic spin density of matter acts as a source for torsion, just as energy-momentum acts as a source for curvature. The equations of motion for the independent connection are algebraic in torsion. Consequently, it does not typically propagate as an independent field; rather, it is determined locally by the spin content of matter and vanishes in vacuum or in the absence of spinorial sources.

In Palatini–Einstein–Cartan form, gravity may be written as the integral of a Lorentz-invariant four-form built from the tetrad and curvature, possibly supplemented by the Holst term [20] involving the Barbero–Immirzi [21] parameter which we will encounter in the following section. In this form the action is polynomial in the basic variables and manifestly invariant under local Lorentz transformations and diffeomorphisms.

As an aside, in formulations with a nonzero cosmological constant, this idea extends naturally to de Sitter or anti-de Sitter gauge groups [22]. For present purposes, however, we take first-order gravity already possessing a local Lorentz gauge structure, with the tetrad and spin connection as its basic variables, curvature as the corresponding field strength, and torsion as the covariant derivative of the tetrad.

Finally, it is important to note that the antisymmetric property of the spin connection allows the fields to be complexified: the connection may then be decomposed into self-dual and anti-self-dual parts

$$\omega^{IJ} = \omega^{+\,IJ} + \omega^{-\,IJ}, \tag{1.161}$$

with

$$\omega^{\pm IJ} := \frac{1}{2}\Big(\omega^{IJ} \mp \frac{i}{2}\,\epsilon^{IJ}{}_{KL}\,\omega^{KL}\Big)\,. \tag{1.162}$$

This split is the bivector analogue of the decomposition into left-handed and right-handed Weyl spinors. The two sectors then become independent. This marks the transition to the chiral formulation. Similarly, we have to carry out the same self-dual and anti-self-dual decomposition in the case of the curvature 2-form

$$R^{IJ} = R^{+\,IJ} + R^{-\,IJ}. \tag{1.163}$$

1.3 Actions of Gravity in $3 + 1$ Dimensions

Following the establishment of the variables of Einstein–Cartan Gravity, this section discusses the action principles that give rise to the relevant field equations and canonical formalisms.

We begin with the first-order Hilbert-Palatini formulation, which may then be extended to the Holst action. The latter term vanishes in the vacuum equations of motion but becomes central in the canonical formulation. We will then introduce the $3 + 1$ decomposition, followed by establishing the structure of the phase space, at which point we are equipped to conduct the Hamiltonian analysis. From this line of development, the Ashtekar, Barbero, and other connection formulations emerge naturally [20, 21].

1.3.1 First-order Palatini–Holst action

Having established the co-tetrad $e^I{}_\mu$ and Lorentz connection $\omega_\mu{}^{IJ}$ as independent variables, the simplest action in four dimensions satisfying this condition is the Hilbert–Palatini action,

$$S_{\mathrm{HP}}[e,\omega] = \frac{1}{4\kappa}\int_{\mathcal{M}} \epsilon_{IJKL}\, e^I \wedge e^J \wedge R^{KL}(\omega), \tag{1.164}$$

where $\kappa = 8\pi G$. As emphasized in the first-order literature, this action is remarkable in that no additional Lagrange multiplier is needed to enforce compatibility between $e^I{}_\mu$ and $\omega_\mu{}^{IJ}$. The corresponding relation follows directly from the equations of motion.

Varying the action with respect to the spin connection yields

$$D\Big(e^I \wedge e^J\Big) = 0, \tag{1.165}$$

which, for a non-degenerate tetrad and in the absence of spinorial torsion sources, recovers the torsion-free condition

$$T^I = de^I + \omega^I{}_J \wedge e^J = 0. \tag{1.166}$$

Variation with respect to the tetrad then reproduces the Einstein equations. In this sense, the Hilbert–Palatini action is classically equivalent to the Einstein–Hilbert action in vacuum, while retaining a more general off-shell structure that is well suited for canonical analysis.

An interesting extension is proposed by the Palatini–Holst action,

$$S_{\mathrm{PH}}[e,\omega] = \frac{1}{4\kappa}\int_{\mathcal{M}} \left(\epsilon_{IJKL}\, e^I \wedge e^J \wedge R^{KL} + \frac{2}{\gamma}\, e_I \wedge e_J \wedge R^{IJ}\right), \tag{1.167}$$

where γ is the Barbero–Immirzi parameter. The second term is parity-odd and does not alter the vacuum Einstein equations when torsion vanishes, but it modifies the canonical structure. To see this, we examine the behavior of each component of the action under parity transformation.

Let $\mathcal{P}$ denote the spatial parity operator, so that in local Minkowski coordinates

$$x^0 \mapsto x^0, \qquad x^i \mapsto -x^i. \tag{1.168}$$

Accordingly, the co-tetrad components transform as

$$e^0 \xrightarrow{\mathcal{P}} e^0, \qquad e^i \xrightarrow{\mathcal{P}} -e^i. \tag{1.169}$$

Equivalently, in internal-index notation, we may write

$$e^I{}_\mu \xrightarrow{\mathcal{P}} P^I{}_J\, e^J{}_\mu, \qquad P^I{}_J = \begin{cases} \delta^I{}_J, & I = 0, \\ -\delta^I{}_J, & I = 1,2,3. \end{cases} \tag{1.170}$$

From Equation 1.169 it is immediately evident that

$$e^0 \wedge e^i \xrightarrow{\mathcal{P}} -\, e^0 \wedge e^i, \qquad e^i \wedge e^j \xrightarrow{\mathcal{P}} +\, e^i \wedge e^j. \tag{1.171}$$

We may now move on to the action of parity on the curvature tensor. Since R^{IJ} transforms as an antisymmetric Lorentz tensor we have

$$R^{IJ} \xrightarrow{\mathcal{P}} P^I{}_K P^J{}_L\, R^{KL}. \tag{1.172}$$

Hence, its mixed and purely spatial components transform as

$$R^{0i} \xrightarrow{\mathcal{P}} -\, R^{0i}, \qquad R^{ij} \xrightarrow{\mathcal{P}} +\, R^{ij}. \tag{1.173}$$

Combining Equation 1.171 and Equation 1.173, the individual pieces in the Holst term behave as

$$e^0 \wedge e^i \wedge R_{0i} \xrightarrow{\mathcal{P}} (-\, e^0 \wedge e^i) \wedge (-\, R_{0i}) = e^0 \wedge e^i \wedge R_{0i}, \tag{1.174}$$

and

$$e^i \wedge e^j \wedge R_{ij} \xrightarrow{\mathcal{P}} (+\, e^i \wedge e^j) \wedge (+\, R_{ij}) = e^i \wedge e^j \wedge R_{ij}. \tag{1.175}$$

Thus, at the level of internal-index contractions alone, the three-form structure $e_I \wedge e_J \wedge R^{IJ}$ is preserved.

However, as parity acts as an orientation-transforming operator, the four-dimensional integration measure changes sign. The Levi–Civita tensor density changes as

$$\epsilon_{\mu\nu\rho\sigma} \xrightarrow{\mathcal{P}} -\, \epsilon_{\mu\nu\rho\sigma}. \tag{1.176}$$

Therefore in the Palatini term $\mathcal{L}_{\mathrm{P}} = \epsilon_{IJKL}\, e^I \wedge e^J \wedge R^{KL}$, the negative sign contributions from the Levi–Civita tensor density and that from orientation reversal cancel so that

$$\mathcal{L}_{\mathrm{P}} \xrightarrow{\mathcal{P}} \mathcal{L}_{\mathrm{P}}. \tag{1.177}$$

Hence, the Palatini term is parity even.

By contrast, the Holst density $\mathcal{L}_{\mathrm{H}} = e_I \wedge e_J \wedge R^{IJ}$ does not contain ϵ_{IJKL}. Therefore, the only sign comes from the reversal of orientation. It follows that

$$\mathcal{L}_{\mathrm{H}} \xrightarrow{\mathcal{P}} -\, \mathcal{L}_{\mathrm{H}}. \tag{1.178}$$

Thus, the Holst term is parity odd.

Consequently, under parity, the full Palatini–Holst action transforms as

$$S[e,\omega] \xrightarrow{\mathcal{P}} \frac{1}{4\kappa}\int_{\mathcal{M}} \left(\epsilon_{IJKL}\, e^I \wedge e^J \wedge R^{KL} - \frac{2}{\gamma}\, e_I \wedge e_J \wedge R^{IJ}\right). \tag{1.179}$$

We conclude that the Palatini–Holst action is not parity invariant for fixed γ, unless the transformation is accompanied by

$$\gamma \xrightarrow{\mathcal{P}} -\gamma. \tag{1.180}$$

We can interpret the previous results in the context of chirality after decomposing the connection into self-dual and anti-self-dual parts,

$$\frac{1}{2}\epsilon^{IJ}{}_{KL}\,\omega^{\pm\,KL} = \pm i\,\omega^{\pm\,IJ}. \tag{1.181}$$

Since parity reverses the sign of the Hodge dual, it exchanges self-dual and anti-self-dual components:

$$\mathcal{P}: \quad \omega^+ \longleftrightarrow \omega^-. \tag{1.182}$$

Accordingly, a term built purely from one chiral sector, for example, $e \wedge e \wedge R(\omega^+)$, is mapped into the corresponding expression involving the opposite chiral sector,

$$e \wedge e \wedge R(\omega^+) \;\xrightarrow{\mathcal{P}}\; -\, e \wedge e \wedge R(\omega^-), \tag{1.183}$$

where the minus sign again reflects the parity-odd character of the Holst contribution. In this sense, parity does not preserve a single chiral sector separately, but rather exchanges the two sectors. We will return to this point when discussing formulations in which the self-dual and anti-self-dual parts appear with different weights.

As an aside, if we wish to include a cosmological constant, the action is augmented by

$$S_\Lambda = -\frac{\Lambda}{24\kappa}\int_{\mathcal{M}} \epsilon_{IJKL}\, e^I \wedge e^J \wedge e^K \wedge e^L. \tag{1.184}$$

Where the full first-order action takes the schematic form

$$S[e,\omega] = S_{\mathrm{PH}}[e,\omega] + S_\Lambda[e]. \tag{1.185}$$

The purpose of the present chapter, however, is to demonstrate how this framework furnishes a suitable starting point for a $3+1$ decomposition.

1.3.2 The $3+1$ decomposition

Transitioning from the covariant action to the Hamiltonian formulation requires a paradigm shift, viewing the Einstein equations as evolutionary in nature. To achieve this reinterpretation, it is advantageous to decompose spacetime into a foliation of hypersurfaces, thereby separating spatial and temporal components. Then, starting with a specific set of conditions satisfied on the initial hypersurface, we investigate the stability of these properties as they evolve in time.

This view is closely related to the notion of a well-posed Cauchy problem. In the sense introduced by Hadamard, an evolution problem is well posed when three conditions are satisfied: a solution exists for admissible initial data, the solution is unique, and the solution depends continuously on the initial data [23]. These conditions express the predictive content of an evolutionary theory. In general relativity, however, the initial-value problem has a constrained character. The Einstein equations split into constraint equations, which restrict the initial data on a hypersurface, and evolution equations, which determine how that data changes along the chosen foliation. The central aim is then to ensure the validity of initial constraints across all temporal slices.

We assume that the spacetime manifold can be foliated as

$$M \simeq \mathbb{R} \times \Sigma, \tag{1.186}$$

where Σ_t is a family of spacelike hypersurfaces labeled by a coordinate time t. The coordinate t is a parameter labeling the leaves of the foliation. It is not, in general, the proper time measured along the normal direction to the hypersurfaces. The relation between coordinate time and normal proper time is encoded in the lapse function N. If two neighboring hypersurfaces are separated by dt, the proper time elapsed along the unit normal is

$$d\tau = N\,dt. \tag{1.187}$$

Thus, the lapse measures the physical separation between neighboring hypersurfaces in the normal direction.

The coordinate system may also drift tangentially along the hypersurface as we move from Σ_t to Σ_{t+dt}. This tangential displacement is described by the shift vector N^i. Therefore, the "time-flow" vector can be decomposed as

$$\left(\frac{\partial}{\partial t}\right)^a = Nn^a + N^a, \tag{1.188}$$

where n^a is the future-directed unit normal satisfying

$$n^a n_a = -1, \tag{1.189}$$

and N^a is tangent to Σ_t, so that $n_a N^a = 0$. In coordinates adapted to the foliation, $N^a = (0, N^i)$.

The spacetime metric can then be written in ADM form as

$$ds^2 = -N^2 dt^2 + h_{ij}\left(dx^i + N^i dt\right)\left(dx^j + N^j dt\right), \tag{1.190}$$

where h_{ij} is the metric intrinsic to the hypersurface Σ_t. Equivalently, the components of the full spacetime metric are

$$g_{\mu\nu} = \begin{pmatrix} -N^2 + h_{ij}N^iN^j & h_{ij}N^j \\ h_{ij}N^j & h_{ij} \end{pmatrix}, \tag{1.191}$$

and the inverse metric is

$$g^{\mu\nu} = \begin{pmatrix} -\dfrac{1}{N^2} & \dfrac{N^i}{N^2} \\ \dfrac{N^i}{N^2} & h^{ij} - \dfrac{N^iN^j}{N^2} \end{pmatrix}. \tag{1.192}$$

Spatial indices are raised and lowered with h_{ij} and h^{ij}.

The tensor

$$h_{ij} = g_{ij} + n_i n_j \tag{1.193}$$

can also act as the projector onto Σ_t, allowing the spacetime tensors to be decomposed into components normal and tangential to the hypersurface. Figure 1.2 shows this decomposition.

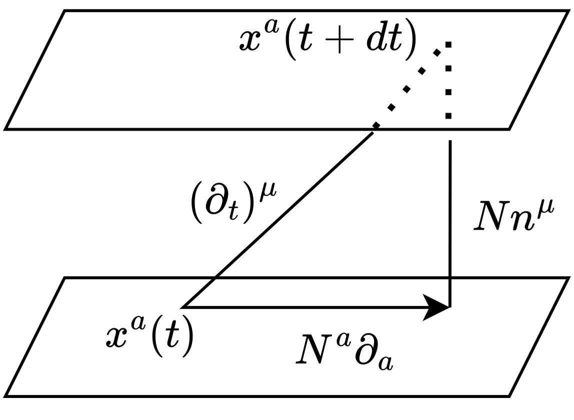


Figure 1.2: Geometric interpretation of the ADM decomposition. The coordinate time-evolution vector $(\partial_t)^\mu$ connects a point $x^a(t)$ on the spatial hypersurface Σ_t to the corresponding point on Σ_{t+dt}. It decomposes into a normal component Nn^μ, controlled by the lapse function N, and a tangential component $N^a\partial_a$, controlled by the shift vector N^a.

These foliations are considered to be embedded in the full spacetime. In order to measure their curvature, we need the so-called extrinsic curvature defined by

$$K_{ij} = -h_i{}^k \nabla_k n_j. \tag{1.194}$$

Geometrically, K_{ij} measures the change of the normal vector as it is parallel transported along the hypersurface, and using $h^k{}_i$ to project the gradient of the normal vector on the hypersurface tells us how it is curved.

A slightly different approach would be to measure the change of the tensor field as it is dragged along the flow of a vector field using the Lie derivative with respect to the normal vector.

$$\mathcal{L}_n h_{ij} = -2K_{ij}. \tag{1.195}$$

From the decomposition (1.188), it follows that

$$\partial_t h_{ij} = -2NK_{ij} + D_i N_j + D_j N_i, \tag{1.196}$$

where D_i is the covariant derivative compatible with h_{ij}. This relation identifies K_{ij} as the normal-time variation of the spatial metric.

The Einstein equations can now be separated into projections normal and tangential to Σ_t. The tangential projection gives the Hamiltonian constraint,

$$^{(3)}R + K^2 - K_{ij}K^{ij} = 16\pi\rho, \tag{1.197}$$

where

$$K = h^{ij}K_{ij}, \qquad \rho = n^\mu n^\nu T_{\mu\nu}. \tag{1.198}$$

The mixed normal-spatial projection gives the momentum constraint,

$$D_j \left(K^{ij} - h^{ij}K\right) = 8\pi J^i, \tag{1.199}$$

with

$$J^i = -h^i{}_\mu n_\nu T^{\mu\nu}. \tag{1.200}$$

These equations restrict the admissible initial data (h_{ij}, K_{ij}) on a spatial hypersurface. The remaining six equations determine the evolution of K_{ij}:

$$\begin{aligned}\partial_t K_{ij} = &-D_i D_j N + N\left({}^{(3)}R_{ij} + KK_{ij} - 2K_{ik}K^k{}_j\right)\\ &+ \mathcal{L}_{\vec{N}}K_{ij} - 8\pi N\left[\bar{S}_{ij} - \frac{1}{2}h_{ij}(\bar{S} - \rho)\right],\end{aligned} \tag{1.201}$$

where

$$\bar{S}_{ij} = h_i{}^\mu h_j{}^\nu T_{\mu\nu}, \qquad \bar{S} = h^{ij}\bar{S}_{ij}. \tag{1.202}$$

The system is therefore supplemented by constraints that must be imposed on the initial hypersurface and preserved by the evolution. For a more in-depth analysis, see Appendix A

In the formalism relevant to our purpose, the same decomposition is performed directly at the level of the tetrad and spin connection

$$e^I{}_\mu = \left(e^I{}_0,\ e^I{}_i\right), \qquad \omega_\mu{}^{IJ} = \left(\omega_0{}^{IJ},\ \omega_i{}^{IJ}\right). \tag{1.203}$$

Analogous to the previous discussion, a useful way to express the tetrad decomposition, without fixing an internal gauge from the outset, is to split the tetrad into parts normal and tangential to the spatial slice. Denoting by N^I an internal timelike unit vector and by $V_a{}^I$ the triad intrinsic to Σ_t, we may write

$$e_0{}^I = N\,N^I + N^a V_a{}^I, \qquad e_a{}^I = V_a{}^I, \tag{1.204}$$

subject to

$$N_I V_a{}^I = 0, \qquad N_I N^I = -1. \tag{1.205}$$

and the spatial metric is induced as

$$h_{ab} = \eta_{IJ} V_a{}^I V_b{}^J. \tag{1.206}$$

At this stage, we may either impose an internal time gauge already at the Lagrangian level, or else keep the full Lorentz covariance manifest until later in the analysis. The time gauge is usually defined by aligning the internal timelike vector with the normal to the hypersurface so that only an internal $SO(3)$ subgroup remains, which in turn significantly simplifies the Legendre transformation.

Nevertheless, to avoid any loss of generality at present, we refrain from imposing this gauge.

1.3.3 Hamiltonian formulation and constrained dynamics

We have performed the $3+1$ decomposition which isolates the variables that evolve along the foliation parameter t from the spatial ones. Once this split has been carried out, the action may be rewritten in a form adapted to canonical analysis, thereby making explicit the phase-space structure of the theory and the constraints associated with its gauge symmetries.

To provide some context, we may recall that Newtonian mechanics was originally reformulated by Joseph-Louis Lagrange, whose work was then further refined by William Rowan Hamilton. While initially intended to simplify the equations of motion, Hamilton's approach unexpectedly paved the way for serendipitous advances in statistical mechanics and quantum mechanics. Sometimes the longest way around is the shortest way home![4] Indeed, the long way brings us to a formalism in which, instead of one second-order equation governing the evolution of degrees of freedom, we have two first-order equations.

In an ordinary mechanical system with canonical variables $q^A(t)$, we pass to the Hamiltonian formulation by defining the canonical momenta

$$p_A := \frac{\partial L}{\partial \dot{q}^A}, \tag{1.207}$$

and then inverting the relation between $\dot{q}^A$ and p_A to implement the Legendre transform[5]. In field theory, the same construction is performed on each spatial slice Σ_t, with fields $\chi^A(t,x)$ replacing the finite-dimensional variables $q^A(t)$.

In this case, if the map

$$(q^i, \dot{q}^i) \longmapsto (q^i, p_i) \tag{1.208}$$

is invertible, then the velocities may be expressed as functions of the canonical variables,

$$\dot{q}^i = \dot{q}^i(q,p), \tag{1.209}$$

and the canonical Hamiltonian is obtained in the usual way:

$$H_{\mathrm{c}}(q,p) = p_i \dot{q}^i - L(q,\dot{q}). \tag{1.210}$$

The action may then be written in phase-space form,

$$S = \int dt \left(p_i \dot{q}^i - H_{\mathrm{c}}(q,p) \right). \tag{1.211}$$

which gives the symplectic structure of the theory.

With no deviation from established logic, we explore the system time evolution. To this end, following the guidelines established by Lagrange, the physicist and mathematician Siméon Denis Poisson was developing a framework to study perturbations of the orbits of celestial bodies. He proposed using the canonical momentum instead of the velocity, which was then extended by Hamilton. He introduced one of the most important quantities at the heart of analytical mechanics, called today the Poisson bracket.

[4] James Joyce said that

[5] As an entertaining aside, Adrien–Marie Legendre conceived his momentous idea for variable transformation to solve an equation proposed by M. de la Grange regarding surfaces in geometry and completely unrelated to mechanics. He wrote in his Memoire: "... I succeeded quite simply through a change of variables ... which might be useful in other occasions". Little did he know!

Given two functions $F(q,p)$ and $G(q,p)$ on phase space,

$$\{F,G\} = \frac{\partial F}{\partial q^i}\frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial G}{\partial q^i}, \tag{1.212}$$

with the fundamental relation

$$\{q^i, p_j\} = \delta^i{}_j. \tag{1.213}$$

Time evolution is then generated by the Hamiltonian flow,

$$\dot{q}^i = \{q^i, H_\mathrm{c}\} = \frac{\partial H_\mathrm{c}}{\partial p_i}, \qquad \dot{p}_i = \{p_i, H_\mathrm{c}\} = -\frac{\partial H_\mathrm{c}}{\partial q^i}. \tag{1.214}$$

But of course, "The worst is not, So long as we can say 'This is the worst'": In gauge theories of gravity, not all fields appear with independent time derivatives in the Lagrangian. Some fields occur only linearly in time derivatives, while others appear without any time derivatives at all. As a result, not all canonical momenta are independent, and relations arise among the phase-space variables. These relations are the *constraints* of the theory.

Let us then recall the general logic of constrained Hamiltonian systems. In other words, cases where the canonical variables satisfy

$$\psi_\alpha(q,p) = 0. \tag{1.215}$$

These relations are the primary constraints of the theory. They define a submanifold of the full phase space, usually called the primary constraint surface. Thus, the physically admissible initial data do not fill the entire canonical phase space, but are restricted to lie on the surface

$$\Gamma_\mathrm{p} = \{(q,p) \mid \psi_\alpha(q,p) = 0\}. \tag{1.216}$$

In the presence of primary constraints, the Hamiltonian must be enlarged by adding the constraints with arbitrary Lagrange multipliers. We therefore introduce the total Hamiltonian

$$H_\mathrm{T} = H_\mathrm{c} + V^\alpha \psi_\alpha, \tag{1.217}$$

where the functions $V^\alpha(t)$ are not fixed a priori. Their arbitrariness reflects the fact that part of the evolution generated by H_T corresponds to gauge freedom rather than to physical change.

The time evolution of any phase-space function $F(q,p)$ is then given by

$$\dot{F} = \{F, H_\mathrm{T}\}. \tag{1.218}$$

To put this section in perspective, the reader might recall that while laying the groundwork for $3+1$ decomposition, we mentioned that one of the requirements of spacetime foliation is the preservation of properties satisfied on one hypersurface through time evolution.

In analogy, here too, the constraints must be preserved by time evolution. Hence, we must impose the consistency conditions

$$\dot{\psi}_\alpha = \{\psi_\alpha, H_{\mathrm{T}}\} \approx 0, \tag{1.219}$$

where the symbol $\approx$ denotes equality on the constraint surface.

These consistency conditions may have different consequences. They may be satisfied identically, they may determine some of the Lagrange multipliers V^α, or they may generate new constraints. If new constraints arise, their preservation must be checked in turn. This iterative procedure continues until no further constraints are produced.

There is a further distinction to be made. We differentiate between first-class and second-class constraints. A constraint ϕ_a is said to be first class if its Poisson bracket with all constraints vanishes weakly,

$$\{\phi_a, \psi_\alpha\} \approx 0. \tag{1.220}$$

First-class constraints generate gauge transformations on phase space. Indeed, if ϕ_a is first class, then an infinitesimal transformation generated by ϕ_a acts on any phase-space function F as

$$\delta_\epsilon F = \epsilon^a \{F, \phi_a\}, \tag{1.221}$$

where ϵ^a are infinitesimal gauge parameters. Such transformations move the system along gauge orbits inside the constraint surface and therefore do not change the physical state.

The remaining constraints are second class. If these are denoted by χ_m, their Poisson bracket matrix is

$$\Delta_{mn} = \{\chi_m, \chi_n\}. \tag{1.222}$$

After all first-class combinations have been separated out, the matrix Δ_{mn} should be invertible on the constraint surface. Second-class constraints do not generate gauge freedom. Instead, they remove redundant variables by restricting the phase space directly.

Let us now concretely consider the collection of fields after the 3+1 decomposition of the action integral denoted by χ^A.

$$\tilde{\mathcal{L}} = \sum_B a_B(\chi, \partial_a\chi)\,(\dot{\alpha}^B)^2 + \sum_C b_C(\chi, \partial_a\chi)\,\dot{\beta}^C - \tilde{U}(\chi, \partial_a\chi), \tag{1.223}$$

Generally, the Lagrangian density may be organized schematically according to its dependence on time derivatives as

$$\begin{aligned}\tilde{\mathcal{L}}_{ext} = \sum_B \tilde{P}_B(\dot{\alpha}^B - V^B) + \sum_C \tilde{P}_C(\dot{\beta}^C - V^C) + \sum_D \tilde{P}_D(\dot{\gamma}^D - V^D) \\ + \sum_B a_B(\chi, \partial_a\chi)\,(V^B)^2 + \sum_C b_C(\chi, \partial_a\chi)\,V^C - \tilde{U}(\chi, \partial_a\chi).\end{aligned} \tag{1.224}$$

where the fields have been separated into three classes:

1. fields α^B that appear quadratically in their time derivatives,
2. fields β^C that appear linearly in their time derivatives,
3. fields γ^D that appear without time derivatives.

Additionally, it is convenient to introduce auxiliary velocity fields together with momenta as independent variables.

The equations of motion obtained by varying the momenta $\tilde{P}_A$ in the extended form enforce

$$\frac{\delta\tilde{\mathcal{L}}_{\text{ext}}}{\delta\tilde{P}_B} = \dot{\alpha}^B - V^B = 0, \qquad \frac{\delta\tilde{\mathcal{L}}_{\text{ext}}}{\delta\tilde{P}_C} = \dot{\beta}^C - V^C = 0, \qquad \frac{\delta\tilde{\mathcal{L}}_{\text{ext}}}{\delta\tilde{P}_D} = \dot{\gamma}^D - V^D = 0. \tag{1.225}$$

Thus,

$$V^B = \dot{\alpha}^B, \qquad V^C = \dot{\beta}^C, \qquad V^D = \dot{\gamma}^D, \tag{1.226}$$

so that the original Lagrangian of Equation 1.223 is recovered.

Additionally, varying (1.224) with respect to V^B gives

$$\frac{\delta\tilde{\mathcal{L}}_{\text{ext}}}{\delta V^B} = -\tilde{P}_B + 2a_B(\chi, \partial_a\chi)V^B = 0. \tag{1.227}$$

Therefore,

$$V^B = \frac{\tilde{P}_B}{2a_B(\chi, \partial_a\chi)}. \tag{1.228}$$

This equation can be solved for V^B. The fields α^B therefore form the regular part of the Legendre transformation.

By contrast, variation with respect to V^C gives

$$\frac{\delta\tilde{\mathcal{L}}_{\text{ext}}}{\delta V^C} = -\tilde{P}_C + b_C(\chi, \partial_a\chi) = 0. \tag{1.229}$$

This equation does not determine V^C. Instead, it gives the constraint

$$C_C(\tilde{P}, \chi, \partial_a\chi) := \tilde{P}_C - b_C(\chi, \partial_a\chi) = 0. \tag{1.230}$$

Similarly, variation with respect to V^D gives

$$\frac{\delta \tilde{\mathcal{L}}_{\text{ext}}}{\delta V^D} = -\tilde{P}_D = 0, \tag{1.231}$$

and therefore

$$C_D(\tilde{P}, \chi) := \tilde{P}_D = 0. \tag{1.232}$$

We see that fields whose time derivatives enter linearly, as well as those whose time derivatives do not appear at all, give rise to constraints.

We can now start to formulate the Hamiltonian of the system by first eliminating the velocity V^B using (1.228). The extended Lagrangian can then be written as

$$\tilde{\mathcal{L}}_{\text{ext}} = \sum_A \tilde{P}_A \dot{\chi}^A - \tilde{\mathcal{H}}, \tag{1.233}$$

where

$$\tilde{\mathcal{H}} = \sum_B \tilde{P}_B V^B + \sum_C \tilde{P}_C V^C + \sum_D \tilde{P}_D V^D - \sum_B a_B (V^B)^2 - \sum_C b_C V^C + \tilde{U}. \tag{1.234}$$

For the quadratic sector we find

$$\begin{aligned} \tilde{P}_B V^B - a_B (V^B)^2 &= \tilde{P}_B \left(\frac{\tilde{P}_B}{2a_B} \right) - a_B \left(\frac{\tilde{P}_B}{2a_B} \right)^2 \\ &= \frac{\tilde{P}_B^2}{2a_B} - \frac{\tilde{P}_B^2}{4a_B} = \frac{\tilde{P}_B^2}{4a_B}. \end{aligned} \tag{1.235}$$

Thus the part of the Hamiltonian density associated with the regular sector is

$$\tilde{\mathcal{H}}_0 = \sum_B \frac{\tilde{P}_B^2}{4a_B(\chi, \partial_a \chi)} + \tilde{U}(\chi, \partial_a \chi). \tag{1.236}$$

The remaining terms are

$$\sum_C \tilde{P}_C V^C - \sum_C b_C V^C = \sum_C V^C \left(\tilde{P}_C - b_C(\chi, \partial_a \chi) \right), \tag{1.237}$$

and

$$\sum_D \tilde{P}_D V^D = \sum_D V^D \tilde{P}_D. \tag{1.238}$$

Using the constraint definitions (1.230) and (1.232), the Hamiltonian density therefore becomes

$$\tilde{\mathcal{H}} = \tilde{\mathcal{H}}_0(\tilde{P}_A, \chi^A, \partial_a \chi^A) + \sum_C V^C C_C(\tilde{P}_A, \chi^A, \partial_a \chi^A) + \sum_D V^D C_D(\tilde{P}_A, \chi^A). \tag{1.239}$$

Consequently, the Lagrangian density takes the canonical Hamiltonian form

$$\tilde{\mathcal{L}} = \sum_A \tilde{P}_A \dot{\chi}^A - \tilde{\mathcal{H}}(\tilde{P}_A, \chi^A, \partial_a \chi^A, V^C, V^D). \tag{1.240}$$

Equations (1.240) and (1.239) define the constraint surface inside the phase space coordinatized by

$$(\chi^A, \tilde{P}_A).$$

From a field theory perspective, for two functionals $F(\chi, \tilde{P})$ and $G(\chi, \tilde{P})$ on phase space, the Poisson bracket is defined by

$$\{F, G\} = \int d^3x \sum_A \left(\frac{\delta F}{\delta \chi^A} \frac{\delta G}{\delta \tilde{P}_A} - \frac{\delta G}{\delta \chi^A} \frac{\delta F}{\delta \tilde{P}_A} \right). \tag{1.241}$$

which implies the relations

$$\{\chi^A(x), \tilde{P}_B(y)\} = \delta^A{}_B \, \delta^{(3)}(x-y), \qquad \{\chi^A(x), \chi^B(y)\} = 0, \qquad \{\tilde{P}_A(x), \tilde{P}_B(y)\} = 0. \tag{1.242}$$

The Hamiltonian functional is then obtained by integrating the Hamiltonian density over the spatial hypersurface,

$$H = \int_\Sigma d^3x \, \tilde{\mathcal{H}}. \tag{1.243}$$

Stationarity of the action with respect to variations of the canonical variables $(\chi^A, \tilde{P}_A)$ gives Hamilton's equations

$$\dot{\chi}^A = \left\{ \chi^A, \int_\Sigma d^3x \, \tilde{\mathcal{H}} \right\}, \tag{1.244}$$

and

$$\dot{\tilde{P}}_A = \left\{ \tilde{P}_A, \int_\Sigma d^3x \, \tilde{\mathcal{H}} \right\}. \tag{1.245}$$

These equations express time evolution as Hamiltonian flow on the constrained phase space. More generally, for any phase-space functional $F(\chi, \tilde{P})$, its time evolution is

$$\dot{F} = \left\{ F, \int_\Sigma d^3x \, \tilde{\mathcal{H}} \right\}. \tag{1.246}$$

As mentioned before, if a constraint C is imposed on the initial hypersurface, consistency requires that the following remain weakly satisfied under time evolution:

$$\dot{C} = \left\{ C, \int_\Sigma d^3x \, \tilde{\mathcal{H}} \right\} \approx 0. \tag{1.247}$$

1.4 Ashtekar Formulation of General Relativity

In the mid-1980s, Abhay Ashtekar presented a revolutionary reformulation of general relativity that has profoundly influenced the development of canonical quantum gravity, particularly loop quantum gravity (LQG). Unlike the traditional Arnowitt-Deser-Misner (ADM) formulation, which uses the spatial metric h_{ab} and extrinsic curvature K_{ab} as canonical variables, the Ashtekar formulation employs a connection and its conjugate momentum reminiscent of Yang-Mills gauge theory. These variables delineate how interactions are mediated by force carriers, thereby establishing essential connections.

To gain a better intuition about this approach, consider transporting a neutral particle along a specified curve from a point A to a point B within a general spatial continuum. This task should ostensibly be straightforward from a mathematical perspective. However, if it happens that an electromagnetic field is introduced and that particle of interest is charged, the operation becomes slightly more complicated in the sense that the propagation will now depend on the path configuration. The wave function undergoes a phase alteration, otherwise known as a gauge transformation. The mediators of such transformations are known as connections or vector potentials because traditionally they were used to denote the potential of an electromagnetic field.

Chen-Ning Yang first presented the idea of such vector bosons at a conference where another prominent figure could be seen amongst the audience: Wolfgang Pauli, an intellectually brilliant, albeit very cutting, character [6] for a deliberate lack of a better word, who remained skeptical about the mass of these new intermediaries throughout the entirety of Yang's talk. Needless to say, Yang's ideas turned out to be foundational for theoretical physics [24].

Ashtekar, by whom the above anecdote was once recounted, had the brilliant idea to extend the same formalism in the presence of a gravitational field. To achieve this, he drew inspiration from the work of Ezra T. Newman and Roger Penrose, who demonstrated that a specific sector of general relativity could be articulated in terms of gauge bosons with particular helicity properties. Indeed, for a zero-rest-mass particle the spin is either alongside or opposite to its 4-velocity. Technically, the helicity configuration is called self-dual or anti-self-dual. In these helicity sectors, the Einstein equations become an integrable system, although they remain nonlinear. The helicity state in terms of the connection propagates through the wave function of a massive spinning particle in the gravitational field and will depend on the path.

The preceding sections have introduced the formulation of gravity in terms of a cotetrad $e^I{}_\mu$ and spin connection $\omega_\mu{}^{IJ}$, together with the $3+1$ decomposition and the Hamiltonian description of constrained systems. We now use these structures to introduce

[6] Quite literally as it happened! He kept interrupting Yang so many times that it ended up preventing him from finishing his talk entirely.

the Ashtekar variables.

The central idea is to replace the metric phase-space variables of the Arnowitt-Deser-Misner ADM formulation [25] by a connection-triad pair. In this form, the configuration variable is a connection on the spatial hypersurface, and its conjugate momentum is a densitized triad. The index notation is as follows. We use $a, b, c, \ldots = 1, 2, 3$ for spatial coordinate indices and $I, J, K, \ldots = 1, 2, 3$ for internal indices The original complex Ashtekar connection arises from a self-dual splitting of the Lorentz connection as

$$A^I{}_a = \Gamma^I{}_a + iK^I{}_a. \tag{1.248}$$

where the extrinsic curvature K_{ab} is given by

$$K^I{}_a = K_{ab} e^{bI}, \tag{1.249}$$

We may note that e^{ai} is the inverse triad satisfying

$$e^I{}_a e^{aJ} = \delta^{IJ}, \qquad e^I{}_a e^b{}_I = \delta^b_a. \tag{1.250}$$

The connection in this form is the self-dual part of the full spacetime connection when restricted to the spatial slice. The factor of i ensures that A^I_a transforms as an $SL(2, \mathbb{C})$ connection.

The conjugate momentum to the connection is the densitized triad. This is the tensor-density type object that was discussed in the section regarding differential forms.

$$E^a{}_I = \det(e)\, e^a{}_I. \tag{1.251}$$

Equivalently, in terms of the spatial triad,

$$E^a{}_I = \frac{1}{2}\epsilon^{abc}\epsilon_{IJK} e^J{}_b e^K{}_c. \tag{1.252}$$

The densitized triad carries density weight $+1$. It determines the inverse spatial metric through

$$E^a{}_I E^b{}_J \delta^{IJ} = h\, h^{ab}, \qquad h := \det(h_{ab}). \tag{1.253}$$

For the complex self-dual variables, the fundamental Poisson bracket is

$$\left\{A^I{}_a(x), E^b{}_J(y)\right\} = 8\pi G\, \delta^I{}_J \delta^b{}_a \delta^{(3)}(x, y), \tag{1.254}$$

with the remaining elementary brackets vanishing,

$$\{A^I{}_a(x), A^J{}_b(y)\} = 0, \qquad \{E^a{}_I(x), E^b{}_J(y)\} = 0. \tag{1.255}$$

In this self-dual/anti-self-dual case, the Barbero–Immirzi parameter introduced in the Holst action corresponds formally to $\gamma = \pm i$.

A key challenge of this approach, however, is that while the complex self-dual connection leads to polynomial constraints, we must introduce certain reality conditions to ensure the physical phase space is restricted to configurations where the metric and extrinsic curvature are real.

Given the preceding mathematical prelude, we can now examine the curious case of a spontaneously broken gauge theory of the Lorentz group

Chapter 2

Hamiltonian Formulation of Gravity as a Spontaneously-Broken Gauge Theory of the Lorentz Group

"There are mysteries which men can only guess at, which age by age they may solve only in part."
— Bram Stoker

2.1 Chiral Gravitational Action

Previously, we established a description of gravity in terms of the co-tetrad $e^I{}_\mu$ and the Lorentz-valued spin connection $\omega_\mu{}^{IJ}$. We also introduced the self-dual and anti-self-dual decomposition of antisymmetric Lorentz tensors. This chapter endeavors to present a more global approach by embedding the previous results in a framework where the coordinate-like fields $\phi^I(x)$ allow a theory originally written on a fixed background to be expressed in a diffeomorphism-invariant form.

The motivation for this idea may be understood from the perspective of parametrized field theory [11]. Consider, for instance, how in ordinary mechanics, time is often treated as an external parameter. The dynamical variables q^i evolve with respect to a preferred time coordinate τ. But from this new point of view, we may instead promote τ to a dynamical field $\tau(\lambda)$[1]. The theory then acquires a reparametrization symmetry, although no new physical degree of freedom has been introduced. The τ field exhibits an interesting behavior. It is compelled to act like a new degree of freedom, but the imposed constraint on the Hamiltonian $\mathcal{H} = 0$ in the model prevents it from doing so. The original dynamics are then recovered if a gauge such as $\tau = \lambda$ is accessible.

[1]Promoting time to a field such that $\tau(\lambda)$ gains its own equation of motion is called parametrized particle mechanics.

The same logic can be applied to a relativistic field theory. Hence, a field theory originally written on a fixed Minkowski background may be rewritten in a diffeomorphism-invariant form by constructing the metric as

$$g_{\mu\nu} \longrightarrow \eta_{IJ}\partial_\mu\phi^I\partial_\nu\phi^J. \tag{2.1}$$

Here, the fields $\phi^I = \phi^I(x)$. If the field configuration $\phi^I = x^I$ can be reached[2]by an admissible gauge choice, then the original Minkowski-space theory is recovered, with x^I denoting the Minkowski coordinates. To provide a better foundation for this proposition, we may invoke a tenet of particle physics – particularly true within the spatial scale characteristic of particle interactions such as those observed in particle accelerators – where the curvature of spacetime is approximated as negligible. The action written on a Minkowski background will then take the form

$$S[\tilde{\varphi}] = \int d^4X \left(-\eta^{IJ}\partial_I\tilde{\varphi}\,\partial_J\tilde{\varphi} - V(\tilde{\varphi})\right). \tag{2.2}$$

where η is the Minkowski metric and X^I are Minkowski coordinates. It is pertinent to mention two symmetries of this model. First, we note that the action is invariant under the coordinate transformation

$$X^I \longrightarrow \Lambda^I{}_J X^J, \tag{2.3}$$

where $\Lambda^I{}_J$ is a matrix representing a Lorentz transformation and does not depend on position in spacetime and thus denotes a global symmetry. In addition, we can establish the symmetry under

$$X^I \longrightarrow X^I + P^I, \tag{2.4}$$

where again P^I does not depend on position in spacetime. In parametrized field theories, X^I are promoted to dynamical fields which will henceforth be called ϕ^I. This yields the following action

$$S_{\text{PFT}}[\tilde{\varphi}, \phi^I] = \int d^4x \sqrt{-\mathcal{G}}\,\left(-\mathcal{G}^{\mu\nu}\partial_\mu\tilde{\varphi}\,\partial_\nu\tilde{\varphi} - V(\tilde{\varphi})\right). \tag{2.5}$$

where x^μ denotes general coordinates and $\mathcal{G}_{\mu\nu} = \eta_{IJ}\phi^I_\mu\phi^J_\nu$ where $\phi^I_\mu = \partial_\mu\phi^I$. Furthermore, this new functional of ϕ^I and $\tilde{\varphi}$ now has more symmetries. Specifically, it is invariant under spacetime diffeomorphism, i.e., it remains unchanged under the coordinate transformation $x^\mu \to \tilde{x}^\mu$. Thus, we have created a local symmetry visible in GR theories starting from a non-gravitational formalism where $\mathcal{G}$ has zero curvature.

To draw a parallel with the current framework, we have designed the model from the

[2]The author requests the reader to remain patient regarding the technical details. They will be provided shortly.

standpoint of a non-gravitational model whose global symmetry

$$\phi^I \longrightarrow \Lambda^I{}_J \phi^J, \tag{2.6}$$

is promoted to a local one by introducing the spin connection. Therefore, using $(\phi^I, \omega_\mu{}^{IJ})$ as the basic variables with

$$D_\mu \phi^I = \partial_\mu \phi^I + \omega^I{}_{J\mu} \phi^J := e^I{}_\mu, \tag{2.7}$$

we have effectively recovered the tetrad field.

Additionally, to make contact with the $3+1$ description introduced earlier, we can reconstruct ϕ^I as

$$\phi^I = \phi_{(N)} N^I + \varphi^I, \tag{2.8}$$

where $\varphi_I N^I = 0$. As a reminder, N^I is the internal timelike vector associated with the foliation, while φ^I lies in the internal spatial hypersurface orthogonal to N^I.

From our co-tetrad re-identification, we obtain

$$\phi_I e^I{}_i = \phi_I D_i \phi^I = \frac{1}{2} \partial_i \phi^2, \tag{2.9}$$

where we introduce $\phi^2 := \eta_{IJ} \phi^I \phi^J$ and using the orthogonality condition $N_I e^I{}_i = 0$, we derive the spatial component of the internal field such that $\phi_I e^I{}_i = \varphi_I e^I{}_i$. In terms of the spatial metric, this gives

$$\varphi_I = \frac{1}{2} h^{ij} e_{Ii} \partial_j \phi^2 \tag{2.10}$$

where h_{ij} denotes the metric on spatial slices. The remaining component of ϕ^I, namely its projection along N^I, is then fixed by the identity

$$\phi^2 = -\phi_{(N)}^2 + \varphi_I \varphi^I, \tag{2.11}$$

which yields

$$\phi_{(N)} = -\xi \sqrt{-\phi^2 + \frac{1}{4} h^{ij} \partial_i \phi^2 \partial_j \phi^2}, \qquad \xi = \mp 1. \tag{2.12}$$

The two signs correspond to the two possible orientations of ϕ^I relative to the internal normal.

2.1.1 Self-dual and anti-self-dual Einstein–Cartan action

We begin with the Einstein–Cartan Palatini action in four dimensions,

$$S_{\mathrm{EC}}[e, \omega] = \frac{1}{64\pi G} \int d^4x \, \tilde{\epsilon}^{\mu\nu\alpha\beta} \epsilon_{IJKL} e^I{}_\mu e^J{}_\nu R_{\alpha\beta}{}^{KL}(\omega), \tag{2.13}$$

where $\tilde{\epsilon}^{\mu\nu\alpha\beta}$ is the Levi-Civita density and

$$R_{\alpha\beta}{}^{IJ}(\omega) = 2\partial_{[\alpha}\omega^{IJ}{}_{\beta]} + 2\omega^{I}{}_{K[\alpha}\omega^{KJ}{}_{\beta]} \tag{2.14}$$

is the curvature associated with ω_μ^{IJ} in index notation. We note that the spacetime diffeomorphism symmetry is equivalent to that of General Relativity. Moreover, since the spin connection is antisymmetric in its Lorentz indices, it admits a decomposition into self-dual and anti-self-dual components. To achieve this, it is useful to introduce the projectors

$$\mathcal{K}^{\pm}{}_{IJKL} := \frac{1}{2}\left(\epsilon_{IJKL} \mp 2i\eta_{I[K}\eta_{L]J}\right) \tag{2.15}$$

The corresponding self-dual and anti-self-dual Einstein–Cartan actions composed of the two independent fields are given by [10]

$$\begin{aligned} S_{\mathrm{EC}\pm} &= \frac{1}{32\pi G}\int d^4x\, \tilde{\epsilon}^{\mu\nu\alpha\beta}\mathcal{K}^{\pm}{}_{IJKL}e^{I}{}_{\mu}e^{J}{}_{\nu}R_{\alpha\beta}{}^{KL}(\omega) \\ &= \frac{1}{32\pi G}\int d^4x\, \tilde{\epsilon}^{\mu\nu\alpha\beta}\epsilon_{IJKL}e^{I}{}_{\mu}e^{J}{}_{\nu}R^{\pm\,KL}_{\alpha\beta}(\omega) \\ &= \frac{1}{32\pi G}\int d^4x\, \tilde{\epsilon}^{\mu\nu\alpha\beta}\epsilon_{IJKL}e^{I}{}_{\mu}e^{J}{}_{\nu}R^{\pm\,KL}_{\alpha\beta}(\omega^{\pm}). \end{aligned} \tag{2.16}$$

These actions yield the complexified Einstein equations. It is important to note that each sector independently reproduces standard General Relativity upon imposition of reality conditions.

Given the family of metrics in Equation 2.1, we may construct Lagrangian densities which are quadratic with respect to the covariant derivative of the scalar field and linear with respect to the spin connection. The simplest of such actions with non-trivial gravitational dynamics is given by

$$S[\phi,\omega] = \frac{\alpha}{2}\int d^4x\, \tilde{\epsilon}^{\mu\nu\alpha\beta}\left(\epsilon_{IJKL} + \frac{2}{\gamma}\eta_{K[I}\eta_{J]L}\right)D_\mu\phi^I D_\nu\phi^J R_{\alpha\beta}{}^{KL}(\omega), \tag{2.17}$$

where α is an overall normalization factor related to the inverse of Newton's constant at the GR limits. As such, we will fix $\alpha = 1$, representing our choice of units. As for γ, it plays a role analogous to the Barbero–Immirzi parameter of the Palatini–Holst action.

Using the self-dual and anti-self-dual decomposition, the action may be written as

$$S[\phi,\omega^+,\omega^-] = \int d^4x\, \tilde{\epsilon}^{\mu\nu\alpha\beta}\epsilon_{IJKL}D_\mu\phi^I D_\nu\phi^J\left(g_+R^{+\,KL}_{\alpha\beta}(\omega^+) + g_-R^{-\,KL}_{\alpha\beta}(\omega^-)\right), \tag{2.18}$$

where the chiral couplings are given by

$$g_\pm = \frac{\alpha}{2}\left(\frac{\gamma \mp i}{\gamma}\right). \tag{2.19}$$

Equivalently,

$$\alpha = g_+ + g_-, \qquad \gamma = i\left(\frac{g_+ + g_-}{g_- - g_+}\right). \tag{2.20}$$

Here, at the level of the Lagrangian density, this expression bears a close resemblance to the Palatini action, provided we identify the co-tetrad e^I_μ with the covariant derivative $D_\mu\phi^I$. More specifically, the parameter γ plays a role analogous to the Barbero–Immirzi parameter. Then, the limit $\gamma \to \infty$ corresponds to $g_+ = g_-$, resembling the Palatini case, while the choices $\gamma = i$ and $\gamma = -i$ isolate the self-dual and anti-self-dual sectors, respectively. Indeed, this follows from the fact that the complexified Lorentz algebra decomposes into two independent chiral sectors, corresponding to the eigenspaces of the internal Hodge dual.

Another remarkable feature, to be discussed shortly, is that the degrees of freedom depend essentially on (g_+, g_-).

2.1.2 3+1 decomposition of the Lagrangian density

Given the Lagrangian density, we now prepare the system to analyze its evolution. To this end, we decompose the connection into its temporal and spatial components,

$$\omega^{\pm IJ}_\mu dx^\mu = \Omega^{\pm IJ} dt + \beta^{\pm IJ}_a dx^a, \tag{2.21}$$

$\Omega^{\pm IJ}$ denotes the coefficient of the connection along the coordinate one-form dt, while $\beta^{\pm IJ}_a$ is the pullback of the connection to Σ_t. The spatial curvature of the connection is then defined by

$$R^{\pm IJ}_{ab} \equiv 2\left(\partial_{[a}\beta^{\pm IJ}_{b]} + \beta^{\pm IK}_{[a|}\beta^{\pm J}{}_{K|b]}\right). \tag{2.22}$$

Finally, the co-tetrad is reconstructed from the Lorentz-covariant derivative of the Higgs-like field ϕ^I that plays the role of a symmetry-breaking field

$$e^I_a \equiv \partial_a\phi^I + \beta^I{}_{Ja}\phi^J = D_a\phi^I, \tag{2.23}$$

where $\beta^{IJ}_a = \beta^{+IJ}_a + \beta^{-IJ}_a$.

The Lagrangian density of Equation 2.18 can be reformulated as

$$\mathcal{L} = \tilde{\epsilon}^{\mu\nu\alpha\beta} C_{IJKL} D_\mu\phi^I D_\nu\phi^J R^{KL}_{\alpha\beta}, \tag{2.24}$$

where the internal tensor C_{IJKL} is written in terms of the self-dual and anti-self-dual projectors given in Equation 2.15

$$C_{IJKL} = \epsilon_{IJKL} + \frac{2}{\gamma}\eta_{K[I}\eta_{J]L} = \mathcal{A}\,\mathcal{K}^{+}_{IJKL} + \mathcal{B}\,\mathcal{K}^{-}_{IJKL}. \tag{2.25}$$

The spatial and mixed components of the curvature in terms of the decomposed parameters give

$$R^{\pm KL}_{tc} = \dot{\beta}^{\pm KL}_{c} - \partial_c\Omega^{\pm KL} + \Omega^{\pm K}{}_{E}\beta^{\pm EL}_{c} - \beta^{\pm K}{}_{Ec}\Omega^{\pm EL}. \tag{2.26}$$

Substituting into the Lagrangian density then yields

$$\begin{aligned}\mathcal{L} =& 2\left(\mathcal{A}\mathcal{K}^{+}_{IJKL} + \mathcal{B}\mathcal{K}^{-}_{IJKL}\right) e^J_a \left(\partial_b\beta^{KL}{}_c + \beta^{KE}{}_b\beta_E{}^L{}_c\right)\tilde{\epsilon}^{abc}\dot{\phi}^I \\ &+ \left(\mathcal{A}\mathcal{K}^{+}_{IJKL} + \mathcal{B}\mathcal{K}^{-}_{IJKL}\right) e^I_a e^J_b \tilde{\epsilon}^{abc}\dot{\beta}^{KL}_c \\ &+ 2\left(\mathcal{A}\mathcal{K}^{+}_{IJKL} + \mathcal{B}\mathcal{K}^{-}_{IJKL}\right)\Omega^I{}_M\phi^M e^J_a \left(\partial_b\beta^{KL}{}_c + \beta^{KE}{}_b\beta_E{}^L{}_c\right)\tilde{\epsilon}^{abc} \\ &+ \left(\mathcal{A}\mathcal{K}^{+}_{IJKL} + \mathcal{B}\mathcal{K}^{-}_{IJKL}\right) e^I_a e^J_b \tilde{\epsilon}^{abc}\left(-\partial_c\Omega^{KL} + \Omega^{KE}\beta_E{}^L{}_c - \beta^{KG}{}_c\Omega_G{}^L\right).\end{aligned} \tag{2.27}$$

where $\mathcal{A}$ and $\mathcal{B}$ are obtained by

$$\mathcal{A} = \frac{\gamma - i}{\gamma} = 2g_+, \qquad \mathcal{B} = \frac{\gamma + i}{\gamma} = 2g_- \tag{2.28}$$

For a better visual experience, the tetrads are not written in their decomposed form. From the above, we note that $(\phi^I, \beta^{\pm IJ}_a)$ appear linearly in their time derivatives and $\Omega^{\pm IJ}$ appears with no time derivative. We can, therefore, deploy the methodology introduced in the previous chapter, using Equation 1.224. We introduce the auxiliary velocity fields $(V^I, V^{\pm IJ}_a)$ which are constrained to be equal to the corresponding Lagrange multiplier $(P_I, P^{\pm a}_{IJ})$, i.e., canonical momenta conjugate to ϕ^I and $\beta^{\pm KL}_c$.

The momentum conjugate to ϕ^I is

$$\begin{aligned}P_I :=& 2\mathcal{A}\,\mathcal{K}^{+}_{IJKL}\left[e^J_a\left(\partial_b\beta^{+KL}{}_c + \beta^{KE}{}_b\beta_E{}^L{}_c\right)\tilde{\epsilon}^{abc}\right] \\ &+ 2\mathcal{B}\,\mathcal{K}^{-}_{IJKL}\left[e^J_a\left(\partial_b\beta^{-KL}{}_c + \beta^{KE}{}_b\beta_E{}^L{}_c\right)\tilde{\epsilon}^{abc}\right].\end{aligned} \tag{2.29}$$

Similarly, the momenta conjugate to the connections are

$$P^{+c}_{KL} := \mathcal{A}\,\mathcal{K}^{+}_{IJKL}\left[e^I_a e^J_b \tilde{\epsilon}^{abc}\right]^{+}, \tag{2.30}$$

and

$$P^{-c}_{KL} := \mathcal{B}\,\mathcal{K}^{-}_{IJKL}\left[e^I_a e^J_b \tilde{\epsilon}^{abc}\right]^{-}. \tag{2.31}$$

The Lagrangian can then be put in the form containing the auxiliary velocities and the corresponding Lagrange multipliers.

$$\begin{aligned}\mathcal{L} =& P_I\dot{\phi}^I + P^{+c}_{KL}\dot{\beta}^{+KL}_c + P^{-c}_{KL}\dot{\beta}^{-KL}_c \\ &+ \Omega^{+KL}\left[\partial_c P^{+c}_{KL} + P^{+c}_{KE}\beta^{+L}_{E\ \ c} - P^{+c}_{EL}\beta^{+E}{}_{Kc} + \left(P_{[K}\phi_{L]}\right)^+\right] \\ &+ \Omega^{-KL}\left[\partial_c P^{-c}_{KL} + P^{-c}_{KE}\beta^{-L}_{E\ \ c} - P^{-c}_{EL}\beta^{-E}{}_{Kc} + \left(P_{[K}\phi_{L]}\right)^-\right] \\ &- V^I C_I - V^{+KL}_c C^{+c}_{KL} - V^{-KL}_c C^{-c}_{KL}. \end{aligned} \tag{2.32}$$

In passing to this form, the terms involving $\partial_c\Omega^{\pm KL}$ have been integrated by parts:

$$-P^{\pm c}_{KL}\partial_c\Omega^{\pm KL} = -\partial_c\left(P^{\pm c}_{KL}\Omega^{\pm KL}\right) + \left(\partial_c P^{\pm c}_{KL}\right)\Omega^{\pm KL}, \tag{2.33}$$

and the total derivative is discarded as a boundary term. We may therefore identify the Lorentz-gauge constraints

$$\mathcal{G}^+_{KL} = \partial_c P^{+c}_{KL} + P^{+c}_{KE}\beta^{+E}{}_{Lc} - P^{+c}_{EL}\beta^{+E}{}_{Kc} + \left(P_{[K}\phi_{L]}\right)^+, \tag{2.34}$$

$$\mathcal{G}^-_{KL} = \partial_c P^{-c}_{KL} + P^{-c}_{KE}\beta^{-E}{}_{Lc} - P^{-c}_{EL}\beta^{-E}{}_{Kc} + \left(P_{[K}\phi_{L]}\right)^- \tag{2.35}$$

which can be put in the more compact form

$$\mathcal{G}^\pm_{KL} = D^{(\beta^\pm)}_c P^{\pm c}_{KL} + \left(P_{[K}\phi_{L]}\right)^\pm. \tag{2.36}$$

Here $D^{(\beta^\pm)}$ is the spatial covariant derivative compatible with β, β^+ and β^-. The remaining primary constraints are

$$C_I = P_I - 2\mathcal{A}\,\mathcal{K}^+_{IJKL}e^J_a R^{+KL}_{bc}\tilde{\epsilon}^{abc} - 2\mathcal{B}\,\mathcal{K}^-_{IJKL}e^J_a R^{-KL}_{bc}\tilde{\epsilon}^{abc}, \tag{2.37}$$

$$C^{+c}_{KL} = P^{+c}_{KL} - \mathcal{A}\,\mathcal{K}^+_{IJKL}\left[e^I_a e^J_b\tilde{\epsilon}^{abc}\right]^+, \tag{2.38}$$

$$C^{-c}_{KL} = P^{-c}_{KL} - \mathcal{B}\,\mathcal{K}^-_{IJKL}\left[e^I_a e^J_b\epsilon^{abc}\right]^-. \tag{2.39}$$

The Lagrangian density therefore takes the canonical form

$$\mathcal{L} = P_I\dot{\phi}^I + P^{+c}_{KL}\dot{\beta}^{+KL}_c + P^{-c}_{KL}\dot{\beta}^{-KL}_c - \widetilde{\mathcal{H}} + \partial_a\widetilde{\ell}^a, \tag{2.40}$$

where $\partial_a\widetilde{\ell}^a$ is a boundary term, and,

$$\widetilde{\mathcal{H}} = -\Omega^{+KL}\mathcal{G}^+_{KL} - \Omega^{-KL}\mathcal{G}^-_{KL} + V^I\widetilde{C}_I + V^{+KL}_c C^{+c}_{KL} + V^{-KL}_c C^{-c}_{KL}. \tag{2.41}$$

2.1.3 Evolution of constraints

We must now require that the set of constraints be preserved under time evolution [26], and consistency requires their time derivatives to vanish weakly on the constraint surface.

$$\dot{\mathcal{G}}^{\pm}_{IJ} \approx 0, \qquad \dot{C}_I \approx 0, \qquad \dot{C}^{\pm a}_{IJ} \approx 0. \tag{2.42}$$

It is useful to recall how the time evolution of any phase-space function is determined. Once the Hamiltonian form of the theory has been obtained, the canonical variables and their momenta obey Hamilton's equations. Therefore, for a quantity f as a function of the phase-space fields, $f = f(q,p)$, the time derivative is fixed by the chain rule,

$$\dot{f} = \frac{\partial f}{\partial q}\dot{q} + \frac{\partial f}{\partial p}\dot{p}. \tag{2.43}$$

Equivalently, using Hamilton's equations, this may be written as

$$\dot{f} = \{f, H\}. \tag{2.44}$$

This is the Poisson bracket defined in Chapter 1. Thus, our goal should be to recover the equations for $\dot{p}$ and $\dot{q}$.

For our purpose, the first step is to classify each type of constraint. As it turns out, $\mathcal{G}^{\pm}_{IJ}$ generate self-dual and anti-self-dual Lorentz transformations, respectively. Their preservation is therefore guaranteed weakly when both constraints are smeared and their Poisson brackets with the remaining constraints close up to boundary terms. The relevant identities, as well as an illustrative example of the computation, are given in the Appendix of [10].

Evolution of the remaining constraints yields

$$\dot{C}_I \approx -V_a^{+KL} W^{a+}_{KLI} - V_a^{-KL} W^{a-}_{KLI}, \tag{2.45}$$

$$\begin{aligned}\dot{C}^{+e}_{IJ} \approx\ & \left[g_+ Y^{de}_{IJKL} + g_+ Y^{de}_{KLIJ}\right]^+ V_d^{+KL} \\ & + \left[g_+ Y^{de}_{IJKL} + g_- Y^{de}_{KLIJ}\right]^+ V_d^{-KL} + W^{e+}_{IJM} V^M,\end{aligned} \tag{2.46}$$

$$\begin{aligned}\dot{C}^{-e}_{IJ} \approx\ & \left[g_- Y^{de}_{IJKL} + g_- Y^{de}_{KLIJ}\right]^- V_d^{-KL} \\ & + \left[g_- Y^{de}_{IJKL} + g_+ Y^{de}_{KLIJ}\right]^- V_d^{+KL} + W^{e-}_{IJM} V^M.\end{aligned} \tag{2.47}$$

The Lorentz tensors appearing in these equations are

$$Y^{de}_{IJKL} = Y^{de}_{[IJ][KL]} := 4\tilde{\epsilon}^{dbe}\epsilon_{MIJ[K}\phi_{L]}e^{M}_{b}, \tag{2.48}$$

$$W^{e\pm}_{IJK} := \Big[2(g_{\mp} - g_{\pm})\tilde{\epsilon}^{ebc}\epsilon_{KMN[I}\phi_{J]}R^{\mp MN}_{bc} + g_{\pm}\tilde{\epsilon}^{ebc}\phi_{K}\epsilon_{IJMN}R^{\pm MN}_{bc}\Big]^{\pm}. \tag{2.49}$$

These equations should be interpreted as the consistency conditions for the multipliers $V^{\pm IJ}_{a}$: the quantities V^{I}, V^{+IJ}_{a}, and V^{-IJ}_{a} entered the Hamiltonian as arbitrary multipliers; therefore, the evolution of the corresponding constraints is interpreted as a consistency condition. Thus, we should determine whether the arbitrariness remains as gauge freedom, is partially fixed, or gives rise to further constraints.

To this end, we note that the first equation is not independent of the other two. Indeed, contracting the second and third preservation equations with V^{+IJ}_{e} and V^{-IJ}_{e}, respectively, gives

$$V^{+IJ}_{e}\dot{C}^{+e}_{IJ} + V^{-IJ}_{e}\dot{C}^{-e}_{IJ} = -V^{I}\dot{C}_{I}. \tag{2.50}$$

Thus, once the preservation of C^{+a}_{IJ} and C^{-a}_{IJ} has been imposed, consistent evolution of C_{I} is guaranteed. This finding is contingent upon V^{I} containing the information associated with the lapse and shift, namely the freedom to choose the foliation of spacetime.

Moreover, recalling the second condition for a well-posed Cauchy problem, we require that each $V^{\pm IJ}_{e}$ be uniquely determined. To attain that goal, we make use of the internal normal vector established previously by projecting the velocity field tangential to it.

$$V^{\pm IJ}_{a} = [N^{[I}\mathcal{V}^{\pm J]}_{a}] \tag{2.51}$$

where $\mathcal{V}^{\pm I}_{a}N_{I} = 0$. The question of establishing uniqueness is, therefore, shifted to the tangential component, $\mathcal{V}^{\pm I}_{a}$ along this normal vector. Projecting the constraint-propagation equations along N^{I}, we obtain

$$0 \approx ig_{+}\tilde{\epsilon}^{dbe}\frac{1}{2}\partial_{b}\phi^{2}\,\eta_{IJ}V^{+J}_{d} + (g_{+} - g_{-})N_{L}N_{J}Y^{de}_{[LI]^{+}[JK]^{-}}\mathcal{V}^{-K}_{d} + W^{e+}_{KIJ}N^{K}V^{J}. \tag{2.52}$$

$$0 \approx -ig_{-}\tilde{\epsilon}^{dbe}\frac{1}{2}\partial_{b}\phi^{2}\,\eta_{IJ}V^{-J}_{d} + (g_{-} - g_{+})N_{L}N_{J}Y^{de}_{[LI]^{-}[JK]^{+}}\mathcal{V}^{+K}_{d} + W^{e-}_{KIJ}N^{K}\mathcal{V}^{J}. \tag{2.53}$$

where we make use of the fact that $\phi_{K}e^{K}_{b} = \frac{1}{2}\partial_{b}\phi^{2}$.

The case $g_{+} = g_{-} = \frac{1}{2}$

As presenting the material lengthily would have impeded comprehension, we may strive to separate the antecedent section into a more accessible and intelligible format for the benefit of both the reader and the author.

We first consider the special case of $g_{+} = g_{-} = \frac{1}{2}$ where the mixed terms in the projected propagation equations vanish. The two chiral sectors, therefore, decouple at the level of the multiplier equations. Further defining

$$R_{bc}^{\pm IJ} = \left[N^{[I} R_{bc}^{\pm J]}\right]^{\pm}, \qquad N_I R_{bc}^{\pm I} = 0, \tag{2.54}$$

the projected equations reduce to

$$0 \approx -\tilde{\epsilon}^{dbe} \partial_b \phi^2 \delta_{IJ} V_d^{+J} + \tilde{\epsilon}^{ebc} R_{Ibc}^{+} \phi_J V^J, \tag{2.55}$$

$$0 \approx -\tilde{\epsilon}^{dbe} \partial_b \phi^2 \delta_{IJ} V_d^{-J} + \tilde{\epsilon}^{ebc} R_{Ibc}^{-} \phi_J V^J. \tag{2.56}$$

where $\delta_{IJ} := \eta_{IJ} + N_I N_J$ is the projector onto the internal subspace orthogonal to N^I. Equations (2.55) and (2.56) are linear inhomogeneous equations for $\mathcal{V}_a^{+I}$ and $\mathcal{V}_a^{-I}$, each of which may be regarded as a 9-dimensional vector. Their uniqueness, therefore, hinges on whether the corresponding coefficient matrix

$$M^{de}{}_{IJ} := \tilde{\epsilon}^{dbe} \partial_b \phi^2 \delta_{IJ}. \tag{2.57}$$

is invertible. As the matrix $M^{de}{}_{IJ}$ has three null vectors of the form

$$\partial_a \phi^2 \, \bar{S}^I, \qquad \bar{S}^I N_I = 0 \tag{2.58}$$

it cannot be invertible. In vector language, the action of this matrix is equivalent to the spatial cross product with $\nabla_a \phi^2$ and therefore by construction annihilates any vector components parallel to $\nabla \phi^2$. A question we then ask is whether the undetermined direction imposes any new constraints on the system.

We can use the fact that for any system, $M\mathcal{V} = \bar{S}$, if M admits any null eigenvectors, then the equation is solvable only if $\bar{S}$ is orthogonal to the corresponding left-null vectors, i.e. the set of all vectors that result in the zero vector upon multiplication with the matrix. We can test this property by contracting the propagation equations with these null eigenvectors

$$\tilde{\epsilon}^{ebc} R_{bc}^{\pm IJ} \partial_e \phi^2 \approx 0. \tag{2.59}$$

However, due to the following identity, we can conclude that Equation 2.59 does not impose any new secondary constraint because it is itself implied by the primary constraints of the system.

$$D_c^{(\beta^{\pm})} C_{IJ}^{\pm c} \approx \mathcal{G}_{IJ}^{\pm} + i g_{\pm} \tilde{\epsilon}^{bca} R_{IJca}^{\pm} \partial_b \phi^2. \tag{2.60}$$

Thus, once the primary constraints hold, the equation (2.59) is already satisfied. The null directions of $M^{de}{}_{IJ}$, instead, indicate that not all components of $\mathcal{V}_a^{\pm I}$ are fixed by the propagation equations.

There are parts of $\mathcal{V}_a^{\pm I}$ that we can determine, however, and they are obtained by projecting orthogonally to $\partial_a \phi^2$. We define

$$P^a{}_b := \delta^a_b - \frac{\partial_b \phi^2 \, \partial^a \phi^2}{\partial_c \phi^2 \, \partial^c \phi^2}, \tag{2.61}$$

where spatial indices are raised with h^{ab}. Acting on the propagation equations with the complementary projector gives

$$\bar{\mathcal{V}}_a^{\pm I} := P^d{}_a \mathcal{V}_d^{\pm I} \approx \frac{2}{\partial_c \phi^2 \, \partial^c \phi^2} \partial^d \phi^2 \, (\phi_J \mathcal{V}^J) R^{\pm I}_{da}. \tag{2.62}$$

where the barred Lagrange multiplier signifies the part that is solved.

The generic case $g_+ \neq g_-$

In contrast to the previous case, the terms proportional to $g_+ - g_-$ mix the self-dual and anti-self-dual multipliers. In this case, in the same manner as the previous case the uniqueness needs to be investigated. In the notation of the projected equations, the relevant matrix to be inverted is

$$N_L N_J Y^{de}_{[LI]^+[JK]^-}. \tag{2.63}$$

Further calculation shows that this matrix is generically invertible and, therefore, all components of $\mathcal{V}_a^{+I}$ and $\mathcal{V}_a^{-I}$ can be determined from the propagation equations. Symbolically, we may write

$$V_a^{\pm IJ} = Z^{\pm IJ}{}_{Ka} V^K, \tag{2.64}$$

where $Z^{\pm IJ}{}_{Ka}$ depends on the fields

$$e_a^I, \qquad R_{ab}^{+IJ}, \qquad R_{ab}^{-IJ}, \qquad \phi^I.$$

Thus, we can conclude that the generic case as well does not generate secondary constraints.

Hamilton's equations of motion

We can determine Hamilton's equations of motion for any value of (g_+, g_-), given by

$$\dot{\phi}^I = V^I - \left(\Omega^{+IJ} + \Omega^{-IJ}\right) \phi_J, \tag{2.65}$$

$$\dot{\beta}_c^{+IJ} \approx D_c^{(\beta^+)} \Omega^{+IJ} + V_c^{+IJ}, \tag{2.66}$$

$$\dot{\beta}_c^{-IJ} \approx D_c^{(\beta^-)} \Omega^{-IJ} + V_c^{-IJ}, \tag{2.67}$$

$$\dot{P}_K \approx D_a^{(\beta)} \text{☿}^a{}_K + \left(\Omega^{+I}{}_K + \Omega^{-I}{}_K\right) \; P_I, \tag{2.68}$$

$$\dot{P}^{+a}_{KL} \approx -4 \left[D_b^{(\beta+)} \left(V^I g_+ \epsilon_{IJKL} \tilde{\epsilon}^{eba} e_e^J\right)\right]^+ - \left(2\Omega^+_{[L}{}^J P^{+a}_{K]J} + \text{☿}^a{}_{[K}\phi_{L]}\right)^+, \tag{2.69}$$

$$\dot{P}^{-a}_{KL} \approx -4 \left[D_b^{(\beta-)} \left(V^I g_- \epsilon_{IJKL} \tilde{\epsilon}^{eba} e_e^J\right)\right]^- - \left(2\Omega^-_{[L}{}^J P^{-a}_{K]J} + \text{☿}^a{}_{[K}\phi_{L]}\right)^-. \tag{2.70}$$

where we define [3]

$$\text{☿}^a{}_K := 2\left[V^I \left(g_+ R^{+JP}_{bc} + g_- R^{-JP}_{bc}\right) + 2\left(g_+ V_b^{+IJ} + g_- V_b^{-IJ}\right) e_c^P\right] \epsilon_{IJKP} \tilde{\epsilon}^{abc}. \tag{2.71}$$

As expected, any additional terms involving the derivatives of the Lagrange multipliers will be proportional to the corresponding constraints and will, therefore, vanish on constraint surfaces.

2.1.4 Algebra of constraints

Having resolutely resolved the propagation of constraints, we can now classify the constraints by the action of their Poisson bracket with other constraints. We denote the dimensionality of the space by $\mathbf{P}$, the number of first-class constraints by $\mathbf{F}$, and the number of second-class constraints by $\mathbf{S}$. The number of degrees of freedom per spatial points is defined to be

$$DOF = \frac{1}{2}(\mathbf{P} - 2\mathbf{F} - \mathbf{S}) \tag{2.72}$$

This number represents the number of freely specifiable data required to determine a physically distinct solution of the equations of motion, where solutions related by gauge transformations are identified. Since the phase space of the theory is, in general, complex, this number should be understood as the number of complex degrees of freedom of the model.

The case $g_+ \neq g_-$

For the general case $g_+ \neq g_-$, the classification of constraints is illustrated in Figure 2.1. Given the classification of constraints, we can now count how many complex degrees of freedom the model possesses. The dimensionality of the phase space per spatial point is

$$\mathbf{P} = 8\,(\phi^I, P_I) + 18\,(\beta_a^{+IJ}, P^{+a}_{IJ}) + 18\,(\beta_a^{-IJ}, P^{-a}_{IJ}) = 44$$

[3] By now, the reader will have observed the significant number of notations employed within these discussions. As the creation of new symbols presented practical challenges, this designation - Neptune- is dedicated to the city where I conducted my doctoral studies.

The number of first-class constraints is

$$\mathbf{F} = 3\,(\mathcal{G}^{+IJ}) + 3\,(\mathcal{G}^{-IJ}) + 4\,(\mathscr{H}^{I}) = 10$$

and the number of second-class constraints is

$$\mathbf{S} = 9\,(C^{+a}_{IJ}) + 9\,(C^{-a}_{IJ}) = 18$$

The number of degrees of freedom per spatial point is therefore

$$DOF = \frac{1}{2}(\mathbf{P} - 2\mathbf{F} - \mathbf{S}) = 3 \tag{2.73}$$

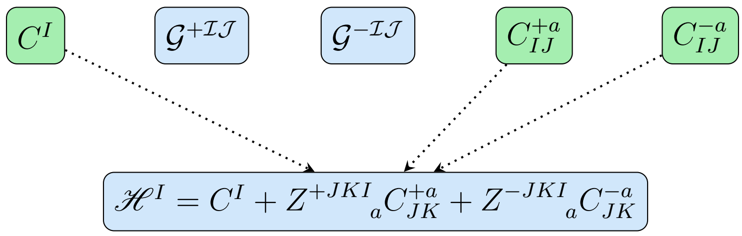


Figure 2.1: The structure of constraints in the case $g_+ \neq g_-$. First-class constraints are blue, whilst constraints that are individually second-class are shown as green. The constraint analysis reveals that a linear combination of second-class constraints yields the first-class constraints $\mathscr{H}^I$.

Case $g_+ = g_-$

For the special case $g_+ = g_-$, the classification of constraints is illustrated in Figure 2.2. As in the previous case, the dimensionality of the phase space per spatial point is

$$\mathbf{P} = 8\,(\phi^I, P_I) + 18\,(\beta^{+IJ}_a, P^{+a}_{IJ}) + 18\,(\beta^{-IJ}_a, P^{-a}_{IJ}) = 44$$

However, now the number of first-class constraints is

$$\mathbf{F} = 3\,(\mathcal{G}^{+IJ}) + 3\,(\mathcal{G}^{-IJ}) + 4\,(\mathscr{H}^{I}) + 3(\partial_a\phi^2\tilde{C}^{+a}_{IJ}) + 3\,(\partial_a\phi^2\tilde{C}^{-a}_{IJ}) = 16$$

and the number of second-class constraints is

$$\mathbf{S} = 6\,(\mathcal{P}^a{}_b C^{+b}_{IJ}) + 6\,(\mathcal{P}^a{}_b C^{-b}_{IJ}) = 12$$

The number of degrees of freedom per spatial point is therefore

$$DOF = \frac{1}{2}(\mathbf{P} - 2\mathbf{F} - \mathbf{S}) = 0 \tag{2.74}$$

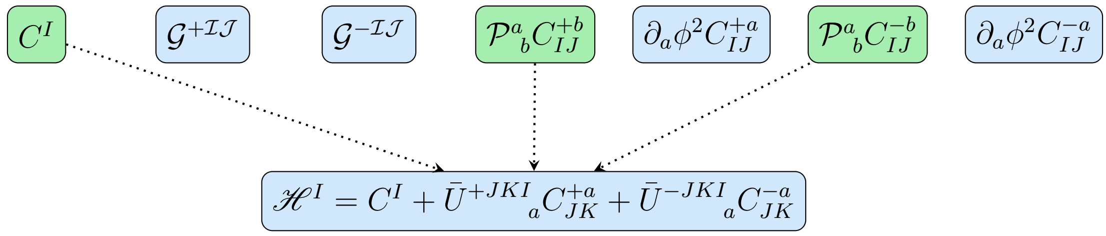


Figure 2.2: The structure of constraints in the case $g_+ = g_-$. First class constraints are blue whilst second class constraints are green. Unlike in the case $g_+ \neq g_-$, a subset of the individual $\tilde{C}^{\pm a}_{IJ}$ constraints are first class. As in the $g_+ \neq g_-$ constraint analysis reveals that a linear combination of individually second class constraints yields the first class constraints $\tilde{\mathscr{H}}^I$. We have defined $\bar{U}^{\pm IJ}{}_{Ka} = \frac{2}{(\partial_b\phi^2\partial^b\phi^2)}\partial^c\phi^2\phi_K R^{\pm IJ}{}_{ca}$

This is a special case in which the theory possesses no degrees of freedom. This is because we have an additional symmetry under the following field transformation

$$\phi^I \to \phi^I, \quad \omega_\mu^{IJ\pm} \to \omega_\mu^{IJ\pm} + \partial_\mu\phi^2\xi^{\pm IJ} \tag{2.75}$$

Therefore, the action changes only by a total derivative.

2.2 Discussion on Reality Conditions and the General Relativity Limit

2.2.1 Reality conditions

The Hamiltonian analysis above was carried out on the complexified phase space. Indeed, because of the inherent complexity of the theory, in principle, the Hamiltonian may generate classical time evolution in which fields become complex even if they are initially real. The choice of complex numbers within the initial data of fields may then result in complex fields in spacetime, such as a complex spacetime metric. The physical requirement is that h_{ij} be real on the initial hypersurface, and that this property be preserved by Hamiltonian evolution. This requirement proves nontrivial. Since the multipliers $V_a^{\pm IJ}$ are determined by the constraint-propagation equations, their dependence on the complex couplings g_+ and g_- can induce an imaginary part in $\dot{h}_{ij}$, even when h_{ij} is initially real. Indeed, recalling the form of the co-tetrads and the spatial metric, using

Hamilton's equations we find schematically

$$\partial_t h_{ij} = 2\eta_{IJ} e^I_i \left(-\partial_j V^J + (Z_j^{+JKL} + Z_j^{-JKL}) V_L \phi_K + \beta^J{}_{Kj} V^K \right). \tag{2.76}$$

Thus, the preservation of the reality of h_{ij} depends on the solved form of $Z_a^{\pm JKL}$. In the generic case this solved form is not simple, and a complete set of reality conditions is correspondingly difficult to state in closed form. The situation simplifies in the chiral limits $(g_+, g_-) = (1, 0)$ and $(g_+, g_-) = (0, 1)$ which correspond to $\gamma^2 = -1$. These are the self-dual and anti-self-dual Ashtekar limits. In these cases, the model makes direct contact with the complex chiral formulation of General Relativity, where the role of reality conditions is already well understood.

The perturbative analysis around a Minkowski background gives a useful diagnostic. For general γ, tensor perturbations do not propagate with the same dynamics as in General Relativity. Only when $\gamma^2 = -1$ does the quadratic action reduce to the standard spin-two propagation of General Relativity [10].

In particular, the Euler-Lagrange equations following from the Lagrangian density in Equation 2.17 have a solution $R^{IJ}{}_{\mu\nu}(\omega) = 0$ [11]. Thus a gauge can be found where $\omega^{IJ}_\mu \stackrel{*}{=} 0$ and $g_{\mu\nu} \stackrel{*}{=} \eta_{IJ}\partial_\mu \phi^I \partial_\nu \phi^J$. If $R^{IJ}{}_{\mu\nu}(\omega) = 0$ then it turns out that ϕ^I is otherwise undetermined by the field equations and hence ϕ^I can take a profile where it forms a set of Minkowski coordinates such that $g_{\mu\nu} = \eta_{\mu\nu}$, i.e., Minkowski space is a solution to the theory for general values of γ. Now we restrict ourselves to a wedge where $\phi^2 = \eta_{IJ}\phi^I\phi^J < 0$ and adopt a Lorentz gauge where $\bar{\phi}^I \stackrel{*}{=} T\delta^I_0$. Then, using T as a time coordinate and denoting x^a as spatial coordinates on the surface of constant time, we have

$$ds^2 = -dT \otimes dT + T^2 \delta_{ij} \mathfrak{E}^i_a \mathfrak{E}^j_b dx^a \otimes dx^b \tag{2.77}$$

where $i, j = 1, 2, 3$ index spatial coordinates, $\mathfrak{E}^1_a dx^a = d\chi, \mathfrak{E}^2_a dx^a = \sinh(\chi) d\theta, \mathfrak{E}^3_a dx^a = \sinh(\chi)\sin(\theta) d\phi$. Therefore, we can identify the region $\phi_I \phi^I < 0$ in Minkowski spacetime coordinatized by (t, x^a) as an open Friedmann-Robertson-Walker universe with scale factor $a = T$. This is a Milne wedge and without loss of generality, we choose the upper Milne wedge of Minkowski spacetime. Consider now the following small perturbations to the spin connection

$$\delta\omega^{0i} = \frac{1}{2}\mathfrak{H}^i{}_j \mathfrak{E}^j \tag{2.78}$$

$$\delta\omega^{ij} = \mathfrak{a}^{ij} dt + \epsilon^{ijl} \mathfrak{W}_l{}^k \mathfrak{E}_k \tag{2.79}$$

where $(\mathfrak{H}^{ij}, \mathfrak{a}^{ij}, \mathfrak{W}^{ij})$ are symmetric and traceless. These perturbations produce a perturbed spatial metric:

$$\delta g_{ab} = \mathfrak{H}^{jk}\mathfrak{E}_{ja}\mathfrak{E}_{kb} \tag{2.80}$$

It can be shown that the perturbation $(\mathfrak{a}_{ij}, \mathfrak{W}_{ij})$ can be solved for algebraically in the field equations in terms of first derivatives of $\mathfrak{H}_{ij}$ and so eliminated from the variational principle to recover the following Lagrangian density

$$\frac{1}{2}\tilde{\mathcal{L}}_{(\mathfrak{H})} = a^3\bigg(-\frac{1}{\gamma^2}\dot{\mathfrak{H}}_{ij}\dot{\mathfrak{H}}^{ij} - \frac{1}{a^2}h^{ab}\mathcal{D}_a\mathfrak{H}^{ij}\mathcal{D}_b\mathfrak{H}_{ij} - \frac{2k}{a^2}\mathfrak{H}_{ij}\mathfrak{H}^{ij}\bigg) \tag{2.81}$$

where h^{ab} is the unperturbed inverse spatial metric on the hypersurface and $\mathcal{D}_a$ is a spatial derivative.

This identifies the chiral cases $(g_+, g_-) = (1, 0)$ and $(0, 1)$, where $\gamma^2 = -1$, as the physically distinguished limits in which the model admits a direct General-Relativistic interpretation. For the case $g_+ = g_-$, corresponding to $\gamma \to \infty$, the kinetic term for $\mathfrak{H}$ vanishes. This coincides with the previous result that the theory has no propagating degrees of freedom.

2.2.2 The chiral limit and the extension of General Relativity

We now focus on the case $(g_+ = 1, g_- = 0)$, understanding that the opposite case is obtained by exchanging the self-dual and anti-self-dual sectors. In this limit, the action depends dynamically only on the self-dual connection ω^{+IJ}. The anti-self-dual connection no longer carries an independent curvature term and is instead fixed through the constraint structure. Indeed, the constraint system can be reduced by solving the second-class constraints associated with the anti-self-dual variables. The remaining canonical structure is then expressed in terms of the self-dual Ashtekar connection and its conjugate momentum, together with the pair associated with ϕ^2. Therefore, as mentioned in the previous section, the case $\gamma^2 = -1$ reproduces the canonical structure of Ashtekar's self-dual formulation of General Relativity, although supplemented by the additional field ϕ^2 [10].

Finally, $\phi^2 := \eta_{IJ}\phi^I\phi^J$ remains in the canonical description and behaves as an additional scalar clock variable. In the reduced Hamiltonian formulation, this field appears in a way analogous to a pressureless dust time variable when $\phi^2 < 0$.

Thus, the chiral model reproduces the gravitational sector of self-dual General Relativity, but extends it by retaining the scalar degree of freedom, since, in general, the magnitude of ϕ^2 varies throughout spacetime.

Chapter 3

Prelude to Cosmology: Geometry, Expansion, and Observables

And that inverted Bowl we call The Sky, Whereunder crawling coop't we live and die, Lift not thy hands to it for help – for It Rolls impotently on as Thou or I.
— Omar Khayyám

We are quite obliged to observe that the Universe has become increasingly spacious, ostentatiously making up for lost time!

As of this writing, we happen to be only three years away from the centenary of Edwin Hubble's detection of the Universe's expansion. Although there are many ways to approach the past 13.8 billion years, here we aim to lay the groundwork for placing the previous chapters within a broader cosmological framework. Building upon the cosmological principle on a smooth, homogeneous, and isotropic background, we will then employ perturbation theory to discuss the early stages of the Universe. Finally, we will conclude this chapter by introducing some of the key observational methods, namely, weak lensing and gravitational waves, crucial for probing the distribution of matter and large-scale structure.

3.1 Cosmology at Large Scales

Although the distribution of matter is highly structured on galactic and cluster scales, observations indicate that this complexity becomes statistically smoother when averaged over sufficiently large distances [27]. This behavior is formulated as the cosmological principle: on large spatial scales, the Universe is homogeneous and isotropic. In other words, no position or direction is preferred over others.

At the time when Einstein first applied general relativity to cosmology, the prevailing view was that the Universe should be static on large scales. This expectation led him to introduce the cosmological constant, whose role was to balance the gravitational attraction of matter and allow a stationary cosmological solution. However, a few years later, Aleksandr Aleksandrovich Friedmann demonstrated that depending on the matter content and spatial curvature [28], the Universe may indeed expand, contract, or remain static.

Modern cosmology today advocates for an expanding universe, the rate of which is inferred observationally through the relation between distance and redshift. This is, of course, much easier said than done. Measuring distances is among the central difficulties of astronomy: we must first identify objects whose intrinsic luminosity or characteristic scale is known, and then compare this intrinsic quantity with what is observed. Once distances to galaxies are estimated, their spectra show a systematic redshift, indicating that distant galaxies are receding from us with a velocity approximately proportional to their distance.

Admittedly, general relativity cautions us that spacetime can deviate from flat geometry. Fortunately, we need very little of the full theory here, and the only allowed dynamical degree of freedom in the background geometry is the time-dependent scale factor $a(t)$. We may think of the expansion as assigned to the evolution of spatial distances themselves. For an object with comoving position $\mathbf{x}$, the physical position is

$$\mathbf{x}_{\mathrm{phys}}(t) = a(t)\mathbf{x}. \tag{3.1}$$

Differentiating then gives

$$\mathbf{v}_{\mathrm{phys}} = \dot{a}\,\mathbf{x} + a\dot{\mathbf{x}} \tag{3.2}$$

which can be written as

$$\mathbf{v}_{\mathrm{phys}} = H\mathbf{x}_{\mathrm{phys}} + \mathbf{v}_{\mathrm{pec}}, \tag{3.3}$$

where

$$H(t) = \frac{\dot{a}(t)}{a(t)} \tag{3.4}$$

is the Hubble parameter, and $\mathbf{v}_{\mathrm{pec}} = a\dot{\mathbf{x}}$ is the peculiar velocity. The first term is the velocity associated with the homogeneous expansion of the background spacetime. The second term describes motion relative to the cosmological frame, usually caused by local gravitational interactions.

For nearby galaxies, peculiar velocities can be comparable to the velocity due to expansion. On sufficiently large scales, the expansion term dominates. Evaluating the Hubble parameter today and neglecting peculiar velocities gives Hubble's law,

$$v_{\mathrm{phys}} = H_0 d. \tag{3.5}$$

To get a rather rough estimate, the inverse of the Hubble parameter defines a characteristic cosmological time scale,

$$t_H = H_0^{-1}. \tag{3.6}$$

Using the observed value $H_0 = 68.4^{+1.0}_{-0.8}\mathrm{km\,s^{-1}\,Mpc^{-1}}$ [29], sets a cosmological distance scale[1]at,

$$d_H \sim 10^{26}\,\mathrm{m}. \tag{3.7}$$

Of course, more accurate notions of distance and horizon size require the full expansion history as we will discuss shortly.

3.1.1 The Friedmann–Lemaître–Robertson–Walker (FLRW) Universe

Generally, the metric is obtained by solving the Einstein equations. However, the FLRW metric is traditionally derived by geometrical considerations and the notion of cosmic time provided by the Weyl postulate: "... the world lines of the stars [in later contexts, galaxies] form a sheaf [bundle], which rises in a given direction from the infinitely distant past, and spreads out over the hyperboloid [representing de Sitter's model] in the direction of the future, getting broader and broader converging towards the past." [30]

To describe an expanding universe, the Minkowski metric, as we have encountered before, is appropriate for a local description of spacetime, but it needs to be modified to capture the expansion effect. We replace the flat spatial metric dx^2 with a homogeneous isotropic metric candidate - preferably one supported by observational data - which is then multiplied by a time-dependent scale factor $a(t)$. The resulting spacetime metric, known as the FLRW metric, takes the form

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2/R^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right]. \tag{3.8}$$

where the time coordinate t is cosmic time. It is the proper time measured by observers at rest with respect to the average cosmic expansion. The factor $k = +1, 0, -1$ labels the three possible spatial geometries, and the coordinates (r, θ, ϕ) are comoving coordinates. The physical distance between two comoving points is proportional to the scale factor. Schematically,

$$d_{\mathrm{phys}}(t) = a(t)d_{\mathrm{co}}. \tag{3.9}$$

Thus the expansion of the Universe is encoded geometrically in the growth of $a(t)$. Com-

[1]To aid the intuition about this magnitude Douglas Adams once offered the following insight: " Gigantic multiplied by colossal multiplied by staggeringly huge is the sort of concept we are trying to get across here."

monly, we set $a(t_0) = 1$, for t_0 denoting the present time. With this convention, earlier times correspond to $a(t) < 1$.

Having established the dynamical degree of freedom, the next natural question would be to investigate its time evolution. From the perspective of general relativity, the answer is provided by Einstein's field equations, but useful intuition can already be obtained from Newtonian arguments.

Consider a homogeneous spherical region of physical radius $R(t)$, filled with matter density $\rho(t)$. A test particle of mass m placed on the boundary feels only the gravitational attraction of the mass enclosed inside the sphere. This follows from the Newtonian shell theorem; in the relativistic treatment, the analogous statement is related to Birkhoff's theorem. Given the enclosed mass of the sphere

$$M = \frac{4\pi}{3} R^3 \rho \tag{3.10}$$

the Newtonian equation of motion for the test particle is

$$\frac{d^2R}{dt^2} = -\frac{GM}{R^2}. \tag{3.11}$$

Multiplying by $\frac{dR}{dt}$ on both sides and integrating gives an energy equation per unit mass,

$$\frac{1}{2}\dot{R}^2 - \frac{GM}{R} = U, \tag{3.12}$$

where U is an integration constant. The first term is the kinetic energy per unit mass associated with the expansion, while the second term is the gravitational potential energy per unit mass. The sign of U determines the qualitative behaviour of the solution:

$$\begin{cases} U > 0 & \Rightarrow \quad \text{Expansion can continue indefinitely,} \\ U < 0 & \Rightarrow \quad \text{Expansion may stop and reverse,} \\ U = 0 & \Rightarrow \quad \text{Assymptotic expansion at infinity.} \end{cases}$$

To connect this with cosmology, we may write the physical radius as

$$R(t) = a(t)R_0, \tag{3.13}$$

where R_0 is a fixed comoving radius. Substituting this into the energy equation gives

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{2U}{R_0^2 a^2}. \tag{3.14}$$

which already has the familiar structure of the Friedmann equation. In the relativistic theory, the integration constant is reinterpreted geometrically as the spatial curvature term. The Friedmann equation is then

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{3.15}$$

in units where $c = 1$, with the curvature scale absorbed into the definition of the spatial coordinate. The relativistic derivation provides an understanding of the integration constant in terms of spatial geometry following from Einstein's equations

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{3.16}$$

applied to the FLRW metric for a homogeneous and isotropic stress-energy tensor. The symmetry assumptions force the matter content to take the form of a perfect fluid, characterized at the background level by an energy density $\rho(t)$. Therefore, the formalism can be applied to relativistic particles.

Observational evidence seems to strongly favor a spatially flat universe, $k = 0$ [31]. Therefore, by evaluating the Hubble parameter today, we can estimate the critical density of matter as

$$\begin{aligned}\rho_c &\sim 5.1\,\text{protons}\,\text{m}^{-3},\\ &\sim 1.3\times 10^{11}\,M_\odot\,\text{Mpc}^{-3}.\end{aligned} \tag{3.17}$$

This amounts to approximately 5 protons per cubic meter, assuming we have distributed the matter content of the Universe uniformly, or alternatively, one galaxy per unit volume with sides equal to $Mpc = 3.08 \times 10^{22} m$. Incidentally, this unit measures the average distance between galaxies, which is the reason for its popular use in astronomy. Thus, the large-scale universe is very dilute.

The Friedmann equation alone does not form a complete dynamical system. To solve for $a(t)$, we must also know how the energy density changes as the Universe expands. This requires the conservation equation in an expanding background,

$$\dot{\rho} + 3H(\rho + P) = 0. \tag{3.18}$$

which corresponds to a fluid with an equation of state

$$P = \omega\rho, \tag{3.19}$$

that integrates to

$$\rho(a) = \rho_0 a^{-3(1+\omega)}. \tag{3.20}$$

where different components evolve differently. Specifically, we are referring to ordinary matter, dark matter, radiative elements, and dark energy. As we saw earlier, the ordinary matter makes up a very small fraction of the Universe, about 5% to be precise. The greater matter contribution comes from the elusive entity designated dark matter, making up roughly 27% of the matter content. The term was first coined by Fritz Zwicky, the visionary Swiss astronomer whose widely recognized incensed portrait continues to provide subtle comfort to researchers in this field. As we regress to earlier times, the energy density of the Universe becomes dominated by relativistic components. This is to reference the so-called radiation era. Finally, there is dark energy (?) [32][2]

We can now turn to the evolution of the specific constituents. Here, we note that a more formal derivation of the following segment is possible; however, the intuitive grasp is already quite clear. Consider a fixed comoving box filled with some component of the cosmic fluid. Given the assumption above, its physical volume grows as

$$V(t) \propto a^3(t). \tag{3.21}$$

The energy density is thus obtained as

$$\rho = \frac{E}{V}. \tag{3.22}$$

Pressureless matter

For pressureless non-relativistic matter, with an equation of state $\omega = 0$, the energy of each particle is dominated by its rest mass and, therefore, does not change as the volume increases. The density is thus obtained by

$$\rho_m = \frac{E}{V} \propto a^{-3}. \tag{3.23}$$

Substituting into the flat Friedmann equation and integrating, we obtain the time dependence of the scale factor for a matter-dominated universe.

$$\left(\frac{\dot{a}}{a}\right)^2 \propto a^{-3} \implies a(t) \propto t^{2/3} \tag{3.24}$$

Radiation

Radiation corresponds to relativistic particles such as photons with the equation of state $\omega = 1/3$. As such, it presents a slightly more sophisticated scenario. Indeed, as we increase the size of the box, the wavelength of each particle also stretches out. From

[2]It may amuse the reader that the foundational Figure 1 of the paper was inexplicably rendered in Comic Sans!

quantum mechanics, we know that the energy is inversely proportional to wavelength $E_\gamma \propto \lambda^{-1}$, therefore, we have an additional power of the scale factor in the denominator.

$$\rho_r \propto a^{-3}a^{-1} = a^{-4}. \tag{3.25}$$

Using the Friedmann equation gives

$$\left(\frac{\dot{a}}{a}\right)^2 \propto a^{-4} \implies a(t) \propto t^{1/2}. \tag{3.26}$$

This is to say, radiation dilutes faster than matter. Although its contribution is small today, it was dominant at sufficiently early times.

Dark energy and the cosmological constant

The late-time universe is dominated by a component whose energy density changes very slowly with expansion. We can consider the simplest form of such constituents: the cosmological constant. In a universe dominated by a cosmological constant, the Friedmann equation becomes

$$H^2 = \frac{8\pi G}{3}\rho_\Lambda = \text{constant}. \tag{3.27}$$

Thus

$$\frac{\dot{a}}{a} = H = \text{constant}, \tag{3.28}$$

and the scale factor grows exponentially,

$$a(t) \propto e^{Ht}. \tag{3.29}$$

The solution for a cosmological constant with equation of state $\omega_\Lambda = -1$ is known as de Sitter space. This produces accelerated expansion. Table 3.1 gather the results of the discussions above.

Era	$\boldsymbol{\omega}$	$\boldsymbol{\rho(a)}$	$\boldsymbol{a(t)}$	$\boldsymbol{a(\tau)}$
Matter	0	$\rho_m \propto a^{-3}$	$a \propto t^{2/3}$	$a \propto \tau^2$
Radiation	$\frac{1}{3}$	$\rho_r \propto a^{-4}$	$a \propto t^{1/2}$	$a \propto \tau$
Cosmological constant	-1	$\rho_\Lambda = \text{constant}$	$a \propto e^{Ht}$	$a \propto -\tau^{-1}$

Table 3.1: Single-component solutions of the flat Friedmann equation. For a perfect fluid with constant equation-of-state parameter w, the density scales as $\rho \propto a^{-3(1+w)}$. For $w \neq -1$, one obtains $a(t) \propto t^{2/[3(1+w)]}$ and $a(\tau) \propto \tau^{2/(1+3w)}$. τ denotes conformal time.

At this juncture, we should backtrack a little to emphasize the observational data that supports this line of reasoning. For more than two decades, we have had strong evidence that the Universe is going through an accelerated expansion phase. The 2011 Nobel Prize in Physics was awarded to Saul Perlmutter, Brian P. Schmidt, and Adam G. Riess for discovering that the expansion of the Universe is accelerating through observations of distant type Ia supernovae. Notably, in a purely decelerating universe, the relation between luminosity distance and redshift would follow a different curve. The observed supernova residuals indicated that distant supernovae were dimmer than expected in a matter-only universe, implying that the expansion has accelerated at late times, as illustrated in Figure 11 of [33]. Since April 3^{rd} 2024, there have been new results from the Dark Energy Spectroscopic Instrument (DESI) collaboration on Baryon Acoustic Oscillations (BAO) concerning the evolution of dark energy that may hint at the slowing down of the expansion of the Universe [34]. The data from the Dark Energy Survey (DES) observations seemed to back up DESI's data [35].

The second Friedmann equation concerning the acceleration equation - more formally obtained from the spatial part of the Einstein equations $G^i{}_j = (8\pi G)T^i{}_j$ - is given by

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P), \tag{3.30}$$

Otherwise known as the Raychaudhuri equation, it implies that, with the exception of the de Sitter solutions, all solutions above admit singularities in the past as long as the matter obeys strong energy conditions (SEC), $0 \leq \rho + 3P$ so that $\ddot{a} \leq 0$. Famously proven by Hawking and Penrose [36], this singularity is a generic feature of all Big Bang cosmologies. To more clearly characterize the behavior of the scale factor, we might expand the Hubble parameter as

$$a(t) = 1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0^2 (t - t_0)^2 + \cdots . \tag{3.31}$$

where the dimensionless quantity

$$q(t) = -\frac{\ddot{a}a}{\dot{a}^2} = -\frac{\ddot{a}}{aH^2} \tag{3.32}$$

is the deceleration parameter. It measures whether the expansion is slowing down or speeding up. If $q > 0$ then $\ddot{a} < 0$ and the expansion is decelerating. The opposite conclusion is drawn for the opposite behavior.

3.2 The Hot Universe and Primordial Expansion

Having surreptitiously inserted the word Big Bang into the preceding discussion, we shall now explicate the concept in significantly greater detail. The term itself was minted by Fred Hoyle during his BBC interview, where he deemed the idea to be "unsatisfactory".

The moment is, in truth, too far in the past for our knowledge to have a satisfactory grasp on. The concept of the Big Bang, on the other hand, was proposed by Georges Lemaître, who, in a letter to Einstein, famously wrote: "Your theory suggests there was a day without a yesterday". Indeed, if the Universe is expanding, then reverse engineering seems to point towards past epochs where conformal distances were less than they are today.

The moment zero being currently inaccessible, let us start shortly after. At $t = 10^{-19}$s, we would find ourselves in a universe with almost equal amounts of matter and antimatter, which is reasonable, since annihilation would ensure radiative dominance. However, for a yet unconfirmed reason, a fraction of matter survived the annihilation [37], enabling the future of cosmic evolution. Taking a logarithmic leap in time would bring us to the Quantum Chromodynamics (QCD) phase transition at $t = 10^{-8}s$ [38], when quarks and gluons became confined into hadrons. By the time the Universe is $1s$ old, it is "cool"[3] enough for neutrinos to decouple due to the weak scale interactions becoming ineffective at keeping them inside the thermal bath, thereby producing a cosmological neutrino background [39]. In fact, the strength of the weak interaction is highly dependent on temperature and weakens as it decreases. An epochal span later, at $t = 3\ min$, we expect the formation of light elements such as hydrogen and helium. Arguably, one of the signal achievements of modern cosmology is to predict the relative abundance of these particles via the Big Bang Nucleosynthesis (BBN) [40]. For instance, there is roughly a three-to-one ratio between hydrogen and helium. After a perceived infinity, when the Universe is $370,000$ years old, the temperature finally falls below the ionization energy scale, making the formation of neutral atoms possible [41]. This moment is referred to as recombination. Incidentally, it also marks the moment when photons can escape the matter plasma to make the Cosmic Microwave Background (CMB). Finally, in what seemed to be yesterday, matter yielded to gravity to form the first stars.

Notwithstanding its initial clarity, the hot Big Bang model alone proves insufficient in explaining the Universe we currently observe. In the subsequent section, we will point out some of the most important shortcomings of this model that will ultimately lead to the hypothesis of a primordial accelerated expansion.

3.2.1 The Flatness and Horizon Problems

The Friedmann equations provide the starting point for identifying two of the central puzzles of the standard hot Big Bang picture. For simplicity, in this section, we reduce the discussion to units in which $8\pi G = M_{\mathrm{Pl}}^{-2} = 1$ and $c = \hbar = 1$. The Friedmann equation

[3]The interpretive responsibility of this word rests solely on the reader

may then be written as

$$H^2 = \frac{1}{3}\rho - \frac{k}{a^2}. \tag{3.33}$$

The critical density is defined by $\rho_{\text{crit}} = 3H^2$ and the density parameter by $\Omega = \frac{\rho}{\rho_{\text{crit}}}$. Dividing Equation 3.33 by H^2 gives

$$1 - \Omega = -\frac{k}{(aH)^2}. \tag{3.34}$$

Thus, the deviation from spatial flatness is controlled by the expression $(aH)^{-1}$, called the comoving Hubble radius. For a perfect fluid with equation of state 3.20, the comoving Hubble radius evolves as

$$(aH)^{-1} \propto a^{\frac{1+3w}{2}}. \tag{3.35}$$

Therefore, for all conventional matter satisfying $w > -\frac{1}{3}$, the comoving Hubble radius increases with time. It follows from Eq. (3.34) that any small deviation from $\Omega = 1$ also grows with time. Observations indicate that the total density parameter today is close to unity, $\Omega_0 \simeq 1$. However, in a universe filled only with ordinary matter or radiation, $\Omega = 1$ is an unstable fixed point. A small early deviation from flatness would be amplified during cosmic evolution. The observed near-flatness today, therefore, requires the early universe to have been extraordinarily close to $\Omega = 1$. This fine-tuning issue is referred to as the flatness problem.

The same behavior of the comoving Hubble radius leads to the horizon problem. Let us define the conformal time as

$$\tau = \int^t \frac{dt'}{a(t')}. \tag{3.36}$$

For a universe dominated by a fluid with constant ω, we find

$$\tau \sim \int da\, (a^2 H)^{-1}. \tag{3.37}$$

When $w > -1/3$, tracing τ back to the initial singularity, we find regions on the surface of last scattering whose past light cones do not overlap. This is problematic when applied to the cosmic microwave background because the CMB is observed to have an almost uniform temperature across the sky, with fluctuations of order $\frac{\Delta T}{T} \sim 10^{-5}$. Yet, in the standard decelerating hot Big Bang picture, the sky could have been divided into approximately 10^4 causally disconnected patches with no proposed mechanism within this framework that could have equilibrated their temperatures.

Yet, from a skeptical perspective, one might argue that homogeneity and flatness are merely the simplest aesthetically pleasing scenarios that any unknown theory could have picked out. What we cannot dismiss, however, is that the Universe contains density fluctuations that are highly correlated over distances that could not have been causal. As David Tong eloquently put it: "These detailed correlations make it more difficult to appeal to a creator without sounding like a young earth creationist arguing that the fossil record was planted to deceive us".

3.2.2 Theory of inflation

Revisiting the Big Bang puzzles

Ideated by physicists in the 1980s, cosmological inflation is a hypothesis that addresses the causality issues mentioned in the previous section and posits a phase before the hot Big Bang scenario in which the homogeneity of the Universe was generated [6, 42].

Given that both the flatness and horizon problems arise because ordinary matter and radiation obey an equation of state $w > -\frac{1}{3}$ that satisfies $\rho + 3P > 0$, the second Friedmann equation

$$\frac{\ddot{a}}{a} = -\frac{1}{6}(\rho + 3P) \tag{3.38}$$

would imply a decelerated expansion with $\ddot{a} < 0$. Inflation reverses this condition. If the early universe underwent a phase with $w < -\frac{1}{3}$, following the same line of reasoning results in an accelerated expansion with $\ddot{a} > 0$ and the decrease of the comoving Hubble radius $\frac{d}{dt}(aH)^{-1} < 0$. Firstly, from Eq. (3.34), we see that if $(aH)^{-1}$ decreases, then $|1 - \Omega|$ is driven toward zero. Inflation, therefore, makes spatial flatness an attractor rather than an unstable condition. Secondly, accelerated expansion changes the causal history of the observable universe. Regions that appear widely separated on the CMB sky can originate from a single causally connected patch before inflation. During inflation, this patch is stretched to scales much larger than the Hubble radius. After inflation ends, the usual radiation- and matter-dominated expansion begins, and these regions later re-enter the horizon. The observed uniformity of the CMB can then be understood as the result of causal contact in the pre-inflationary phase. In conformal time, this corresponds to pushing the initial time effectively toward

$$\tau_{\text{initial}} \to -\infty, \tag{3.39}$$

as in quasi-de Sitter expansion. In other words, the causal structure is modified so that the past light cones of widely separated observed regions can intersect at sufficiently early times.

Inflation also explains how correlations can exist on scales that appear superhorizon at recombination. The key point is the distinction between a fixed comoving scale and the time-dependent comoving Hubble radius. Consider a perturbation with fixed comoving wavelength λ. During inflation, $(aH)^{-1}$ decreases, so a scale that initially lies inside the

Hubble radius can later exit it:

$$\lambda < (aH)^{-1} \quad \longrightarrow \quad \lambda > (aH)^{-1}.$$

After inflation ends, the Universe enters the radiation- and matter-dominated eras, during which the comoving Hubble radius grows. The same mode can then re-enter the Hubble radius at a later time.

Physics of inflation

It is not physically tenable to invoke an arbitrary amount of inflation, as previously mentioned issues impose specific constraints on the dynamics of the inflationary epoch. The amount of inflation is typically measured by the number of e-folds,

$$N_{\text{tot}} \equiv \ln\left(\frac{a_e}{a_i}\right), \tag{3.40}$$

where a_i and a_e denote the scale factor at the beginning and end of inflation. During quasi-de Sitter expansion, the physical Hubble rate is nearly constant,

$$H_i \simeq H_e. \tag{3.41}$$

The decrease of the comoving Hubble radius is then essentially the increase of the scale factor

$$\frac{(a_i H_i)^{-1}}{(a_e H_e)^{-1}} \simeq \frac{a_e}{a_i} = e^{N_{\text{tot}}}. \tag{3.42}$$

A conservative requirement for solving the horizon problem is that the present comoving Hubble radius should have been inside the comoving Hubble radius at the beginning of inflation:

$$(a_0 H_0)^{-1} < (a_i H_i)^{-1}. \tag{3.43}$$

This condition ensures that the entire presently observable universe originated from a region that could have been causally connected before inflation.

The precise lower bound on N_{tot} depends on the energy scale of inflation and on the subsequent reheating history [43]. If reheating occurs efficiently and the Universe becomes radiation-dominated at temperature T_R, a typical estimate gives

$$N_{\text{tot}} \gtrsim 64 + \ln\left(\frac{T_R}{10^{15}\,\text{GeV}}\right). \tag{3.44}$$

Under this assumption, solving the horizon problem requires roughly

$$N_{\text{tot}} \sim 50 - 60 \tag{3.45}$$

e-folds of inflation [44]. Fewer e-folds may be sufficient for lower reheating temperatures, while the precise value depends on the details of the post-inflationary thermal history.

Slow-roll dynamics

The kinematic discussion above establishes that a viable inflationary epoch necessitates a sufficient temporal extent. This requires the background quantities to vary slowly, even though the scale factor grows rapidly. The first Hubble slow-roll parameter,

$$\epsilon = -\frac{\dot{H}}{H^2} = -\frac{d\ln H}{dN}, \tag{3.46}$$

measures the fractional change of H per e-fold of expansion. The condition $\epsilon < 1$ ensures accelerated expansion, while the stronger condition, $\epsilon \ll 1$, corresponds to a nearly constant Hubble rate, and the limiting case $\epsilon \to 0$, results in the exact de Sitter expansion. A realistic inflationary scenario, however, cannot remain exactly de Sitter, since inflation must eventually end. The relevant background, usually called quasi-de Sitter, ensures that H is nearly constant for many Hubble times, but changes slowly enough to allow a graceful exit into the hot Big Bang phase. Then, the persistence of inflation is controlled by the variation of ϵ expressed as

$$\eta \equiv \frac{d\ln\epsilon}{dN} = \frac{\dot{\epsilon}}{H\epsilon}. \tag{3.47}$$

where the condition, $|\eta| \ll 1$, ensures that ϵ itself changes slowly from one e-fold to the next. The conditions on both ϵ and η are geometric conditions and model-independent.

In addition, the inflationary phase provides one of the predominant mechanisms that account for large-scale structure. In fact, the wavelength of the primordial quantum fluctuations gets stretched out during the inflationary epoch and ultimately results in classical fluctuations responsible for structure formation. In the next section, we will delineate the physics of the cosmological perturbations.

3.2.3 The perturbed universe: Scalar-Vector-Tensor decomposition

As was mentioned before, observations of the CMB reveal a high degree of homogeneity across the sky characterized by a remarkably small magnitude of perturbations. Consequently, an analytical strategy would then consist of decomposing physical quantities into homogeneous background components and perturbation-dependent terms. For any relevant cosmological quantity $X(t, \mathbf{x})$, such as the metric, matter density, pressure, or matter fields, we can write

$$X(t, \mathbf{x}) = \bar{X}(t) + \delta X(t, \mathbf{x}), \tag{3.48}$$

where $\bar{X}(t)$ is the homogeneous background quantity and $\delta X(t, \mathbf{x})$ is the perturbation. Linear perturbation theory assumes $|\delta X| \ll |\bar{X}|$.

It is important to note that, although we may choose a given family of comoving observers common to the background, in the perturbed universe, there is no unique way to define the threading of spacetime. For instance, a time coordinate shift $t \to \tilde{t} = t + \alpha(t, \mathbf{x})$ can generate an apparent density perturbation $\bar{\rho}(t) \to \bar{\rho}(\tilde{t}) = \bar{\rho}(t) + \dot{\bar{\rho}}\, \alpha$ even in an exactly homogeneous universe [45]. Conversely, a real density perturbation can be removed locally by choosing hypersurfaces of constant density. For this reason, one must either fix a gauge carefully or work with gauge-invariant combinations of perturbations.

To further explicate this, we note that to define a perturbation, we must decide which point of the perturbed spacetime is to be compared with which point of the background spacetime. More specifically, suppose a background point $\bar{P}$ is assigned coordinates x^μ. A gauge choice selects a point $\hat{P}$ in the perturbed spacetime with the same coordinate labels and then defines the perturbation of a quantity X by $\delta X(P) = X(P) - \bar{X}(\bar{P})$, then a different gauge choice selects a nearby point $\tilde{P}$ in the perturbed spacetime to correspond to the same background point $\bar{P}$. The difference between $\hat{P}$ and $\tilde{P}$ is then described by an infinitesimal vector field $\xi^\mu(x)$ which itself lives on the background.

Suppose the quantity in question is a scalar denoted here as s. We may find the relation between $s(\tilde{P})$ and $s(\hat{P})$ by performing a first-order Taylor expansion of the scalar field between the two nearby points [46]

$$s(\tilde{P}) = s(\hat{P}) + \left.\frac{\partial s}{\partial x^\alpha}\right|_{\hat{P}} \left[x^\alpha(\tilde{P}) - x^\alpha(\hat{P})\right]. \tag{3.49}$$

where $\xi^\alpha = x^\alpha(\tilde{P}) - x^\alpha(\hat{P})$. Thus, $\partial_\alpha s\ \xi^\alpha$ quantifies the changes in the scalar field when we move from $\tilde{P}$ to $\hat{P}$. Additionally, in cosmology, the background scalar is only time-dependent. Thus the shift becomes

$$\partial_\alpha \bar{s} \xi^\alpha = \dot{\bar{s}} \xi^0 \tag{3.50}$$

where we approximate $\partial_\alpha s(\hat{P}) = \partial_\alpha \bar{s}(\bar{P})$. Returning to the example above regarding density perturbation, we see that comparing the perturbed density on a slightly different time slice automatically results in sampling a different value of the background density.

Geometrically, an infinitesimal gauge transformation is generated by a vector field ξ^μ.

$$\tilde{x}^\mu = x^\mu + \xi^\mu. \tag{3.51}$$

In other words, ξ^μ - tangent to the integral curves of the vector field - creates a diffeomorphism describing how points are displaced on the manifold. To connect with the material of the first chapter, diffeomorphisms on a manifold form a group, and infinitesimal changes generate the transformation. Thus, at first order, any tensorial object's perturbation is controlled by the Lie derivative of the corresponding background tensor. For any tensor T

$$T = \bar{T} + \delta T, \tag{3.52}$$

and with the convention in Eq. (3.51),

$$\delta\tilde{T} = \delta T - \mathcal{L}_\xi \bar{T}. \tag{3.53}$$

Furthermore, the symmetries of spatially flat background spacetime allow for the decomposition of quantities into independent scalar, vector, and tensor components. Given that these modes decouple and do not mix in the linearly perturbed equation of motion, we can perform the perturbative analysis on each object separately. The methodology for finding the expressions for the perturbation is identical to that in the scalar case. Taking into account the same consideration for the background and the displacement, we find the following results for each case.

For a scalar $s = \bar{s} + \delta s$,

$$\delta\tilde{s} = \delta s - \xi^\alpha \partial_\alpha \bar{s}. \tag{3.54}$$

For a vector $v^\mu = \bar{v}^\mu + \delta v^\mu$,

$$\delta\tilde{v}^\mu = \delta v^\mu - \xi^\alpha \partial_\alpha \bar{v}^\mu + \bar{v}^\alpha \partial_\alpha \xi^\mu. \tag{3.55}$$

For a covariant rank-two tensor $B_{\mu\nu} = \bar{B}_{\mu\nu} + \delta B_{\mu\nu}$,

$$\delta\tilde{B}_{\mu\nu} = \delta B_{\mu\nu} - \xi^\rho \partial_\rho \bar{B}_{\mu\nu} - \bar{B}_{\rho\nu}\partial_\mu \xi^\rho - \bar{B}_{\mu\rho}\partial_\nu \xi^\rho. \tag{3.56}$$

For a mixed tensor $A^\mu{}_\nu = \bar{A}^\mu{}_\nu + \delta A^\mu{}_\nu$,

$$\delta\tilde{A}^\mu{}_\nu = \delta A^\mu{}_\nu - \xi^\rho \partial_\rho \bar{A}^\mu{}_\nu + \bar{A}^\rho{}_\nu \partial_\rho \xi^\mu - \bar{A}^\mu{}_\rho \partial_\nu \xi^\rho. \tag{3.57}$$

Forecasting the direction of future analysis, we shall more carefully examine the mixed tensor case. A mixed rank-two tensor compatible with spatial isotropy must have a spatial part proportional to the identity

$$\bar{A}^{\alpha}{}_{\beta} = \begin{pmatrix} \bar{A}^{0}{}_{0} & 0 \\ 0 & \frac{1}{3}\delta^{i}{}_{j}\,\bar{A}^{k}{}_{k} \end{pmatrix}. \tag{3.58}$$

The factor 1/3 appears because $\bar{A}^{k}{}_{k}$ denotes the spatial trace. Hence the spatial background tensor is distributed equally along the three spatial directions. Considering the perturbed tensor, we note that the spatial perturbation $\delta A^{i}{}_{j}$ may be split into its trace and traceless parts

$$\delta A^{k}{}_{k} = \delta^{j}{}_{i}\,\delta A^{i}{}_{j}, \tag{3.59}$$

and

$$\delta \tilde{A}^{i}{}_{j} \equiv \delta A^{i}{}_{j} - \frac{1}{3}\delta^{i}{}_{j}\,\delta A^{k}{}_{k}. \tag{3.60}$$

By construction,

$$\delta \tilde{A}^{i}{}_{i} = 0. \tag{3.61}$$

This decomposition is useful because the traceless part of the tensor is gauge invariant at first order. To see this we consider the spatial transformation

$$\delta A^{i}{}_{j} \to \delta A^{i}{}_{j} - \frac{1}{3}\delta^{i}{}_{j}\frac{d}{dt}\left(\bar{A}^{k}{}_{k}\right)\xi^{0}. \tag{3.62}$$

in Equation 3.57 the terms involving derivatives of ξ^{i} cancel because the background spatial tensor is proportional to $\delta^{i}{}_{j}$:

$$\xi^{i}{}_{,k}\bar{A}^{k}{}_{j} - \xi^{k}{}_{,j}\bar{A}^{i}{}_{k} = \frac{1}{3}\bar{A}^{l}{}_{l}\left(\xi^{i}{}_{,j} - \xi^{i}{}_{,j}\right) = 0. \tag{3.63}$$

Therefore, the gauge transformation of $\delta A^{i}{}_{j}$ is proportional only to $\delta^{i}{}_{j}$. It changes only the trace, and so taking the traceless part removes this contribution.

Decomposition of the Fourier modes

Previous results also imply that the linearized equations are diagonal in Fourier space. A perturbation around the translation invariant background may then be expanded as

$$\delta X(t,\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\,\delta X_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{3.64}$$

At linear order, different Fourier modes do not interact. Each mode $\mathbf{k}$ may therefore be studied independently. Furthermore, rotational invariance allows us to classify perturbations by how they transform under rotations about the wave vector $\mathbf{k}$. If a perturbation mode transforms as

$$\delta X_{\mathbf{k}} \to e^{im\psi}\delta X_{\mathbf{k}} \tag{3.65}$$

under a rotation by angle ψ around $\mathbf{k}$, then m is called its helicity. Scalar, vector, and tensor perturbations have helicities

$$m = 0, \qquad m = \pm 1, \qquad m = \pm 2, \tag{3.66}$$

respectively. Differently put, the scalar–vector–tensor decomposition generates a map into irreducible helicity sectors of the spatial rotation group.

FLRW background and perturbation

An important application of the previous results consists of the metric perturbation

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu}, \tag{3.67}$$

we obtain

$$\delta\tilde{g}_{\mu\nu} = \delta g_{\mu\nu} - \mathcal{L}_\xi \bar{g}_{\mu\nu}. \tag{3.68}$$

Since

$$(\mathcal{L}_\xi \bar{g})_{\mu\nu} = \bar{\nabla}_\mu \xi_\nu + \bar{\nabla}_\nu \xi_\mu, \tag{3.69}$$

the metric perturbation transforms as

$$\delta\tilde{g}_{\mu\nu} = \delta g_{\mu\nu} - \bar{\nabla}_\mu \xi_\nu - \bar{\nabla}_\nu \xi_\mu. \tag{3.70}$$

For a spatially flat FLRW background, the spatial background metric is

$$\bar{g}_{ij} = a^2(t)\delta_{ij}. \tag{3.71}$$

Following the previous discussion, perturbations can be decomposed into pieces that transform as scalars, vectors, and tensors under spatial rotations. Each sector has a distinct physical implication that we will discuss shortly.

The scalar sector of the perturbed FLRW metric may be written as

$$ds^2 = -(1 + 2\Phi)dt^2 + 2a(t)\partial_i B\, dx^i dt + a^2(t)\left[(1 - 2\Psi)\delta_{ij} + 2\partial_i\partial_j C\right] dx^i dx^j. \tag{3.72}$$

where the perturbation Φ modifies the time-time component of the metric and therefore the relation between coordinate time and proper time. The function B is the scalar part of the shift. The perturbation Ψ changes the local spatial scale factor and is therefore associated with scalar curvature perturbations of the spatial hypersurfaces. Finally, C describes the scalar shear deformation of the spatial metric

Equation 3.72 is the most general scalar deformation of a spatially flat FLRW metric at first order. The full metric perturbation also contains vector and tensor pieces. In general, any spatial vector can be decomposed into a gradient part and a divergence-free part

$$B_i = -B_{,i} + B_i^V, \qquad \partial^i B_i^V = 0. \tag{3.73}$$

Similarly, the symmetric traceless tensor E_{ij} can be decomposed into scalar, vector, and tensor parts,

$$C_{ij} = C_{ij}^S + C_{ij}^V + C_{ij}^T, \tag{3.74}$$

where

$$C_{ij}^S = C_{,ij} - \frac{1}{3}\delta_{ij}\nabla^2 C, \tag{3.75}$$

$$C_{ij}^V = -\frac{1}{2}\left(C_{i,j} + C_{j,i}\right), \qquad \partial^i C_i = 0, \tag{3.76}$$

and

$$\partial^i C_{ij}^T = 0, \qquad \delta^{ij} C_{ij}^T = 0. \tag{3.77}$$

Schematically, for the metric we obtain

$$g_{0i} = a(t)\left(\partial_i B + B_i^V\right), \qquad \partial^i B_i^V = 0, \tag{3.78}$$

and

$$g_{ij} = a^2(t)\left[(1 - 2\Psi)\delta_{ij} + 2\partial_i\partial_j C + 2\partial_{(i} C_{j)}^V + h_{ij}\right], \tag{3.79}$$

where we have $\partial^i h_{ij} = 0$ and $h^i{}_i = 0$.

Physically, the scalar sector is responsible for density perturbations and structure formation. It couples to perturbations in the energy density, pressure, velocity potential, and anisotropic stress. The vector sector is described by transverse vectors such as B_i^V and C_i^V. These represent rotational or vortical perturbations. In standard inflationary cosmology, they are usually neglected because they are not efficiently generated and typically decay with the expansion. The tensor sector is described by the transverse-traceless perturbation h_{ij}. It contains two independent degrees of freedom, corresponding to the two polarizations of gravitational waves. Unlike scalar perturbations, tensor perturbations are gauge-invariant at linear order. They therefore have a direct physical interpretation as gravitational waves propagating on the expanding FLRW background.

Finally, we can count the degrees of freedom. The metric perturbation has ten independent components. These split into

$$\begin{aligned} &4 \quad \text{scalar components},\\ &4 \quad \text{vector components},\\ &2 \quad \text{tensor components}. \end{aligned}$$

It is important to note that gauge freedom removes unphysical components from the scalar and vector sectors, while the tensor sector is already gauge-invariant at first order.

There are two particularly important gauge-invariant quantities that are worth discussing further. Firstly, the curvature perturbation on uniform-density hypersurfaces typically defined as [47, 48],

$$\zeta = -\Psi - H\frac{\delta\rho}{\dot{\bar{\rho}}}. \tag{3.80}$$

on constant density hypersurfaces where, $\delta\rho = 0$. The transformation of $\delta\rho$ is then canceled by the gauge transformation of Ψ.

Second is the comoving curvature perturbation,

$$\mathcal{R} = \Psi - H\frac{\delta q}{\bar{\rho} + \bar{p}} \tag{3.81}$$

measured on comoving hypersurfaces where the scalar momentum perturbation vanishes.

In the special case of a single-field inflation, where the inflaton perturbation $\delta\phi$ carries the momentum perturbation, the equation above can be reformulated as [41]

$$\mathcal{R} = \Psi + H\frac{\delta\phi}{\dot{\bar{\phi}}}. \tag{3.82}$$

For a scalar field, we often also introduce the gauge-invariant inflaton perturbation on spatially flat slices [49],

$$Q = \delta\phi + \frac{\dot{\bar{\phi}}}{H}\Psi. \tag{3.83}$$

The variable Q is closely related to the Mukhanov–Sasaki variable used in the quantization of scalar perturbations. We note that it combines matter and metric fluctuations into a gauge-invariant quantity.

The variables ζ and $\mathcal{R}$ are related by the linearized Einstein equations. On super-horizon scales, $k \ll aH$ gradient terms are negligible, and for adiabatic perturbations, we have

$$\zeta \simeq -\mathcal{R}, \tag{3.84}$$

up to sign conventions. The curvature perturbation is, therefore, conserved outside the horizon for adiabatic modes.

3.2.4 Omnium rerum principia parva sunt[4]

Presently, we have assigned our contemporary epoch the title of precision cosmology. Yet somewhat ironically, the orders of magnitude we routinely discard as systematics in our

[4] The beginnings of all things are small —Cicero

experiments once, portentously, harbored the seed of our cosmos.

Previously, we have defined a mathematical framework for cosmological perturbations. Here, we endeavor to explain how these small deviations become the large-scale structure observed today.

The relevant quantity is the density contrast,

$$\delta(t, \mathbf{x}) = \frac{\rho(t, \mathbf{x}) - \bar{\rho}(t)}{\bar{\rho}(t)}. \tag{3.85}$$

Linear perturbation theory applies as long as $|\delta| \ll 1$. Thus, structure formation becomes nonlinear when the density contrast reaches order unity,

$$\delta \sim 1,$$

at which point overdense regions detach from the background expansion and collapse gravitationally.

A useful intuition can be gained by considering a spherical overdensity embedded in an otherwise homogeneous universe. If the region has density larger than the cosmic average, it behaves locally like a separate universe with positive spatial curvature. Its evolution may therefore be understood through a local Friedmann equation: the overdense region expands more slowly than the background, eventually turns around, and may collapse [48].

In a static universe, gravitational attraction would lead to an exponential instability,

$$\delta(t) \propto e^{t/\tau}. \tag{3.86}$$

The expanding universe changes this conclusion. Expansion acts as a friction term, reducing the efficiency of gravitational collapse. Different behavior can be seen during different cosmological eras. During matter domination, for instance, the growing mode evolves as a power law,

$$\delta(t) \propto t^{2/3} \propto a(t), \tag{3.87}$$

whereas during radiation domination, the growth of matter perturbations is strongly suppressed. In the simplest approximation,

$$\delta(t) \propto \ln t. \tag{3.88}$$

These perturbations are therefore almost frozen. Thus, efficient gravitational clustering begins only after the Universe becomes matter-dominated. Before equality, radiation pressure and the rapid expansion prevent matter perturbations from growing significantly.

In order to produce a more realistic framework, we should mention that density perturbations are not exactly spherical. Therefore, Fourier decomposition proves indispensable to describe individual modes,

$$\delta(t, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \delta_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{3.89}$$

or, schematically,

$$\delta = \sum_{\mathbf{k}} \delta_{\mathbf{k}} \sin(\mathbf{k} \cdot \mathbf{x} + \varphi_{\mathbf{k}}). \tag{3.90}$$

As mentioned before, in linear order, each Fourier mode evolves independently, and each mode satisfies the equation of motion of a spherically symmetric overdensity. Moreover, it should be emphasized that the physical scale of a perturbation is finite and is therefore associated with a causal scale. In cosmology, this is expressed through horizon crossing.

The power spectrum describes the statistical distribution of these Fourier amplitudes is

$$P(k) = |\delta_k|^2. \tag{3.91}$$

More precisely, it is defined through the two-point function in Fourier space,

$$\langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{k}') P(k). \tag{3.92}$$

This offers a description of the strength of these fluctuations as a function of scale, where observationally large scales are probed most cleanly by the CMB, intermediate scales by galaxy clustering, and smaller scales by probes such as the Lyman-α fluctuations.

Additionally, we can depict the causal history through the shape of the power spectrum. Small-scale modes enter the horizon earlier, during radiation domination, and their growth is suppressed. Large-scale modes enter later, after matter domination begins, and therefore experience less suppression. The turnover in the matter power spectrum corresponds roughly to the horizon scale at matter-radiation equality.

The photon-baryon fluid provides an important description. Before recombination, photons and baryons are tightly coupled by electromagnetic interactions and behave approximately as a single fluid. Gravity pulls matter into overdense regions, while photon pressure resists compression. The result is an acoustic oscillator [50, 51]. Compression and rarefaction of the photon-baryon plasma produce sound waves, whose phase at recombination is imprinted in the CMB anisotropy spectrum. Reminiscent of a sound cavity, the generated sound waves will act as standing waves, creating phenomena such as interference. This explains the acoustic peaks of the CMB power spectrum. Their positions and relative heights depend on the composition of the Universe [39]. Baryons enhance compressional peaks; dark matter determines the gravitational potentials; radiation controls the early expansion; and dark energy affects distances to the last-scattering surface.

On very small scales, fluctuations are damped. The primordial plasma is not a perfect fluid at arbitrarily short distances. Photons have a finite mean free path and diffuse from hot regions into cold regions, smoothing temperature fluctuations. This diffusion damping, often called Silk damping [52], suppresses CMB anisotropies at high multipoles. Free-streaming neutrinos also modify the evolution of perturbations and contribute to damping and phase shifts [53].

3.2.5 Power Spectra and Scale Dependence

Having introduced the statistical description of primordial perturbations in terms of their correlation functions, we may now parse the general description to isolate individual contributions.

Scalar perturbations

For scalar perturbations, the relevant gauge-invariant quantity is usually the comoving curvature perturbation $\mathcal{R}$. Its two-point function defines the scalar power spectrum [45].

$$\langle \mathcal{R}_{\mathbf{k}} \mathcal{R}_{\mathbf{k}'} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{k}') P_{\mathcal{R}}(k). \tag{3.93}$$

Appendix B provides a more detailed discussion of this notation. The dimensionless scalar power spectrum is then defined as

$$\Delta_s^2(k) \equiv \Delta_{\mathcal{R}}^2(k) = \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k) \tag{3.94}$$

and the variance of $\mathcal{R}$ is obtained by integrating over logarithmic intervals in wavenumber:

$$\langle \mathcal{R}^2 \rangle = \int d\ln k \, \Delta_{\mathcal{R}}^2(k). \tag{3.95}$$

If the perturbations are Gaussian, the power spectrum contains all statistical information. Non-Gaussianity is encoded in higher-order correlation functions, such as the bispectrum and trispectrum. In the single-field slow-roll models, primordial non-Gaussianity is typically small, while more general models may produce larger higher-order correlations. The scale dependence of scalar perturbations is described by the scalar spectral index,

$$n_s - 1 = \frac{d \ln \Delta_s^2}{d \ln k}. \tag{3.96}$$

Exact scale invariance corresponds to

$$n_s = 1, \tag{3.97}$$

and the running of the spectral index is

$$\alpha_s = \frac{dn_s}{d\ln k}. \tag{3.98}$$

Around a pivot scale $k_\star$, denoting the reentry of modes in the horizon that experiment can probe quite accurately [31] - the spectrum is commonly parametrized as

$$\Delta_s^2(k) = A_s(k_\star)\left(\frac{k}{k_\star}\right)^{n_s(k_\star)-1+\frac{1}{2}\alpha_s(k_\star)\ln(k/k_\star)}. \tag{3.99}$$

Here A_s is the scalar amplitude, n_s measures the tilt, and α_s measures the running of the tilt. We will return to the feasibility of constraining these parameters by observational means in the next chapter dedicated to the case study of a specific inflationary model.

Tensor perturbations

Before defining the tensor power spectrum, it is useful to isolate the tensor sector of the metric perturbation. The line element is given by the transverse-traceless sector [45]

$$ds^2 = -dt^2 + a^2(t)\left(\delta_{ij} + h_{ij}\right)dx^i dx^j. \tag{3.100}$$

As a reminder of the previous section, these two conditions leave two independent propagating degrees of freedom, corresponding to the two tensor polarizations. In Fourier space, the tensor perturbation may be decomposed as

$$h_{ij}(\tau, \mathbf{x}) = \sum_{s=+,\times}\int \frac{d^3k}{(2\pi)^3}\epsilon_{ij}^s(\mathbf{k})h_{\mathbf{k}}^s(\tau)e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{3.101}$$

where $\mathbf{k}$ is the comoving momentum vector and ϵ_{ij}^s are polarization tensors. They satisfy

$$k^i\epsilon_{ij}^s = 0, \qquad \delta^{ij}\epsilon_{ij}^s = 0, \tag{3.102}$$

and are conventionally normalized as

$$\epsilon_{ij}^s\epsilon^{s'\,ij} = 2\delta^{ss'}. \tag{3.103}$$

The two values $s = +, \times$ label the two independent gravitational-wave polarizations.

In the Einstein–Hilbert formalism, the action can be written to second order in h_{ij}, using conformal time $d\tau = dt/a(t)$,

$$S_h^{(2)} = \frac{M_{\rm Pl}^2}{8}\int d\tau\, d^3x\, a^2\left[(h_{ij}')^2 - (\partial_\ell h_{ij})^2\right]. \tag{3.104}$$

where each tensor polarization behaves like a massless scalar degree of freedom in an expanding background, up to an overall normalization fixed by the Planck mass. Introducing the canonical variable

$$v_{\mathbf{k}}^{s} = \frac{aM_{\rm Pl}}{2} h_{\mathbf{k}}^{s}, \tag{3.105}$$

the action becomes

$$S_h^{(2)} = \sum_{s=+,\times} \frac{1}{2} \int d\tau\, d^3k \left[|(v_{\mathbf{k}}^{s})'|^2 - \left(k^2 - \frac{a''}{a}\right) |v_{\mathbf{k}}^{s}|^2 \right]. \tag{3.106}$$

We note that, after decomposing into the two polarization states, tensor perturbations reduce to two independent copies of the canonical quadratic system encountered for scalar fluctuations. This formalism will be expanded upon in the subsequent chapter, where the transition from scale factor a to the variable z is forthcoming. This modification will be thoroughly explicated in the accompanying Appendix D. Additionally, we emphasize that the tensor power spectrum is fixed by the expansion rate during inflation. Since tensor modes are fluctuations of the metric itself, their amplitude is suppressed by $M_{\rm Pl}$. For each polarization, one obtains

$$\Delta_h^2(k) = \frac{k^3}{2\pi^2} P_h(k) = \frac{4}{M_{\rm Pl}^2} \left(\frac{H}{2\pi}\right)^2. \tag{3.107}$$

The total tensor spectrum is then the sum over the two polarizations:

$$\Delta_t^2(k) = 2\Delta_h^2(k) = \frac{8}{M_{\rm Pl}^2} \left(\frac{H}{2\pi}\right)^2. \tag{3.108}$$

Generally, a light degree of freedom during inflation acquires fluctuations of order $H/2\pi$. Tensor perturbations obey the same logic, but their normalization is gravitational, hence the factor of $M_{\rm Pl}^{-2}$ [41]. From an energetic perspective, therefore, larger H produces larger primordial gravitational waves.

For each polarization mode of the transverse-traceless metric perturbation h_{ij}, we define

$$\langle h_{\mathbf{k}} h_{\mathbf{k}'} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{k}') P_h(k), \tag{3.109}$$

where the scale dependence is defined by

$$n_t = \frac{d \ln \Delta_t^2}{d \ln k}. \tag{3.110}$$

Finally, the tensor spectrum is often written as

$$\Delta_t^2(k) = A_t(k_\star) \left(\frac{k}{k_\star}\right)^{n_t}. \tag{3.111}$$

In slow-roll inflation models, the scale dependence arises because the Hubble parameter is not exactly constant. Different Fourier modes leave the Hubble radius at different times, determined approximately by

$$k = aH. \tag{3.112}$$

If the background were exactly de Sitter, H would be constant and the spectra would be exactly scale invariant. In quasi-de Sitter inflation models, however, H changes slowly with time, so modes exiting at different times acquire slightly different amplitudes. This is expressed by writing

$$\frac{d\ln\Delta_s^2}{d\ln k} = \frac{d\ln\Delta_s^2}{dN}\frac{dN}{d\ln k}, \tag{3.113}$$

where $N = \ln a$ is the number of e-folds. Since horizon crossing, $k = aH$, gives $\ln k = N + \ln H$, we have to first order in slow-roll,

$$\frac{dN}{d\ln k} = \left(1 + \frac{d\ln H}{dN}\right)^{-1} \simeq 1 + \epsilon. \tag{3.114}$$

More specifically, for a single-field slow-roll inflation, the scalar and tensor spectra at horizon crossing are approximately

$$\Delta_s^2(k) \simeq \frac{1}{24\pi^2 M_{\rm Pl}^4}\frac{V}{\epsilon_V}\bigg|_{k=aH}, \tag{3.115}$$

$$\Delta_t^2(k) \simeq \frac{2}{3\pi^2}\frac{V}{M_{\rm Pl}^4}\bigg|_{k=aH}, \tag{3.116}$$

where ϵ_V is the potential slow-roll parameter explained below. The scalar amplitude is enhanced when the potential is flat, since ϵ_V is small. The tensor amplitude, by contrast, is set directly by the inflationary energy scale.

Another way to describe the essential slow-roll variables is by defining potential slow-roll parameters [44]

$$\epsilon_V = \frac{M_{\rm Pl}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2, \qquad \eta_V = M_{\rm Pl}^2\frac{V_{,\phi\phi}}{V}. \tag{3.117}$$

This description is contingent upon the fact that in the slow-roll limits $\epsilon \ll 1$ we can drop the kinetic term proportional to $\dot{\phi}^2$ from the Friedmann equation and, therefore, get $H^2 \propto V$. To first order in slow-roll, the spectral indices become

$$n_s - 1 = 2\eta_V - 6\epsilon_V, \tag{3.118}$$

and

$$n_t = -2\epsilon_V. \tag{3.119}$$

The tensor-to-scalar ratio is

$$r := \frac{\Delta_t^2}{\Delta_s^2} = 16\epsilon_V. \tag{3.120}$$

Combining Eqs. (3.119) and (3.120) gives the single-field slow-roll consistency relation

$$r = -8n_t. \tag{3.121}$$

These relations show how observations of primordial perturbations constrain inflationary dynamics. The amplitude A_s fixes the overall scale of scalar fluctuations. The tilt n_s measures the departure from exact scale invariance and therefore probes the time dependence of the inflationary background. The tensor amplitude and tensor-to-scalar ratio constrain the energy scale of inflation. In slow-roll models, these observables are directly related to the height, slope, and curvature of the inflationary potential.

3.3 Empirical Connections

Though this section finds its place as the final movement of this chapter, let it not be misconstrued as a reflection of its value. Despite our inherent penchant for theoretical elegance, the Universe declares its own reality. Therefore, our ultimate objective must be to align our predictions with empirical evidence provided by observations.

This section lays the groundwork for two distinct branches of observational methodology, namely, gravitational wave astronomy and gravitational weak lensing. These subjects will receive comprehensive treatment in subsequent chapters, where specific phenomena and case studies will be examined in detail.

3.3.1 Gravitational waves (GW)

For the greater part of modern cosmology - dating from the pivotal discovery of cosmic expansion - the primordial universe has been observationally opaque. The Last Scattering Surface (LSS), marking the recombination era, renders all preceding phenomena inaccessible to electromagnetic inquiry. The recent onset of gravitational waves, however, has the potential to provide a previously inaccessible window to these primordial epochs [54]. Already, it offers a novel lens in strong-field regimes such as compact binary coalescence.

We may recall that gravitational waves are propagating tensor perturbations of the spacetime metric. In the weak-field regime, the metric is written as a small perturbation around Minkowski spacetime,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1, \tag{3.122}$$

where cosmological expansion can be neglected. The physical content of the perturbation

$h_{\mu\nu}$ is isolated by imposing appropriate gauge conditions. In the Lorenz gauge,

$$\partial^{\mu}\bar{h}_{\mu\nu} = 0, \tag{3.123}$$

where

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \qquad h = \eta^{\mu\nu}h_{\mu\nu}, \tag{3.124}$$

the linearized Einstein equations in vacuum reduce to

$$\Box\bar{h}_{\mu\nu} = 0. \tag{3.125}$$

The solutions are given by plane waves,

$$\bar{h}_{\mu\nu} = \mathrm{Re}\left[A_{\mu\nu}e^{ik_{\alpha}x^{\alpha}}\right], \tag{3.126}$$

where $A_{\mu\nu}$ is the polarization tensor and k^{μ} is the wave-vector. Computing the derivatives of the plane wave gives $\partial_{\alpha}\bar{h}_{\mu\nu} = ik_{\alpha}\bar{h}_{\mu\nu}$. The wave equation, therefore, implies $k^{\mu}k_{\mu} = 0$. This expresses the fact that gravitational waves propagate on null hypersurfaces, i.e., at the speed of light in vacuum. Furthermore, the Lorenz condition implies

$$k^{\mu}A_{\mu\nu} = 0. \tag{3.127}$$

This imposes four constraints on the amplitude $A_{\mu\nu}$. Since the perturbation metric is symmetric, $A_{\mu\nu}$ is also symmetric, and thus, it initially has ten independent components. The Lorenz condition reduces these to six. The remaining coordinate freedom can be used to impose the transverse-traceless gauge [55]

$$h = 0, \qquad h_{\mu\nu}u^{\mu} = 0, \qquad k^{\mu}h_{\mu\nu} = 0, \tag{3.128}$$

where u^{μ} is the four-velocity of the observer. The first condition is tracelessness; the second removes components along the observer's time direction; the third imposes transversality with respect to the propagation direction. Together, these conditions reduce the perturbation to two independent physical degrees of freedom. For a wave propagating in the z-direction, we may then take

$$k^{\mu} = \left(\frac{\omega}{c}, 0, 0, \frac{\omega}{c}\right). \tag{3.129}$$

The transverse-traceless perturbation then takes the form

$$h_{\mu\nu}^{\rm TT} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} e^{ik_\alpha x^\alpha}. \tag{3.130}$$

The two independent components h_+ and $h_\times$ are the two polarizations of the gravitational wave. The $+$ polarization stretches one transverse direction while compressing the orthogonal one; the $\times$ polarization does the same pattern after a rotation by 45°.

We may best observe the effect of gravitational waves on freely falling test particles. Locally, these particles would follow geodesics. Gravitational waves are then detected through the relative separation of neighboring geodesics. Consider two nearby freely falling particles separated by a spatial displacement vector ξ^i. In the presence of a gravitational wave, their proper separation changes according to the metric perturbation. For a wave propagating in the z-direction, and restricting to particles in the transverse x-y plane, the spatial line element becomes

$$d\ell^2 = (1 + h_+)\, dx^2 + (1 - h_+)\, dy^2 + 2h_\times dx\, dy. \tag{3.131}$$

For the $+$ polarization, a displacement along the x-axis is stretched while a displacement along the y-axis is compressed, and half a period later the effect reverses. For the $\times$ polarization, the stretching and compression axes are rotated by 45°.

An interferometer such as the Laser Interferometer Gravitational-Wave Observatory (LIGO) or Virgo measures this differential change in its arm length. Such a detector would be sensitive to the so-called strain between two perpendicular arms. We will provide more details on the detector response and their setup shortly.

Generation of gravitational waves

In the weak-field regime, gravitational waves are sourced by time-dependent, non-spherical distributions of stress-energy. The linearized Einstein equations with matter take the form

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu}. \tag{3.132}$$

The solution can be written using a retarded Green function[5], reflecting the fact that changes in the source propagate at the speed of light. Equivalently, the field observed at

[5] More intuitively, a Green function as a solution adds up the effect of every source point s, in the past, weighted by its influence at the present observer's point.

$(t, \mathbf{x})$ is determined by the stress-energy distribution evaluated at the retarded time,

$$t_{\text{ret}} = t - \frac{|\mathbf{x} - \mathbf{x}'|}{c}.$$

Thus,

$$\bar{h}_{\mu\nu}(t, \mathbf{x}) = \frac{4G}{c^4} \int d^3x' \, \frac{T_{\mu\nu}\left(t - \frac{|\mathbf{x}-\mathbf{x}'|}{c}, \mathbf{x}'\right)}{|\mathbf{x} - \mathbf{x}'|}. \tag{3.133}$$

This expression is a spatial integral over the source, with the retarded time condition selecting the part of the source lying on the past light cone of the observation event [55]. The field observed at a given time is therefore determined by the state of the source at the appropriate retarded time. The dominant radiative contribution is quadrupolar. Monopole radiation is absent because total mass-energy is conserved; similarly, dipole radiation is absent because momentum conservation forbids radiation from the centre-of-mass motion. The first non-vanishing radiative multipole is the quadrupole moment,

$$Q_{ij} = \int \rho \left(x_i x_j - \frac{1}{3}\delta_{ij} r^2 \right) d^3x. \tag{3.134}$$

The gravitational-wave amplitude is therefore controlled by the second time derivative of the quadrupole moment, i.e., the Einstein quadrupole formula at the level of the emitted waveform is

$$h_{ij}^{\text{TT}} \sim \frac{2G}{c^4 r} \ddot{Q}_{ij}^{\text{TT}} \left(t - \frac{r}{c} \right). \tag{3.135}$$

This is the leading-order result in the radiation zone for a slow-moving, weakly self-gravitating source and underscores the role of binary systems as efficient gravitational-wave sources. Here, the superscript TT denotes the use of the transverse-traceless gauge. This should be distinguished from the transverse-traceless projection sometimes used in the construction of radiative fields from isolated sources; the two notions coincide only under additional assumptions, for example, in the appropriate far-zone radiative limit. A binary system has a time-dependent quadrupole moment. If the two masses orbit each other, the mass distribution is not spherically symmetric, and its quadrupole varies periodically. The resulting gravitational wave carries away energy and angular momentum, causing the orbit to shrink. In contrast, a perfectly spherical collapse or expansion does not radiate gravitational waves, because it has no time-dependent traceless quadrupole moment.

Energy carried by gravitational waves

While gravitational waves transport energy across space, the theory of General Relativity makes it impossible to define exactly how much energy is localized at specific points in spacetime. The reason is that the equivalence principle forbids any notion of the gravitational energy density at a point and thus, it cannot be represented by a coordinate- independent stress-energy tensor, i.e., by choosing local inertial coordinates, the connection can always be made to vanish at a point. The ambiguity becomes less problematic only after averaging over scales large compared with the wavelength [56]. We can define an effective stress-energy tensor for gravitational waves by averaging over several wavelengths. The Isaacson effective stress-energy tensor is [57]

$$T^{\mathrm{GW}}_{\alpha\beta} = \frac{c^4}{32\pi G}\left\langle \partial_\alpha h^{\mathrm{TT}}_{ij} \partial_\beta h^{ij}_{\mathrm{TT}} \right\rangle . \tag{3.136}$$

The brackets denote an average over scales larger than the gravitational wavelength but smaller than the characteristic scale over which the background changes. The corresponding energy flux is proportional to the square of the time derivative of the wave amplitude

$$\frac{dE}{dt} = \frac{c^3}{32\pi G}\int \left\langle \dot{h}^{\mathrm{TT}}_{ij} \dot{h}^{ij}_{\mathrm{TT}} \right\rangle dA. \tag{3.137}$$

Equivalently, in terms of the two polarizations,

$$\frac{dE}{dt\, dA} = \frac{c^3}{16\pi G}\left\langle \dot{h}^2_+ + \dot{h}^2_\times \right\rangle . \tag{3.138}$$

The dependence on $\dot{h}^2$ means that a static deformation of the metric does not transport gravitational-wave energy. Energy transport is rather associated with oscillatory, propagating tensor perturbations.

The radiative loss of energy from compact binaries was experimentally tested before the direct detection of gravitational waves. The orbital decay of the Hulse–Taylor binary pulsar PSR B1913+16 was found to agree with the energy loss predicted by the quadrupole formula [58]. The first comparison was already consistent with the general-relativistic prediction at roughly the ten-percent level, while later timing analyzes, after accounting for kinematic and Galactic corrections, brought the agreement to the level of about 0.2% [59]. This provided the first strong experimental indirect evidence for gravitational radiation from a compact binary system and confirmed, observationally, that the quadrupole formula correctly describes radiative losses in general relativity. Historically, this should be distinguished from the earlier theoretical work of Trautman, for instance, which had already established that gravitational waves carry energy [60].

Detector response and the LIGO–Virgo setup

Laser interferometers such as LIGO and Virgo are designed to measure the strain produced by a passing gravitational wave. As we mentioned before, the observable is the relative change in arm length, $h \sim \frac{\Delta L}{L}$. A gravitational wave passing perpendicular to the detector plane changes the proper lengths of the two arms in opposite ways. For a pure + polarization aligned with the arms,

$$\frac{\Delta L_x}{L} = \frac{1}{2}h_+, \qquad \frac{\Delta L_y}{L} = -\frac{1}{2}h_+. \tag{3.139}$$

The interferometer measures the differential effect:

$$\frac{\Delta L_x - \Delta L_y}{L} = h_+. \tag{3.140}$$

For a $\times$ polarization, the response depends on the detector orientation relative to the polarization axes.

We emphasize that a network of detectors is crucial to observations because a single detector cannot fully reconstruct the sky position, polarization content, and source parameters of a gravitational-wave event. Different detectors observe the same wave with different arrival times and antenna-pattern responses. The relative time delays localize the source on the sky, while the relative amplitudes and phases help constrain the polarization and inclination of the source.

3.3.2 Gravitational weak lensing

Gravitational weak lensing, predicted by general relativity, manifests as a subtle distortion in the shape of background galaxies. It can be detected statistically across large populations of galaxies and encodes information on the integrated mass distribution and offers a unique window for probing large-scale structure and cosmology. In contrast to strong lensing, where the gravitational field is sufficiently large to generate multiple images, arcs, or Einstein rings, weak lensing produces only small distortions in the apparent shapes and sizes of background galaxies. These distortions are too small to be identified reliably in individual galaxies, but become measurable statistically by averaging over large ensembles of sources.

More specifically, light emitted by a distant source travels through an inhomogeneous universe before reaching the observer. Along the way, gravitational potentials associated with foreground structures deflect the photon trajectory. As a result, the image of the source is displaced, stretched, and magnified. Since galaxies have unknown intrinsic shapes, the lensing signal is extracted statistically from coherent alignments of many background galaxy images. If foreground structures trace the underlying matter distribution, at least in a biased statistical sense, they are expected to be associated with coherent distortion patterns in the shapes of more distant galaxies. The standard example is galaxy–galaxy lensing, where we measure the average tangential alignment of source galaxies around foreground lens galaxies. The same statistical logic can be adapted to more general foreground structures, such as clusters, voids, filaments, and, in the application considered later in this work, projected density ridges.

In the weak-field regime, the perturbed FLRW metric may be written in terms of scalar gravitational potentials, and their accumulated effect along the line of sight produces a small deflection of the apparent source position. The deflection angle is small; therefore, we usually work in the Born approximation, where the deflection is along the unperturbed light path. The lensing effect can then be described by a projected lensing potential [61].

$$\psi(\boldsymbol{\theta}) = \frac{2}{c^2} \int_0^{\chi_s} d\chi \, \frac{\chi_s - \chi}{\chi_s \chi} \Phi(\chi\boldsymbol{\theta}, \chi), \tag{3.141}$$

where χ is comoving distance, χ_s is the comoving distance to the source, $\boldsymbol{\theta}$ is the observed angular position, and Φ is the gravitational potential. In the corresponding convergence kernel, i.e., weighted projection of the matter density contrast δ, the lensing efficiency suppresses matter very close to the observer or to the source and gives the largest weight to intermediate distances along the line of sight.

In the weak lensing formalism we map the true angular position of a source, $\boldsymbol{\beta}$, to the observed image position, $\boldsymbol{\theta}$. Locally, this mapping is described by the Jacobian matrix

$$\mathcal{A}_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j}. \tag{3.142}$$

This matrix is conventionally decomposed as

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}. \tag{3.143}$$

where $\kappa = \frac{1}{2}\nabla_\theta^2 \psi$ is the convergence and the two shear components are

$$\gamma_1 = \frac{1}{2}\left(\partial_1^2 - \partial_2^2\right)\psi, \qquad \gamma_2 = \partial_1 \partial_2 \psi. \tag{3.144}$$

Here, κ describes the isotropic focusing or dilation of images. In the thin-lens limit, it is related to the projected surface mass density, while for large-scale structure, it is more generally a weighted line-of-sight projection of the matter density contrast. The shear γ, by contrast, describes the anisotropic stretching of images. In practice, weak-lensing observations are most directly sensitive to shear, because the intrinsic sizes and luminosities of galaxies are difficult to know accurately, whereas coherent shape distortions can be extracted statistically. For a localized foreground structure, the most useful shear component is the tangential shear, γ_+, which is defined relative to the direction connecting the lens and the source. A spherically symmetric mass distribution produces a tangential shear pattern, while the cross component, $\gamma_\times$, vanishes by symmetry. The cross component is therefore often used as a null test for residual systematics.

Chapter 4

Left-Right Symmetric Model of Inflation

"But nature is a stranger yet;
The ones that cite her most
Have never passed her haunted house,
Nor simplified her ghost."
— Emily Dickinson

Many years later, faced with the empirical constraints, we were to remember that distant Thursday in 2013 when the Planck collaboration discovered that non-minimal coupling[1] models are favored by observations [62].

Admittedly, it is in itself a monumental task to count the arrays of such formulations introduced over the decades [63][2]. In this light and given our previous discussion, we propose to study a left-right symmetric extension to a Higgs inflation model where the essential component centers around a non-minimal coupling of a scalar field and the curvature, with its framework being elevated to that of Einstein–Cartan. The preceding chapters have established the two ingredients required for the present model. Chapter 1 introduced gravity in first-order form, where the co-tetrad e^I and the spin connection ω^{IJ} are treated as the fundamental geometrical variables, and where the complexified Lorentz algebra admits a natural decomposition into self-dual and anti-self-dual sectors. Chapter 3 then introduced the cosmological setting in which this structure will be applied, namely an approximately homogeneous and isotropic inflationary background together with its linear perturbations. At the background level, because direct numerical evaluation is

[1] Between gravity and matter

[2] "You know how sometimes you meet somebody and they're really nice, so you invite them over to your house, and you keep talking with them, and they keep telling you more and more cool stuff? But then at some point you're like, maybe we should call it a day, but they just won't leave, and they keep talking, and as more stuff comes up it becomes more and more disturbing, and you're like, just stop already? That's kind of what happened with inflation." Offered Max Tegmark helpfully!

expensive, the investigation uses an emulator-assisted pipeline combining Mathematica and Markov Chain Monte Carlo (MCMC) sampling to determine the relevant space of inflationary variables given the observational constraints on cosmological parameters.

Additionally, it must be emphasized that the post-inflationary phase, also known as reheating, is crucial for constraining the theoretical framework. This phase is characterized by the oscillation of the inflaton around the minimum of the potential, and it serves as a connection between the slow-roll dynamics and the scales observed in the CMB.

4.1 Introducing the Model

4.1.1 Palatini–Higgs inflation

Before introducing the left-right symmetric inflationary model, useful intuition can be gained from examining the simpler case of the the Higgs inflation. We recall the basic mechanisms in the Palatini formulation [64]. Higgs inflation is compelling because it uses the scalar sector already present in the Standard Model. In its simplest form, the Higgs field is coupled to curvature by a non-minimal interaction of the form

$$\xi H^\dagger H\, R, \tag{4.1}$$

where ξ is a dimensionless non-minimal coupling. Using the unitary gauge, which removes the angular components of the Higgs doublet and working at field values much larger than the electroweak scale, the Higgs doublet may be represented by a single real scalar field h. This is because the Standard Model Higgs is a complex $SU(2)_L$ doublet

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_1 + i\theta_2 \\ h + i\theta_3 \end{pmatrix}. \tag{4.2}$$

It contains four real scalar components. After electroweak symmetry breaking, three of these are Goldstone modes that can then be gauged out so that the Higgs doublet is written around the vacuum expectation value as

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}. \tag{4.3}$$

where v is the electroweak vacuum expectation value and h is the physical Higgs fluctuation. For Higgs inflation, we should typically employ field values, $h \gg v$ much larger than the electroweak scale. We note that the Higgs field points entirely along the physical radial direction in field space.

In scalar-tensor theories, a non-minimally coupled scalar field is most naturally written in the Jordan frame. For a single real scalar field ϕ, the action may be written schematically as

$$S_J = \int d^4x\sqrt{-g}\left[f(\phi)R - \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right]. \tag{4.4}$$

The function $f(\phi)$ determines the effective gravitational coupling. The canonical Einstein–Hilbert form is recovered in the limit where $f(\phi) = M_{\rm Pl}^2/2$. We may use a conformal transformation as a means of locally rescale the metric,

$$g_{\mu\nu} \longrightarrow \tilde{g}_{\mu\nu} = \Omega^2(\phi)g_{\mu\nu}, \tag{4.5}$$

where $\Omega(\phi)$ is a positive scalar function and does not change the spacetime coordinates. It does, however, absorb the field-dependent coefficient of the curvature into the metric, so that the gravitational part of the action assumes the standard Einstein–Hilbert form. In four dimensions, the determinant transforms as

$$\sqrt{-\tilde{g}} = \Omega^4\sqrt{-g}, \tag{4.6}$$

while the inverse metric transforms as

$$\tilde{g}^{\mu\nu} = \Omega^{-2}g^{\mu\nu}. \tag{4.7}$$

In the metric formulation, where the connection is the Levi–Civita connection of the metric, the Ricci scalar also transforms non-trivially as [65]

$$R = \Omega^2\left[\tilde{R} + 6\tilde{\Box}\ln\Omega - 6\tilde{g}^{\mu\nu}\partial_\mu\ln\Omega\partial_\nu\ln\Omega\right], \tag{4.8}$$

Notice that the Ricci scalar produces additional derivative terms involving $\Omega(\phi)$. Since Ω depends on the scalar field, these terms modify the scalar kinetic sector. The conformal factor is chosen so that the gravitational part of the action becomes canonical. For the action (4.4) we shall choose

$$\Omega^2(\phi) = \frac{2f(\phi)}{M_{\rm Pl}^2}. \tag{4.9}$$

The action is then said to be in the Einstein frame

$$S_E = \int d^4x\sqrt{-\tilde{g}}\left[\frac{M_{\rm Pl}^2}{2}\tilde{R} - \frac{1}{2}K(\phi)\tilde{g}^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - U(\phi)\right], \tag{4.10}$$

where the potential is given by

$$U(\phi) = \frac{V(\phi)}{\Omega^4(\phi)}. \tag{4.11}$$

To facilitate comparison, the metric formulation of the kinetic coefficient is

$$K_{\text{metric}}(\phi) = \frac{1}{\Omega^2(\phi)} + \frac{6}{M_{\text{Pl}}^2}\left(\frac{df/d\phi}{\Omega^2(\phi)}\right)^2. \tag{4.12}$$

where the second term is the characteristic contribution generated by the conformal transformation of the Ricci scalar.

The Palatini formulation differs at this point. As we have seen, the metric and connection are treated as independent variables. The Ricci tensor $R_{\mu\nu}(\Gamma)$ is constructed from the independent connection Γ, and, the Palatini Ricci scalar transforms only through the inverse metric contraction,

$$g^{\mu\nu}R_{\mu\nu}(\Gamma) = \Omega^2\tilde{g}^{\mu\nu}R_{\mu\nu}(\Gamma), \tag{4.13}$$

and no additional derivative terms of Ω are generated from the curvature sector. Consequently, in the Palatini case, the Einstein-frame kinetic coefficient takes the simpler form

$$K_{\text{Palatini}}(\phi) = \frac{1}{\Omega^2(\phi)}. \tag{4.14}$$

In the Jordan frame, the Higgs–Palatini inflation action takes the form

$$S_J = \int d^4x\sqrt{-g}\left[\frac{1}{2}\left(M_{\text{Pl}}^2 + \xi h^2\right)g^{\mu\nu}R_{\mu\nu}(\Gamma) - \frac{1}{2}g^{\mu\nu}\partial_\mu h\partial_\nu h - V(h)\right], \tag{4.15}$$

with

$$V(h) = \frac{\lambda}{4}\left(h^2 - v^2\right)^2. \tag{4.16}$$

During inflation the assumption $h \gg v$ implies that $V(h) \simeq \lambda h^4/4$ and in the gravitational part of the action, the effective Planck mass depends on h as

$$M_{\text{eff}}^2(h) = M_{\text{Pl}}^2 + \xi h^2. \tag{4.17}$$

In the presence of non-minimally coupled matter, the Palatini and metric formulations lead to different scalar dynamics [66]. To analyze the inflationary dynamics, we need to perform the conformal transformation mentioned above to the Einstein frame

$$g_{\mu\nu}^{\mathbf{E}} = \Omega^2(h)g_{\mu\nu}^{\mathbf{J}}, \qquad \Omega^2(h) = 1 + \frac{\xi h^2}{M_{\text{Pl}}^2}, \tag{4.18}$$

where $\mathbf{E}$ and $\mathbf{J}$ correspond to the Einstein and Jordan frames, respectively. After the transformation, the action becomes

$$S_{\mathbf{E}} = \int d^4x\sqrt{-g_{\mathbf{E}}}\left[\frac{M_{\text{Pl}}^2}{2}R_{\mathbf{E}} - \frac{1}{2\Omega^2(h)}g_{\mathbf{E}}^{\mu\nu}\partial_\mu h\partial_\nu h - U(h)\right], \tag{4.19}$$

where the Einstein-frame potential is

$$U(h) = \frac{V(h)}{\Omega^4(h)}. \tag{4.20}$$

At this juncture, we are brought to ask, "What win I, if I gain the thing I seek?" Indeed, we notice that gaining a canonical gravitational sector is accomplished at the price of a non-canonical kinetic term for the scalar field. Crucial to this argument is the premise that, under proper assumptions on the Riemannian curvature tensor, these results can be generalized to a multi-field case [65].

In practice, the kinetic term may be brought to canonical form by introducing a new field φ satisfying

$$\frac{d\varphi}{dh} = \frac{1}{\Omega(h)} = \left(1 + \frac{\xi h^2}{M_{\mathrm{Pl}}^2}\right)^{-1/2}. \tag{4.21}$$

This integrates to

$$\varphi = \frac{M_{\mathrm{Pl}}}{\sqrt{\xi}} \sinh^{-1}\left(\frac{\sqrt{\xi} h}{M_{\mathrm{Pl}}}\right), \tag{4.22}$$

or equivalently

$$h = \frac{M_{\mathrm{Pl}}}{\sqrt{\xi}} \sinh\left(\frac{\sqrt{\xi}\varphi}{M_{\mathrm{Pl}}}\right). \tag{4.23}$$

The Einstein-frame action then takes the familiar single-field form

$$S_{\mathbf{E}} = \int d^4x \sqrt{-g_{\mathbf{E}}} \left[\frac{M_{\mathrm{Pl}}^2}{2} R_{\mathbf{E}} - \frac{1}{2} g_{\mathbf{E}}^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi - U(\varphi)\right]. \tag{4.24}$$

In this frame, the quartic Higgs potential becomes

$$U(h) = \frac{\lambda h^4}{4} \left(1 + \frac{\xi h^2}{M_{\mathrm{Pl}}^2}\right)^{-2}. \tag{4.25}$$

In terms of the canonical field φ, this may be written as

$$U(\varphi) = \frac{\lambda M_{\mathrm{Pl}}^4}{4\xi^2} \tanh^4\left(\frac{\sqrt{\xi}\varphi}{M_{\mathrm{Pl}}}\right). \tag{4.26}$$

We note that the Jordan-frame potential is flattened in the Einstein frame by the conformal factor; at large field values we obtain the limit

$$U(\varphi) \longrightarrow \frac{\lambda M_{\mathrm{Pl}}^4}{4\xi^2}, \tag{4.27}$$

which provides the slowly evolving potential required for inflation.

4.1.2 Einstein–Cartan Higgs inflation

The Palatini formulation modified the usual metric formulation of Higgs inflation by treating the connection as an independent variable yet symmetric in its lower indices. As viewed in previous chapters, the Einstein–Cartan formulation enlarges this framework by allowing the independent connection to possess torsion defined as the Lorentz-covariant exterior derivative of the co-tetrad.

Consider the Palatini–Higgs inflation. The gravitational part of the Jordan-frame action is given by

$$S_{\text{grav}} = \int d^4x\sqrt{-g}\,\frac{M_{\text{Pl}}^2 + \xi h^2}{2} R(\Gamma), \tag{4.28}$$

In the Einstein–Cartan formulation, the presence of torsion allows additional dimension-four gravitational densities that would otherwise be trivial or dynamically irrelevant. In particular, the Holst density may contribute once the connection is not restricted to be torsion-free. The gravitational action can therefore be extended to

$$\begin{aligned} S_{\text{grav}}^{\mathbf{EC}} = \frac{1}{2}\int d^4x\sqrt{-g}\,\left(M_{\text{Pl}}^2 + \xi h^2\right) R(\omega) \\ + \frac{1}{4\bar{\gamma}}\int d^4x\sqrt{-g}\,\left(M_{\text{Pl}}^2 + \xi_\gamma h^2\right) \epsilon^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma}(\omega) \end{aligned} \tag{4.29}$$

where ω is taken to be the torsionful spin connection. The first term is the usual non-minimally coupled Higgs-curvature term. The second term is the Holst contribution, with $\bar{\gamma}$ denoting the Barbero–Immirzi parameter and ξ_γ allowing for a Higgs-dependent coefficient. It is instructive to observe that there are additional inputs that can be included, such as the Nieh–Yan contribution [44]. For the purposes of the present work, however, the Nieh–Yan term will not be retained as the future analysis of the left-right symmetric model is organized around the emergence of an effective Barbero–Immirzi, sector.

To analyze the inflationary dynamics, we begin by performing a conformal transformation,

$$g_{\mu\nu} \longrightarrow \Omega^2 g_{\mu\nu}, \qquad \Omega^2 = 1 + \frac{\xi h^2}{M_{\text{Pl}}^2}. \tag{4.30}$$

As stated above, this transformation removes the non-minimal coupling part of the action. Conveniently, the remaining torsionful part of the connection can be solved for algebraically and substituted back into the action. The result is an effective Einstein-frame theory with canonical gravity and a modified scalar kinetic sector,

$$S_{\text{eff}} = \int d^4x\sqrt{-g}\left[\frac{M_{\text{Pl}}^2}{2}\mathring{R} - \frac{1}{2}K(h)g^{\mu\nu}\partial_\mu h\partial_\nu h - \frac{U(h)}{\Omega^4}\right], \tag{4.31}$$

where $\mathring{R}$ denotes the Ricci scalar of the torsion-free connection. The information about the original theory is now encoded in the kinetic function $K(h)$ [67]

For a single scalar field, the full Einstein–Cartan Higgs-inflation theory contains the standard non-minimal Higgs–Ricci coupling $\xi h^2 R$, together with an additional torsional Higgs-dependent Holst coupling controlled by ξ_γ. Having eliminated the torsion, the effective scalar kinetic function becomes

$$K(h) = \frac{1}{\Omega^2}\left[1 + \frac{6h^2\left(\frac{\xi_\gamma - \xi}{\bar{\gamma}}\right)^2}{M_{\rm Pl}^2\Omega^2\left[\Omega^4 + \bar{\gamma}^{-2}\left(1 + \frac{\xi_\gamma h^2}{M_{\rm Pl}^2}\right)^2\right]}\right]. \tag{4.32}$$

However, we may also consider setting the direct Higgs–Holst coupling ξ_γ to zero. The Higgs field then couples non-minimally only to the Ricci scalar, while the Holst sector is controlled solely by the constant Barbero–Immirzi parameter $\bar{\gamma}$. In this case,

$$K_{\rm H}(h) = \frac{1}{\Omega^2}\left[1 + \frac{6\xi^2 h^2}{M_{\rm Pl}^2\bar{\gamma}^2\Omega^2\left(\Omega^4 + \bar{\gamma}^{-2}\right)}\right]. \tag{4.33}$$

Notably, in the limit where the second term is negligible, we recover the Palatini kinetic structure

Next, the canonical inflaton field χ is defined by

$$\frac{d\chi}{dh} = \sqrt{K_{\rm H}(h)}. \tag{4.34}$$

and the Einstein-frame potential is

$$U_E(h) = \frac{U(h)}{\Omega^4}, \tag{4.35}$$

It warrants explicit mention that, its functional dependence on the canonical field χ is altered through Eq. (4.34).

The left-right symmetric model considered below will extend the Einstein–Cartan version of Higgs inflation by allowing the scalar sector to contain fields associated with the left and right gauge sectors. Furthermore, we keep the non-minimal coupling between the Higgs and the Holst sector active so that the Barbero–Immirzi parameter retains its dependence on the scalar fields.

4.1.3 Left-right symmetric models

As we discussed, in a theory coupled to fermions, local Lorentz transformations act on the internal frame, and the co-tetrad provides a map between spacetime tensors and internal Lorentz indices. We then use the spin connection to define the corresponding covariant derivative. Since the Lorentz algebra admits a chiral decomposition after complexification, the spin connection may be separated into self-dual and anti-self-dual parts, $\omega^{IJ} = \omega^{+IJ} + \omega^{-IJ}$ where the two sectors are associated with the two inequivalent chiral spinor representations of the Lorentz group. Thus, in a formulation where both ω^{+IJ} and ω^{-IJ} are retained, the gravitational sector itself carries a structure that can be compared directly with the left- and right-handed matter sectors.

For instance, in the Standard Model of particle physics, the weak interaction acts chirally: left-handed fermions transform non-trivially under $SU(2)_L$, whereas the corresponding right-handed fermions do not participate in the same weak interaction. Left-right symmetric models enlarge the electroweak gauge group so that right-handed fermions are also assigned to non-trivial multiplets

$$SU(2)_L \times SU(2)_R \times U(1)_{B-L}. \tag{4.36}$$

At high energies, this structure treats the left and right sectors more symmetrically, while at lower energies the observed parity violation is recovered through spontaneous symmetry breaking $SU(2)_{(L)} \times SU(2)_{(R)} \times U(1)_{B-L} \to SU(2)_{(L)} \times U(1)_Y$. In this configuration, we have an enlarged $\chi^{(L)(R)}$ Higgs field responsible for the Yukawa couplings to fermions. The scalar sector is equally enlarged to include $\phi^{(R)}$ which can spontaneously break the $SU(2)_R$ symmetry as well as a field $\phi^{(L)}$ which is necessary to preserve the left-right symmetry. Consequently, we recover a richer scalar content than that of the Standard Model that could be coupled non-minimally to gravity.

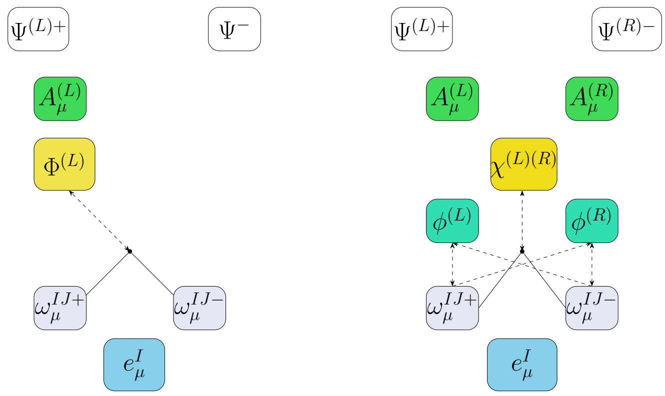


Figure 4.1: *The left part of the figure shows fields relevant to gravitation and the weak interaction in the context of the Standard Model and Einstein-Cartan gravity. Lorentz group indices on fields are shown whilst* (L) *denotes that the field transforms non-trivially under* $SU(2)_{(L)}$*.* $\Psi_a^{(L)}$ *and* $\Psi^{a'}$ *are left and right-handed fermion fields,* $A_\mu^{(L)}$ *is the weak gauge boson and* $\Phi^{(L)}$ *is the standard model Higgs field. The dotted line denotes the non-minimal coupling introduced that can lead to cosmic inflation. The right part of the figure shows a proposed generalization to a left-right symmetric standard model. Now a new* $SU(2)_{(R)}$ *interaction is introduced with corresponding gauge boson* $A_\mu^{(R)}$ *and which right-handed fermions* $\Psi^{a'(R)}$ *can transform non-trivially under. The Higgs sector is more involved, now consisting of fields* $(\chi^{(L)(R)}, \phi^{(L)}, \phi^{(R)})$*. Proposed non-minimal couplings between gravity and Higgs fields* $(\chi^{(L)(R)}, \phi^{(L)}, \phi^{(R)})$ *are denoted by dotted lines.*

As a reminder, in the usual Higgs-inflation scenario in the Einstein-Cartan formalism, the relevant coupling is given by

$$\xi \Phi^\dagger \Phi \, \epsilon_{IJKL} e^I \wedge e^J \wedge R^{KL}(\omega). \tag{4.37}$$

In the left-right symmetric case, however, the spin connection admits chiral components. This expression is invariant under the simultaneous interchange of the left-handed matter sector with the right-handed matter sector, as well as the self-dual and anti-self-dual sectors.

$$(L, +) \longleftrightarrow (R, -),$$

The dynamical fields of the simplified model are taken to be

$$\left(e^I{}_\mu, \omega^{+IJ}{}_\mu, \omega^{-IJ}{}_\mu, \Phi^R, \Phi^L\right). \tag{4.38}$$

The most general relevant non-minimal couplings sensitive to each parameter may be organized in the Lagrangian density

$$\mathcal{L}_{\rm NM} = \frac{1}{4}\left[\xi_1\chi^\dagger\chi + \xi_2\,\mathrm{Re}(\det\chi) + \xi_3\,\mathrm{Im}(\det\chi)\right]\epsilon_{IJKL}e^I\wedge e^J\wedge R^{KL}(\omega)$$
$$+\frac{1}{4}\left[\zeta\,\phi^{(L)}\cdot\phi^{(L)} + \eta\,\phi^{(R)}\cdot\phi^{(R)}\right]\epsilon_{IJKL}e^I\wedge e^J\wedge R^{KL}(\omega^+)$$
$$+\frac{1}{4}\left[\zeta\,\phi^{(R)}\cdot\phi^{(R)} + \eta\,\phi^{(L)}\cdot\phi^{(L)}\right]\epsilon_{IJKL}e^I\wedge e^J\wedge R^{KL}(\omega^-). \tag{4.39}$$

Then, the combined gravitational Lagrangian four-form is given by

$$\mathcal{L}_G = \frac{1}{32\pi G}\epsilon_{IJKL}e^I\wedge e^J\wedge R^{KL}(\omega) + \mathcal{L}_{NM(LR)} \tag{4.40}$$
$$= \frac{\bar{M}_{PL}^2}{4}\left(\epsilon_{IJKL} + \frac{2}{\bar{\gamma}}\eta_{IK}\eta_{JL}\right)e^I\wedge e^J\wedge R^{KL}(\omega) \tag{4.41}$$

Here $\mathcal{L}_{NM}$ denotes the matter-sector Lagrangian including the kinetic and potential terms of the scalar fields $(\chi^{(L)(R)}, \phi^{(L)}, \phi^{(R)})$. We note that, the quantities $\bar{M}_{\rm Pl}^2$ and $\bar{\gamma}$ are no longer constants. They are effective scalar-field-dependent functions determined by the non-minimal couplings. Explicitly,

$$\bar{M}_{\rm Pl}^2 = M_{\rm Pl}^2 + \xi_1\chi^\dagger\chi + \xi_2\,\mathrm{Re}(\det\chi) + \xi_3\,\mathrm{Im}(\det\chi)$$
$$+ (\zeta+\eta)\left[\phi^{(R)}\cdot\phi^{(R)} + \phi^{(L)}\cdot\phi^{(L)}\right], \tag{4.42}$$

$$\bar{\gamma} = -i\,\frac{\bar{M}_{\rm Pl}^2}{(\zeta-\eta)\left[\phi^{(L)}\cdot\phi^{(L)} - \phi^{(R)}\cdot\phi^{(R)}\right]}. \tag{4.43}$$

We will shortly examine the extent to which the background expansion is sensitive to each parameters. We now restrict the model to a spatially flat FLRW background. We choose the co-tetrad

$$e^0 = N(t)\,dt, \tag{4.44}$$
$$e^i = a(t)\delta^i_a\,dx^a, \tag{4.45}$$

with the corresponding spacetime metric as

$$ds^2 = -N^2(t)dt^2 + a^2(t)\delta_{ij}dx^i dx^j, \qquad \sqrt{-g} = Na^3. \tag{4.46}$$

The most general homogeneous and isotropic spin connection may be parametrized by two functions $B(t)$ and $C(t)$,

$$\omega^{0i} = B(t)\,\bar{e}^i, \tag{4.47}$$
$$\omega^{ij} = -C(t)\epsilon^{ijk}\bar{e}_k, \tag{4.48}$$

with indices $i, j, k, ...$ belonging to the representation of $SO(3)$. For the scalar sector, we assume that a gauge may be chosen in which each scalar multiplet is represented by a single real homogeneous component,

$$\phi^{R/L} \longrightarrow \phi^{R/L}(t). \tag{4.49}$$

For instance, if ϕ^L belongs to an adjoint representation of $SU_L(2)$, we may choose a gauge of the form

$$\phi^L = (0, 0, \phi^L). \tag{4.50}$$

In this homogeneous background, and with vanishing background gauge fields, the matter covariant derivatives reduce to ordinary derivatives. Substituting the FLRW ansatz 4.48 into the action gives a reduced background Lagrangian density of the form

$$L_{\rm bg} = 3M_{\rm Pl}^2 \bar{a}\left(\dot{B}\bar{a} + B^2\bar{N} - C^2\bar{N} - \frac{2}{\bar{\gamma}}BC\bar{N} - \frac{1}{\bar{\gamma}}\dot{C}\bar{a}\right) + \bar{a}^3\bar{N}\left(\frac{\Omega^{-2}K_Q}{\bar{N}^2} - \Omega^{-4}V(Q)\right), \tag{4.51}$$

where $Q^A = \{\chi, \phi^{(R)}, \phi^{(L)}\}$ collectively denotes the scalar sector, K_Q is the corresponding quadratic kinetic form. Note that the following conformal transformations have been performed.

$$\bar{e}^I = \Omega e^I, \qquad \Omega = \frac{\bar{M}_{\rm Pl}}{M_{\rm Pl}} \tag{4.52}$$

Then, after the addition of suitable total derivatives, the variables $B(t)$ and $C(t)$ may be eliminated through their own equations of motion in the same manner as before. This yields an effective background dynamics expressed in terms of the scale factor and the scalar fields. In particular, for a single-field trajectory dominated by $\phi^{(R)}$, the reduced Lagrangian takes the schematic form

$$L_{\rm bg} = \bar{a}^3\bar{N}\left[-3M_{\rm Pl}^2\frac{\dot{\bar{a}}^2}{\bar{N}^2\bar{a}^2} + Y(\phi)\frac{\dot{\phi}^2}{\bar{N}^2} - U(\phi)\right], \tag{4.53}$$

where the effective field-space kinetic coefficient is given by

$$Y(\phi) = \frac{1}{V_0(\phi)}\left(1 + \frac{3}{1+\bar{\gamma}^2(\phi)}\frac{1}{V_0(\phi)}\frac{M_{PL}^2}{\phi^2}\right), \tag{4.54}$$

$$V_0(\phi) = 1 + (\zeta+\eta)\frac{\phi^2}{M_{PL}^2}, \tag{4.55}$$

$$\bar{\gamma}(\phi) = i\frac{V_0(\phi)}{(\zeta-\eta)}\frac{M_{PL}^2}{\phi^2} \tag{4.56}$$

Here $U(\phi)$ is the Einstein-frame scalar potential. The viability of inflation is then determined by the shape of $U(\phi)$ together with the non-canonical kinetic factor $Y(\phi)$.

4.2 Perturbative Expansion

Building upon our prior discussion on the mechanism of cosmological perturbation, we shall now perturb the co-tetrad and spin connection around the homogeneous background. The perturbations will be expanded with respect to the background co-tetrad basis $\{\bar{e}^0, \bar{e}^i\}$. Then, the most general linear perturbation can be written as

$$\delta e^0 = \Psi\, \bar{e}^0 + L_i \bar{e}^i, \tag{4.57}$$

$$\delta e^i = N^i \bar{e}^0 + S^{ij} \bar{e}_j. \tag{4.58}$$

Similarly, in its most general form, the perturbation of the spin connection is decomposed as

$$\delta\omega^{0j} = \gamma^i \bar{e}^0 + \xi\, \bar{e}^i + \epsilon^{ijk} \mu_j \bar{e}_k + \tau^{ij} \bar{e}_j, \tag{4.59}$$

$$\delta\omega^{ij} = \epsilon^{ij}{}_k \phi^k \bar{e}^0 + \zeta\, \epsilon^{ij}{}_k \bar{e}^k + \nu^{[i} \bar{e}^{j]} + \epsilon^{ij}{}_k \chi^{kl} \bar{e}_l. \tag{4.60}$$

Under the spatial symmetries of the FLRW background, this decomposition isolates perturbations according to their transformation properties, thereby facilitating their alignment with the scalar-vector-tensor framework established in Chapter 3.

4.2.1 Scalar perturbations

The scalar sector contains perturbations of the tetrad, the spin connection, and the inflaton field that can be recovered from the scalar-vector-tensor decomposition. Schematically, we may write the scalar part of the spatial vector given in Equation 4.57 as

$$L_i = \partial_i L, \tag{4.61}$$

Similarly, we isolate the scalar part S_{ij} of the tetrad perturbation given in 4.58 as

$$S_{ij} = \frac{1}{3}\mathfrak{h}\, \delta_{ij} + \left(\partial_i \partial_j - \frac{1}{3}\delta_{ij}\partial^2\right) \mathcal{S} + \epsilon^{ijk} \partial_k \mathcal{A}, \tag{4.62}$$

To emphasize some of the attributes of the above equation, we note that the first term contains the trace and therefore measures the isotropic scalar deformation of the spatial tetrad. The second term, on the other hand, is the traceless scalar shear, and may be simplified should we choose to work in a gauge that sets $\mathcal{S}$ to zero. The final term is the antisymmetric component that will persist in the most general form of the decomposition in the tetrad language before gauge fixing.

Moreover, the same decomposition may be performed for the scalar parts of the spin connection perturbation, and thus we may use collected field equations for δe^I and $\delta\omega^{IJ}$

to express all scalar perturbations in the spacetime gauge ($\delta\phi \stackrel{*}{=} 0, \mathcal{S} \stackrel{*}{=} 0$) in terms of the perturbation $\mathfrak{h}$.

Note that this formulation is the analogue of the comoving gauge in the single-field description, where the inflaton fluctuation is removed by the time reparametrization, and the scalar degree of freedom is instead carried by the metric perturbation h.

Equation of motion for $\mathfrak{h}$ in Fourier space can then be obtained from the following action

$$S_S[\mathfrak{h}] = \frac{1}{9}\int dt d^3k \bar{N}\bar{a}^3 \frac{Y(\phi)\dot{\phi}^2}{(\dot{\bar{a}}/\bar{a})^2}\left(\frac{1}{\bar{N}^2}\dot{\mathfrak{h}}^*\dot{\mathfrak{h}} - \frac{1}{\bar{a}^2}k^2\mathfrak{h}^*\mathfrak{h}\right) \tag{4.63}$$

Choosing conformal time, $\bar{N} = \bar{a}$, and denoting respective derivatives with prime, we may bring the action into canonical form using the following definitions

$$\tilde{\nu} = \frac{Z}{3}h, \qquad Z^2 = \frac{2\bar{a}^2 Y(\phi)\phi'^2}{\left(\bar{a}'/\bar{a}\right)^2}. \tag{4.64}$$

Then the action becomes

$$S_S[\tilde{\nu}] = \frac{1}{2}\int d\tau\, d^3k \left[\tilde{\nu}^{*\prime}\tilde{\nu}' - \left(k^2 - \frac{Z''}{Z}\right)\tilde{\nu}^*\tilde{\nu}\right]. \tag{4.65}$$

This is the Mukhanov–Sasaki action [41] for a single effective scalar degree of freedom. At the level of scalar perturbations, we are pleasantly surprised to see that the left-right Einstein–Cartan model reduces to the familiar single-field structure, and that the modifications are solely restricted to the kinetic function $Y(\phi)$. All is well that ends well!

4.2.2 Tensor perturbations

Incentivized by the earlier outcome, we now turn our attention to the tensor sector: the transverse and traceless components of the spatial rank-two perturbations. Accordingly, the relevant variables are extracted from

$$S^{ij}, \qquad \tau^{ij}, \qquad \chi^{ij}.$$

We define

$$\mathfrak{S}^{ij} := S^{ij}\Big|_{\mathrm{TT}}, \tag{4.66}$$

$$\mathcal{B}^{ij} := \frac{1}{\bar{\gamma}\,\bar{a}^2}\left(\chi^{ij} + \bar{\gamma}\tau^{ij}\right)\Big|_{\mathrm{TT}}. \tag{4.67}$$

where the transverse-traceless conditions are

$$\mathfrak{S}^{ij} = \mathfrak{S}^{ji}, \qquad \mathfrak{S}^{i}{}_{I} = 0, \qquad \partial_i \mathfrak{S}^{ij} = 0, \tag{4.68}$$

for $\mathfrak{S}$ and similarly for $\mathcal{B}^{ij}$. At quadratic order in perturbations, the tensor variables $\{\mathfrak{S}^{ij}, \mathcal{B}^{ij}\}$ decouple from the scalar and vector sectors such that the tensor action can be studied independently,

$$S_{\mathrm{T}}[\mathfrak{S}, \mathcal{B}] = \int d^4x\, \mathcal{L}_{\mathrm{T}}, \tag{4.69}$$

where $\mathcal{L}_{\mathrm{T}}$ is obtained by substituting Eqs. (4.57)–(4.60) into the full action given by 4.41 and retaining only terms quadratic in the transverse-traceless fields.

In fact, in component form, the part of the action relevant for the tensor sector is built from the two contractions

$$\mathcal{I}_{\epsilon} := \tilde{\epsilon}^{\mu\nu\alpha\beta} \epsilon_{IJKL} \bar{e}^{I}{}_{\mu} \bar{e}^{J}{}_{\nu} R^{KL}{}_{\alpha\beta}(\omega), \tag{4.70}$$

$$\mathcal{I}_{\gamma} := \tilde{\epsilon}^{\mu\nu\alpha\beta} \eta_{IK} \eta_{JL} \bar{e}^{I}{}_{\mu} \bar{e}^{J}{}_{\nu} R^{KL}{}_{\alpha\beta}(\omega), \tag{4.71}$$

corresponding respectively to the Einstein–Cartan and Holst contribution.

The action is then obtained by expanding these objects to second order around the homogeneous background. Since the action contains the product $\bar{e}\,\bar{e}\,R(\omega)$, the terms quadratic in tensor perturbations arise, after some algebra[3], from

$$\delta\bar{e}\,\delta\bar{e}\,\bar{R}, \qquad \bar{e}\,\delta\bar{e}\,\delta R, \qquad \bar{e}\,\bar{e}\,\delta^2 R. \tag{4.72}$$

The third class contains the second-order part of the spin- connection curvature. Thus, we also need the curvature of the spin connection defined by

$$R^{IJ}{}_{\mu\nu} = 2\partial_{[\mu}\omega^{IJ}{}_{\nu]} + 2\omega^{I}{}_{K[\mu}\omega^{KJ}{}_{\nu]}. \tag{4.73}$$

We note that $R^{IJ}(\omega)$ contains a first derivative of the connection, the resulting quadratic action is thus first order in time derivatives of the tensor connection variable. After integrations by parts and after discarding boundary terms, the only time-derivative term may be written as $-M_{\mathrm{Pl}}^2 \bar{a}\, \mathfrak{S}_{ij} \mathcal{B}'^{\,ij}$. The remaining terms contain no time derivatives of $\mathcal{B}_{ij}$ and are collected into a Hamiltonian density $\mathcal{H}(\mathfrak{S}, \mathcal{B})$

$$\mathcal{L}_T = -M_{\mathrm{Pl}}^2 \bar{a}\, \mathfrak{S}_{ij} \mathcal{B}'^{\,ij} - \mathcal{H}(\mathfrak{S}, \mathcal{B}), \tag{4.74}$$

[3]Out of solidarity with all members of the community who have once dutifully tried to verify easily shown results, a more complete derivation is relegated to the Appendix

where the Hamiltonian density is given by.

$$
\begin{aligned}
\mathcal{H} = \frac{M_{\rm Pl}^2}{2}\bar{a}^2 \Bigg[& \frac{1}{\bar{a}^2}\mathcal{B}_{ij}\mathcal{B}^{ij} - \frac{\bar{a}''}{\bar{a}}\mathfrak{S}_{ij}\mathfrak{S}^{ij} \\
& - \frac{1+\bar{\gamma}^2}{\bar{\gamma}^2}\left(2\mathfrak{S}^{ij}\partial_k\partial^k\mathfrak{S}_{ij} + \partial_k\mathfrak{S}_{ij}\partial^k\mathfrak{S}^{ij}\right) \\
& - \frac{2}{\bar{\gamma}}\epsilon^{ijk}\frac{1}{\bar{a}}\mathfrak{S}_{\ell i}\partial_k\mathcal{B}_j{}^{\ell} + \frac{\bar{\gamma}'}{\bar{\gamma}^2}\epsilon^{ijk}\mathfrak{S}_{\ell i}\partial_k\mathfrak{S}_j{}^{\ell}\Bigg].
\end{aligned}
\tag{4.75}
$$

This delineates two distinct analytical perspectives. While the first will be examined rigorously, the latter will be explored through a more qualitative approach. In fact, at the classical level, we may eliminate the connection-sector tensor variable $\mathcal{B}_{ij}$. Its equation of motion is algebraic, and substituting the resulting solution back into the action gives the standard second-order tensor action

$$
S_{\rm T}[\mathfrak{S}] = \frac{M_{\rm Pl}^2}{2}\int d\tau\, d^3x\, \bar{a}^2\left[\mathfrak{S}'_{ij}\mathfrak{S}'^{\,ij} - \partial_k\mathfrak{S}_{ij}\partial^k\mathfrak{S}^{ij}\right]. \tag{4.76}
$$

Thus, in this reduced description, the explicit dependence on $\bar{\gamma}$ disappears from the classical tensor action, providing the same structure as in General Relativity on an FLRW background.

Given that the scalar sector has already been reduced to the Mukhanov–Sasaki form, the leading inflationary observables may be computed using the standard single-field slow-roll formulae, provided they are evaluated with the effective Einstein-frame potential and canonical field normalization of the present model. Accordingly, this line of inquiry establishes the basis of our inaugural approach.

4.3 Inflationary Parameters on FLRW Background

The relevant background quantities to consider are the canonically normalized field χ, defined by

$$
\frac{d\chi}{d\phi} = \sqrt{Y(\phi)}, \tag{4.77}
$$

and the Einstein-frame potential $U(\chi)$. The potential slow-roll parameters are then, to leading order, given in reduced Planck units, $M_{\rm pl} = 1$ by [44]

$$
\epsilon_V = \frac{1}{2}\left(\frac{U_{,\chi}}{U}\right)^2, \qquad \eta_V = \frac{U_{,\chi\chi}}{U}. \tag{4.78}
$$

Finally, the scalar spectral index, tensor-to-scalar ratio, and scalar amplitude are computed as

$$n_s = 1 - 6\epsilon_V + 2\eta_V, \qquad r = 16\epsilon_V, \qquad A_s \simeq \left. \frac{U}{24\pi^2 \epsilon_V} \right|_{\chi=\chi_*}. \tag{4.79}$$

Here χ_* denotes the field value at horizon exit, fixed by the condition that N_* e-folds remain before the end of inflation.

Our final aim is twofold. First, we strive to determine whether the effective single-field model admits viable inflationary trajectories compatible with current observational constraints. Second, we examine the role of the effective Barbero–Immirzi parameter to ascertain whether its evolution changes the observational outcome. The analysis is performed for two distinct branches

$$\bar{\gamma} \in \mathbb{R}$$
$$\bar{\gamma} = i\gamma_I \in i\mathbb{R}$$

The real branch is the direct continuation of the real Holst-inflation background analysis. The purely imaginary branch is motivated by the self-dual/Ashtekar interpretation of the first-order tensor phase space.

In both cases, the background study asks the same question: for which values of the model parameters does the effective Einstein-frame potential support slow-roll inflation with acceptable (n_s, r, A_s)?

4.3.1 Methodology

The general pipeline for both cases is identical. The background analysis was carried out with an emulator-assisted Bayesian pipeline. The exact numerical backend constructs the equations of motion, solves the background scalar-field trajectory, and finally constructs the Einstein-frame potential $U(\chi)$. This will then allow us to compute the slow-roll variables, locate the end of inflation, identify the horizon-exit point, and ultimately find the values of each observable (n_s, r, A_s).

The sampled parameter vector is[4]

$$\text{☕} = \left(\log_{10} q, \log_{10} \gamma_I, u, \phi_0, \dot{\phi}_0\right), \qquad u = \log_{10}\left(\frac{\lambda}{q}\right). \tag{4.80}$$

where $q = \zeta + \eta$, controls the overall scale of the non-minimal coupling. The use of $u = \log_{10}(\lambda/q)$ is motivated by the fact that the viable region is more sharply organized by the ratio λ/q than by λ and q separately. This reparameterization, therefore, gives a better-adapted coordinate system for active learning, regression, and posterior sampling.

[4]Humbly dedicated to the roasted beans that fueled this intellectual journey!

The working prior intervals used in the scan are

$$\log_{10} q \in [-10, 10], \tag{4.81}$$

$$\log_{10} \gamma_I \in [-1, 10], \tag{4.82}$$

$$u \in [-11.5, -10.0], \tag{4.83}$$

$$\phi_0 \in [10, 20], \tag{4.84}$$

$$\dot{\phi}_0 \in [-0.1, 0.1]. \tag{4.85}$$

The observational viability cuts are chosen to be compatible with current CMB constraints [31]

$$n_s = 0.9649 \pm 0.0042 \tag{4.86}$$

at 68% confidence, the tensor-to-scalar ratio close to[5]

$$r_{0.002} < 0.056. \tag{4.87}$$

Accordingly, we use a narrow interval around the observed scalar amplitude.

$$A_s \simeq 2.1 \times 10^{-9} \tag{4.88}$$

The numerical analysis is carried out in multiple stages. First, exact Mathematica evaluations were used to build a dataset of successful and unsuccessful points. A point was classified as successful if the background evolution, canonical normalization, and Einstein-frame potential could all be constructed consistently. It was then classified as viable if the resulting observables lay inside the observationally allowed region. Second, an emulator was trained on these exact evaluations and used to guide a posterior exploration of the viable region.

Then, a tree-histogram gradient-boosted regressor trained on the viable subset was used to predict the observable values as

$$n_s, \qquad \log_{10} r, \qquad \log_{10} A_s. \tag{4.89}$$

We use logarithmic variables for r and A_s because these quantities vary over several orders of magnitude, whereas n_s is naturally treated on a linear scale. The regressors were validated using train-test splits and cross-validation before being used in the posterior exploration.

Posterior sampling was then performed using an affine-invariant ensemble sampler [69]. The likelihood was taken to be Gaussian in n_s and $\log_{10} A_s$, while the tensor-to-

[5]Although the numerical pipeline used the conservative cut $r < 0.06$, we will see that the resulting validated points satisfy the tighter BICEP/Keck 2018 bound [68]– giving $r_{0.05} < 0.036$ at 95% confidence – by a large margin.

scalar ratio was imposed as a hard upper bound $r < 0.06$. Explicitly, the likelihood has the following form

$$\ln \mathcal{L}(\coffeecup) = -\frac{1}{2}\left(\frac{n_s(\coffeecup) - n_s^{\mathrm{obs}}}{\sigma_{n_s,\mathrm{eff}}}\right)^2 - \frac{1}{2}\left(\frac{\log_{10} A_s(\coffeecup) - \log_{10} A_s^{\mathrm{obs}}}{\sigma_{\log_{10} A_s,\mathrm{eff}}}\right)^2, \tag{4.90}$$

with proposals failing the condition $r < 0.06$ rejected. The effective uncertainties combine observational errors with the interpolation error of the regressors,

$$\sigma_{\mathrm{eff}}^2 = \sigma_{\mathrm{obs}}^2 + \sigma_{\mathrm{emu}}^2. \tag{4.91}$$

Significantly, two safeguards were imposed during the MCMC sampling. First, proposals were required to have a sufficiently high probability of belonging to the numerically successful region. Second, a nearest-neighbour trust region was imposed relative to the viable training set. The expected outcome of the pipeline is therefore a posterior map of the viable background region. Finally, a set of representative posterior points was re-evaluated with Mathematica.

The scope of this investigation is, however, subject to some limitations that must be considered. First, the numerical analysis necessarily depends on the choice of finite prior intervals. These domains are chosen to cover the region in which the background equations can be solved reliably and in which viable slow-roll solutions are expected to occur. After all, as Nate Silver articulates in his book *The Signal and the Noise* [6], "Under Bayes' theorem, no theory is perfect, rather it is a work in progress, always subject to further refinement."

A second limitation concerns the finite numerical resolution of the emulator-assisted posterior. The MCMC was run on an interpolating representation of the viable region, with safeguards designed to keep the chain close to points supported by exact numerical evaluations. Increasing the number of exact training points, running longer chains, or using more walkers would improve the smoothness of the posterior and reduce residual sampling noise.

4.3.2 Case of $\bar{\gamma} \in i\mathbb{R}$

To the best of our knowledge, the numerical background analysis presented here has not previously been carried out for an effective left-right symmetric inflationary model with a purely imaginary Barbero–Immirzi parameter. Existing studies of Higgs inflation in first-

[6]For the astute reader, having introduced this book, I feel acutely responsible to caveat that while Nate Silver is a brilliant statistician, his book's chapter on climate change, uncritically amplifies figures like A. Pielke Jr., a climate skeptic tied to the conservative think tanks. Ironically, this makes him the first to fall into the trap of prior biases!

order gravity have, however, examined the effect of Holst coupling on the inflationary background. Still, they do not provide a parameter-space analysis of the imaginary-$\bar{\gamma}$ branch generated by the left-right scalar sector. This sector is relevant from the first-order point of view, since imaginary values of the Barbero–Immirzi parameter are naturally associated with self-dual or Ashtekar-like variables.

As stated before, we wish to determine whether this case admits viable inflationary parameters and the extent to which our observable predictions are affected by the choice of the $\bar{\gamma}$ parameter.

The corner plot 4.2 first gives a global view of the posterior structure in the sampled parameter space. The most pronounced localization occurs in the combination $\log_{10}(\lambda/q)$ which is concentrated around $u \simeq -10.76$. The parameter $\log_{10} q$ is also constrained, with support concentrated around values of order 10^6 in the posterior median, while the initial conditions ϕ_0 and $\dot{\phi}_0$ remain comparatively broad. This suggests that the viable inflationary region is mainly selected by the effective shape and normalization of the Einstein-frame potential, rather than by a finely tuned initial trajectory.

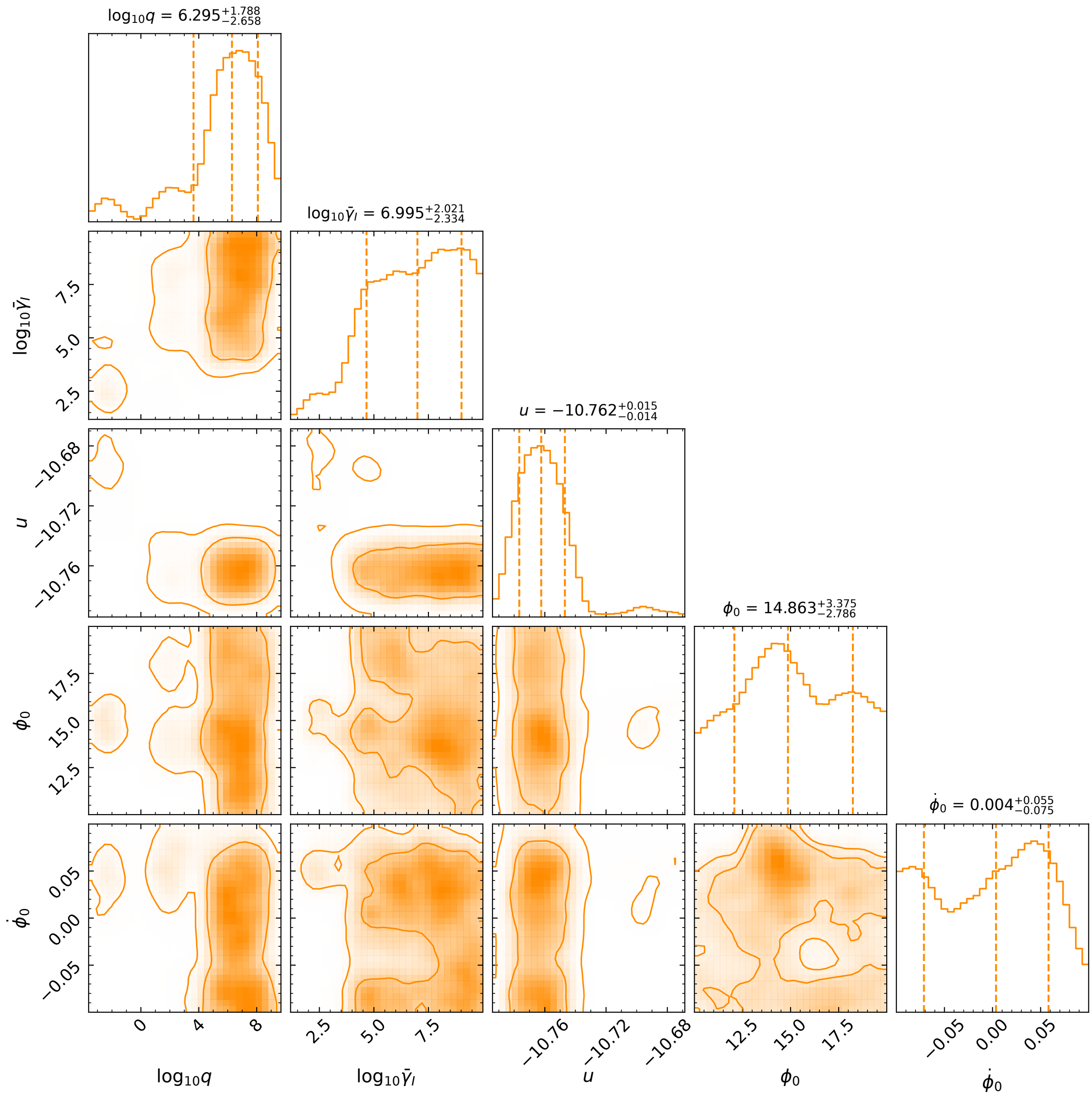


Figure 4.2: Posterior distribution of the sampled parameters in the imaginary-$\bar{\gamma}$ branch. The most pronounced localization occurs in $u = \log_{10}(\lambda/q)$.

The parameter-sensitivity analysed in Figure 4.3 supports this interpretation. The regressor feature-importance plot shows that $\log_{10} q$ is the dominant variable controlling n_s and $\log_{10} r$, while $u = \log_{10}(\lambda/q)$ is by far the dominant variable controlling $\log_{10} A_s$ which is a consequence of u fixing the effective scale of the potential. On the other hand, q controls the flattening of the Einstein-frame potential and hence affects the tilt and tensor suppression. By contrast, the direct importance of $\log_{10} \gamma_I$ is small for all three observables. This indicates that the scalar observables are not primarily shaped by continuous variations of γ_I inside the viable region.

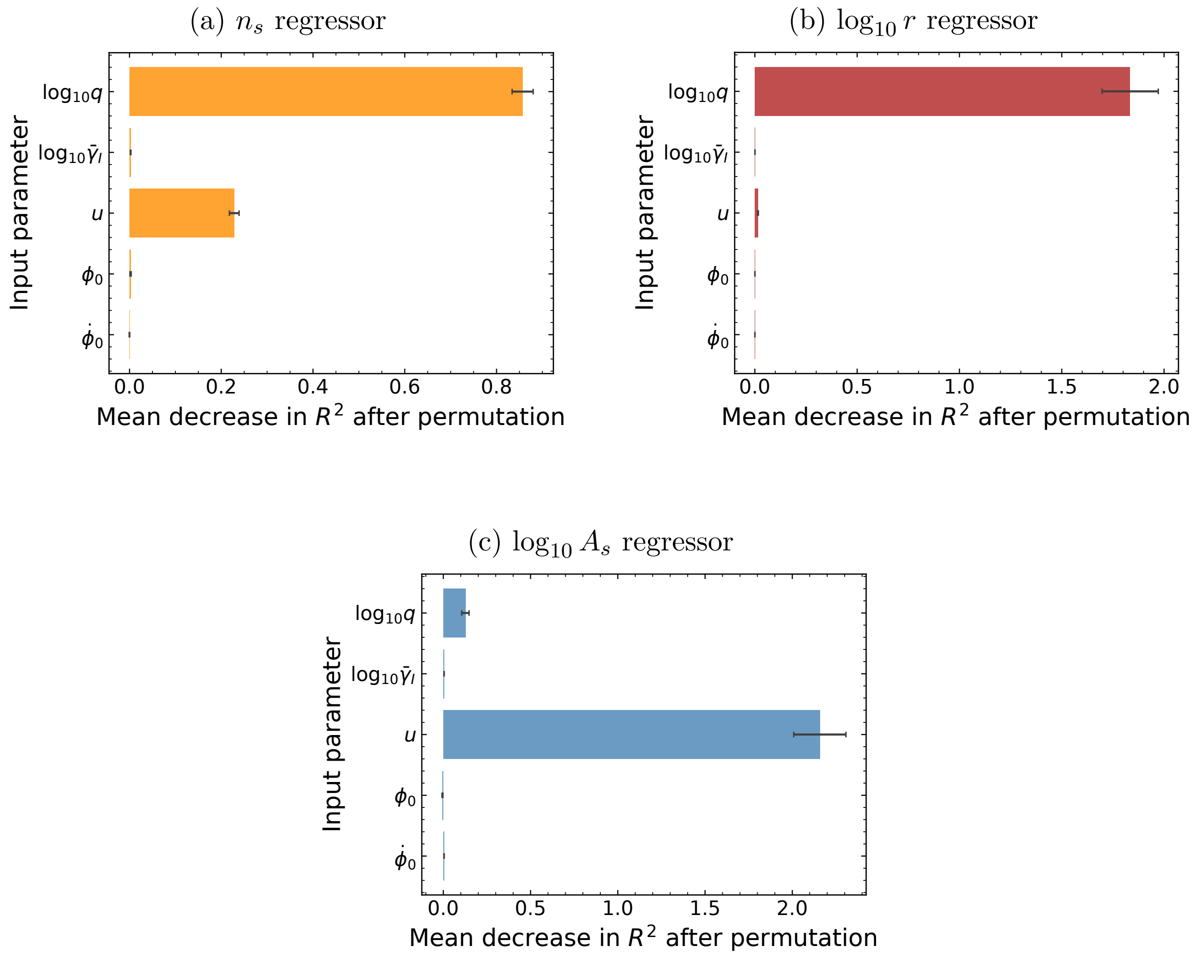


Figure 4.3: Regressor feature importance for the imaginary-$\bar{\gamma}$ branch. Permutation importance of the emulator input parameters for the regressors predicting n_s, $\log_{10} r$, and $\log_{10} A_s$. For each feature, the plotted value is the mean decrease in the test-set R^2 score after randomly permuting that feature, with error bars showing the standard deviation across repeated permutations. The parameter $\log_{10} q$ dominates the prediction of n_s and $\log_{10} r$, while $u = \log_{10}(\lambda/q)$ is the dominant variable controlling $\log_{10} A_s$. The feature importance of $\log_{10} \gamma_I$ is small for all three observables.

The parameter γ_I plays a subtle role here. Comparing the posterior to the uniform prior yields $D_{\rm KL} \simeq 0.409$ where $D_{\rm KL}$ denotes the Kullback–Leibler divergence [70] between the posterior and the prior. In a Bayesian numerical analysis, it measures how much information is gained about a parameter after imposing the data and viability conditions. Thus, $D_{\rm KL} = 0$ would mean that the posterior is indistinguishable from the prior, while a positive value indicates that the constraints have reshaped the prior volume. In the present case, it seems to indicate that the posterior is mildly informative, but not sharply localized.

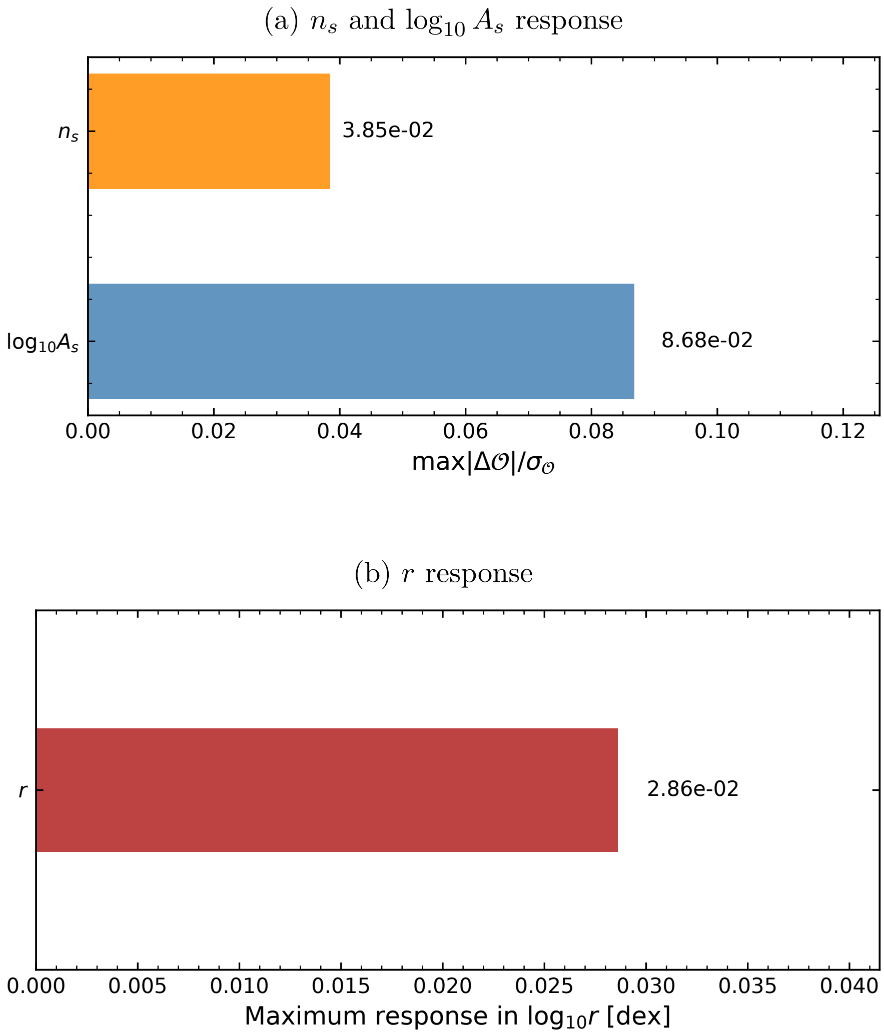


Figure 4.4: Observable response to varying $\log_{10}\bar{\gamma}$ at fixed posterior-median values of the remaining parameters. (a) For n_s and $\log_{10} A_s$, the bar height shows $\max|\Delta\mathcal{O}|/\sigma_{\mathcal{O}}$ where $\mathcal{O}$ denotes the respective observable, i.e., the largest predicted shift normalized by the corresponding observational 1σ uncertainty. (b) The r response is shown separately as $\max|\Delta \log_{10} r|$ in dex, since r is constrained by an upper bound rather than by a Gaussian measurement. One dex corresponds to a factor of ten in r.

At first sight, this may seem in tension with the permutation-importance result. Indeed, if γ_I does not significantly affect n_s, r, or A_s, as shown in Figure 4.4, why is its posterior not uniform? The resolution is mainly rooted in the fact that $\bar{\gamma}$ affects the domain of admissible backgrounds. Some values of γ_I, especially toward the low-γ_I region, make it more likely that the canonical normalization becomes pathological, for example through $Y(\phi) < 0$, or that the canonical field interval lies outside the slow-roll regime. Thus γ_I maintains some influence on whether a point enters that viable region at all.

The quantification of the previous claim was attained by a targeted sensitivity test in which posterior-like samples were selected from viable $\log_{10} q$, u, ϕ_0, and $\dot{\phi}_0$, while only $\log_{10}\gamma_I$ was forced into the low-tail interval.

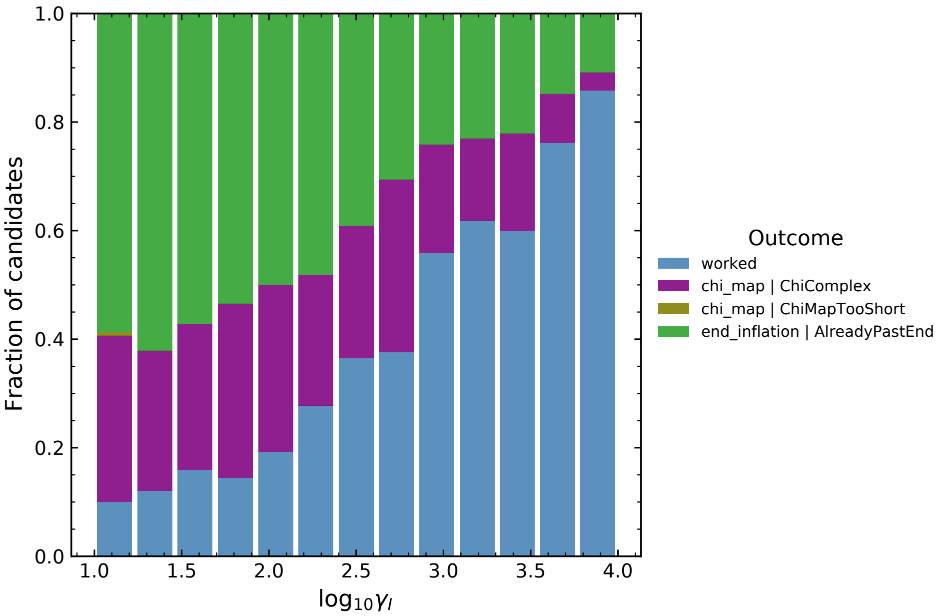


Figure 4.5: Failure-mode composition in the low-γ_I tail test. The vertical axis gives the fraction of tested candidate points, while the horizontal axis shows $\log_{10}\gamma_I$.

These points were then re-evaluated with the full Mathematica solver and classified according to their failure mode to isolate the role of $\bar{\gamma}_I$. We found that if the points fail predominantly at the numerical-evolution stage, the suppression is likely solver-related; if they fail through conditions explicitly defined in Table 4.1, such as `AlreadyPastEnd` or `NoEndFound`, the low-γ_I deformation changes the inflationary trajectory itself; and if they work but are not viable, then low γ_I remains numerically admissible but moves the model outside the observationally allowed region. For failures of type `ChiComplex`, we additionally constructed matched pairs in which each failing low-$\bar{\gamma}_I$ point was duplicated with all parameters fixed except γ_I, which was replaced by a representative high-γ_I value from the posterior bulk. This diagnostic was performed to test whether the complex canonical-field map arises from the low-γ_I deformation itself. In particular, if the low-γ_I point gives $Y(\phi) < 0$ while its high-γ_I twin remains real, the failure indicates a ghost-like region in the kinetic sector rather than a generic numerical integration problem.

Table 4.1: Interpretation of Mathematica failure modes in the low-$\bar{\gamma}_I$ tail test.

Failure mode	**Interpretation**
`AlreadyPastEnd`	The initial condition is already outside the inflationary regime.
`NoEndFound`	The evolution does not reach the end-of-inflation condition, $\epsilon = 1$
`PhiNDSolveFailed`	The scalar-field evolution failed at the numerical ODE-solving stage.
`ChiComplex`	The canonical-field map $\chi(\phi)$ became complex.
`BadPotentialGrid`	The Einstein-frame potential or its interpolation grid could not be constructed reliably.
`ChiMapTooShort`	The available $\chi(\phi)$ map does not cover the range required by the inflationary trajectory and canonical-field interpolation is insufficient.

Finally, Figure 4.6 presents the viable candidate points in the (n_s, r) plane. The result shows that the successful backgrounds are concentrated in a narrow vertical band around $n_s \simeq 0.966$, placing them within the Planck-compatible scalar-tilt region. At the same time, the tensor-to-scalar ratio is strongly suppressed, extending to values many orders of magnitude below the current observational upper bounds.

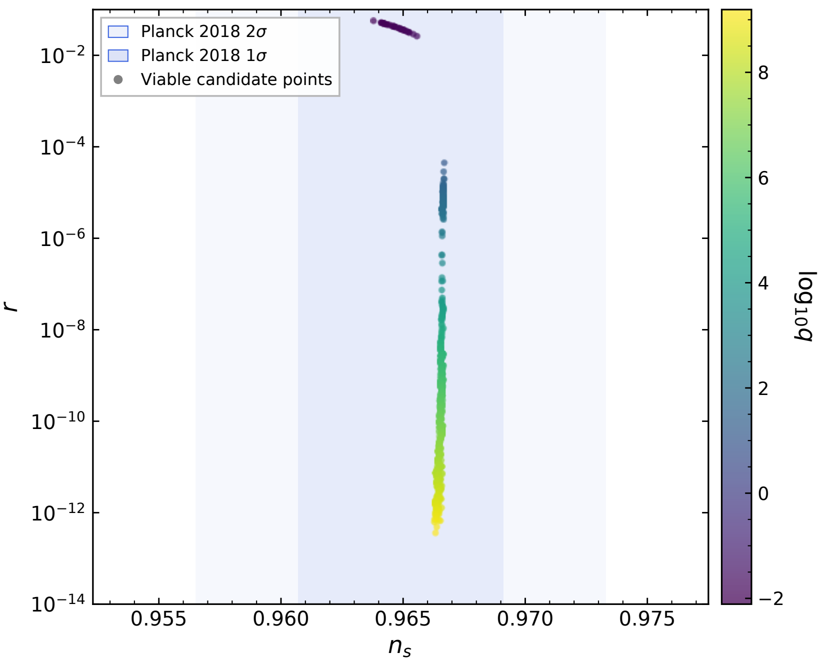


Figure 4.6: Viable candidate points in the (n_s, r) plane. The shaded bands indicate the Planck-compatible scalar-tilt intervals. The viable points are concentrated around $n_s \simeq 0.966$ and predict a strongly suppressed tensor-to-scalar ratio.

The colour scale indicates the value of $\log_{10} q$. The visible ordering of the points shows that q plays an important role in controlling the tensor amplitude: larger values of q correspond to smaller values of r, which is consistent with the feature-importance analysis, where $\log_{10} q$ dominates the prediction of both n_s and $\log_{10} r$.

4.3.3 Case of $\bar{\gamma} \in \mathbb{R}$

This scenario suffers from extreme similarity to the former case. Since the numerical pipeline, observables, and viability criteria are the same, we shall not repeat the full methodological discussion. We will instead seek to isolate the differences between the real and imaginary branches, and to determine whether the real-$\bar{\gamma}$ case displays a comparable sensitivity to the Immirzi sector.

The real branch shows a broadly similar phenomenological behaviour in the (n_s, r) plane seen in Figure 4.7. The viable points again cluster around a Planck-compatible scalar spectral index, $n_s \simeq 0.966$, while the tensor-to-scalar ratio is strongly suppressed across most of the viable region.

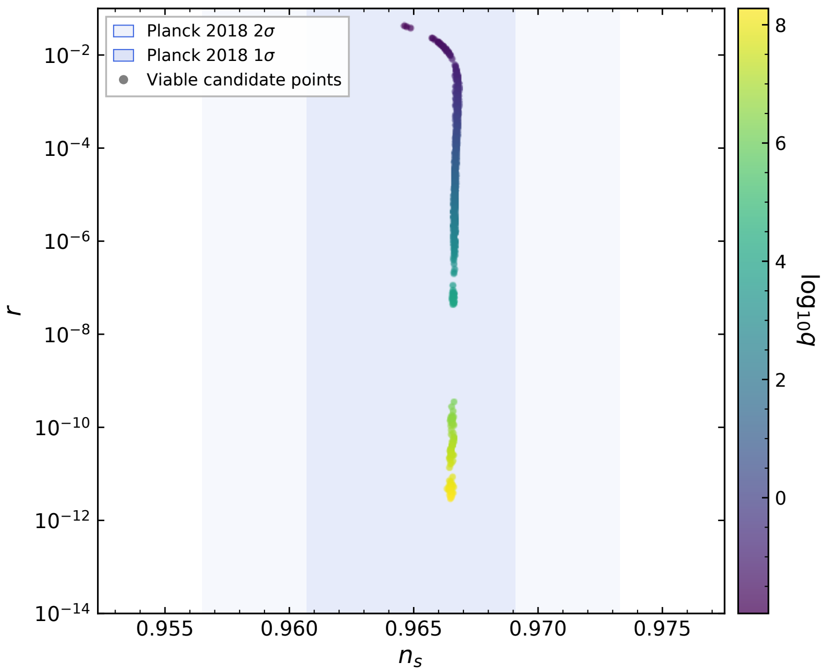


Figure 4.7: Viable candidate points in the (n_s, r) plane for the real $\bar{\gamma}$ branch.

Additionally, we see in Figure 4.8 – as in the imaginary case – that $\log_{10} q$ dominates the prediction of n_s and $\log_{10} r$, while $u = \log_{10}(\lambda/q)$ dominates the prediction of $\log_{10} A_s$. The direct importance of $\log_{10} \bar{\gamma}$ is negligible for all three observables.

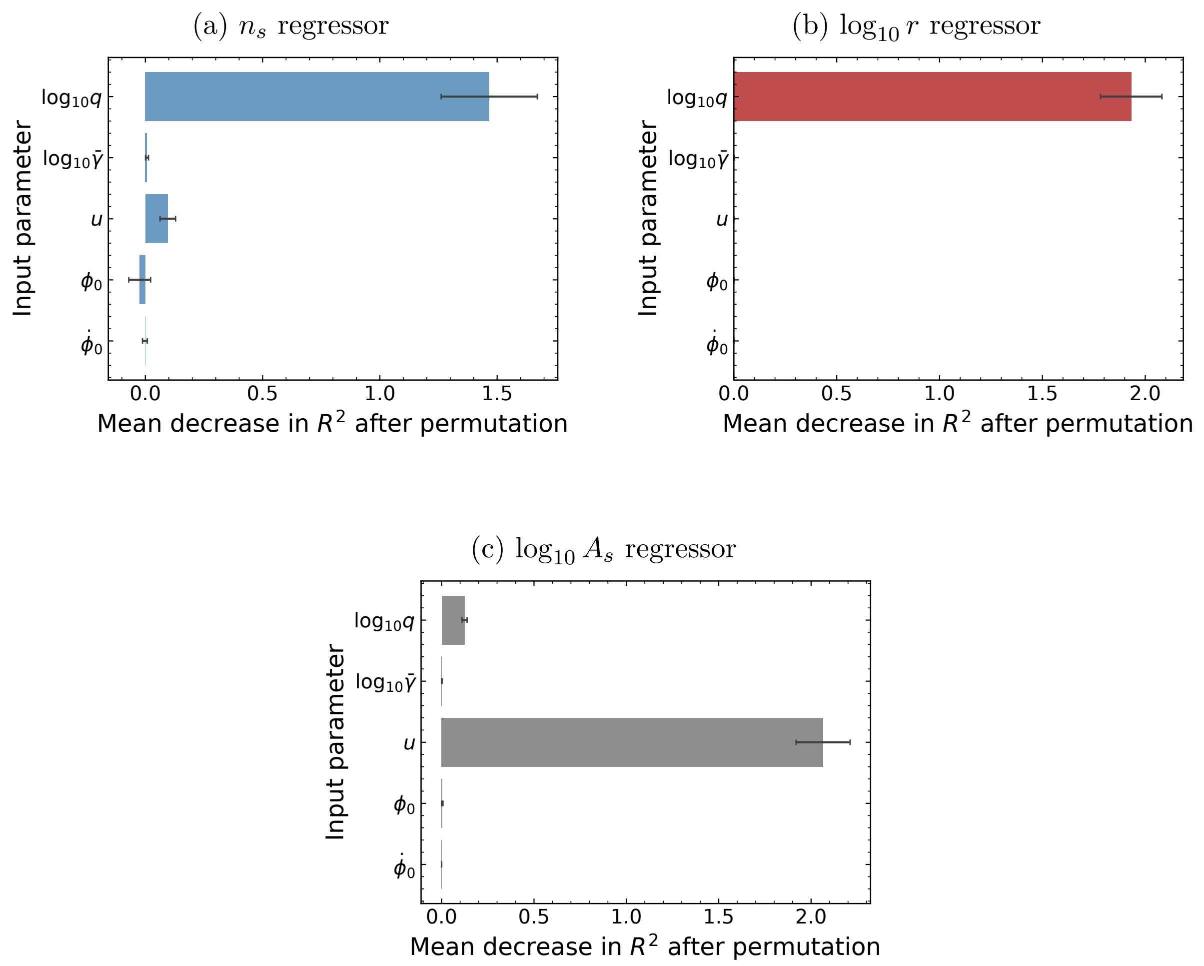
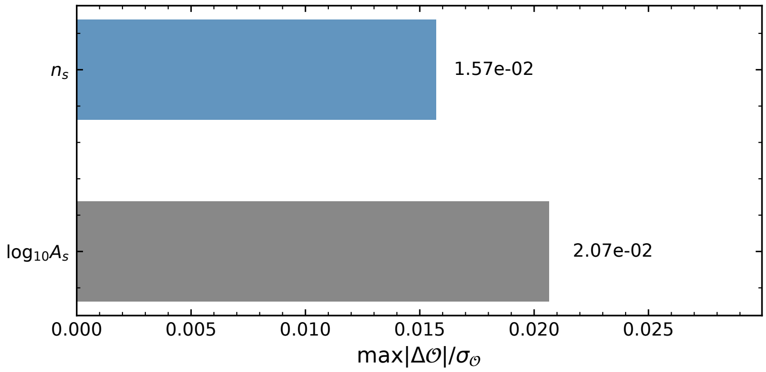


Figure 4.8: Regressor feature importance for the real-$\bar{\gamma}$ branch.

This is further supported by the local response test where varying $\log_{10} \bar{\gamma}$ near the posterior point changes n_s by only 1.57×10^{-2} of the observational 1σ scale, changes $\log_{10} A_s$ by 2.07×10^{-2} of the corresponding scale, and changes $\log_{10} r$ by only 3.34×10^{-2} dex shown in Figure 4.9.

(a) n_s and $\log_{10} A_s$ response



(b) r response

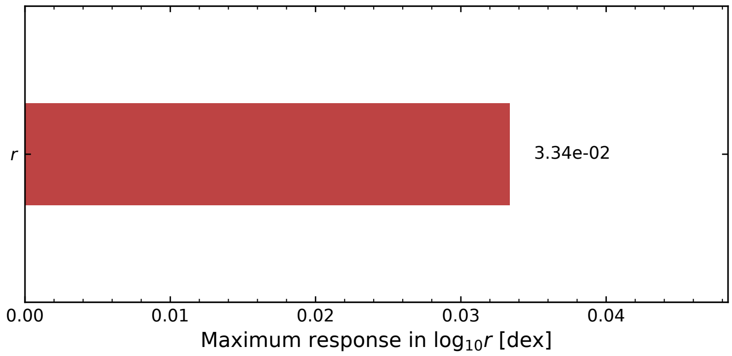


Figure 4.9: Local response of the observables to variations of $\log_{10}\bar{\gamma}$ in the real branch.

The main difference concerns the low-$\bar{\gamma}$ tail test whose analogous diagnostic for real-$\bar{\gamma}$ tail test should be interpreted with some caution. In the imaginary branch, the low-γ_I region showed a substantial suppression of successful points, mostly through domain failures such as a complex canonical field map or trajectories already outside the slow-roll regime. By contrast, in the real-$\bar{\gamma}$ tail, diagnostics of uniformly sampled values concentrated close to the drop in the distribution, in Figure 4.10, indicate that almost all random candidates are successful, with the exception of some isolated failures. This indicates that the real branch is locally stable in the posterior region, but it does not by itself establish that the entire upper-$\bar{\gamma}$ domain is equally viable.

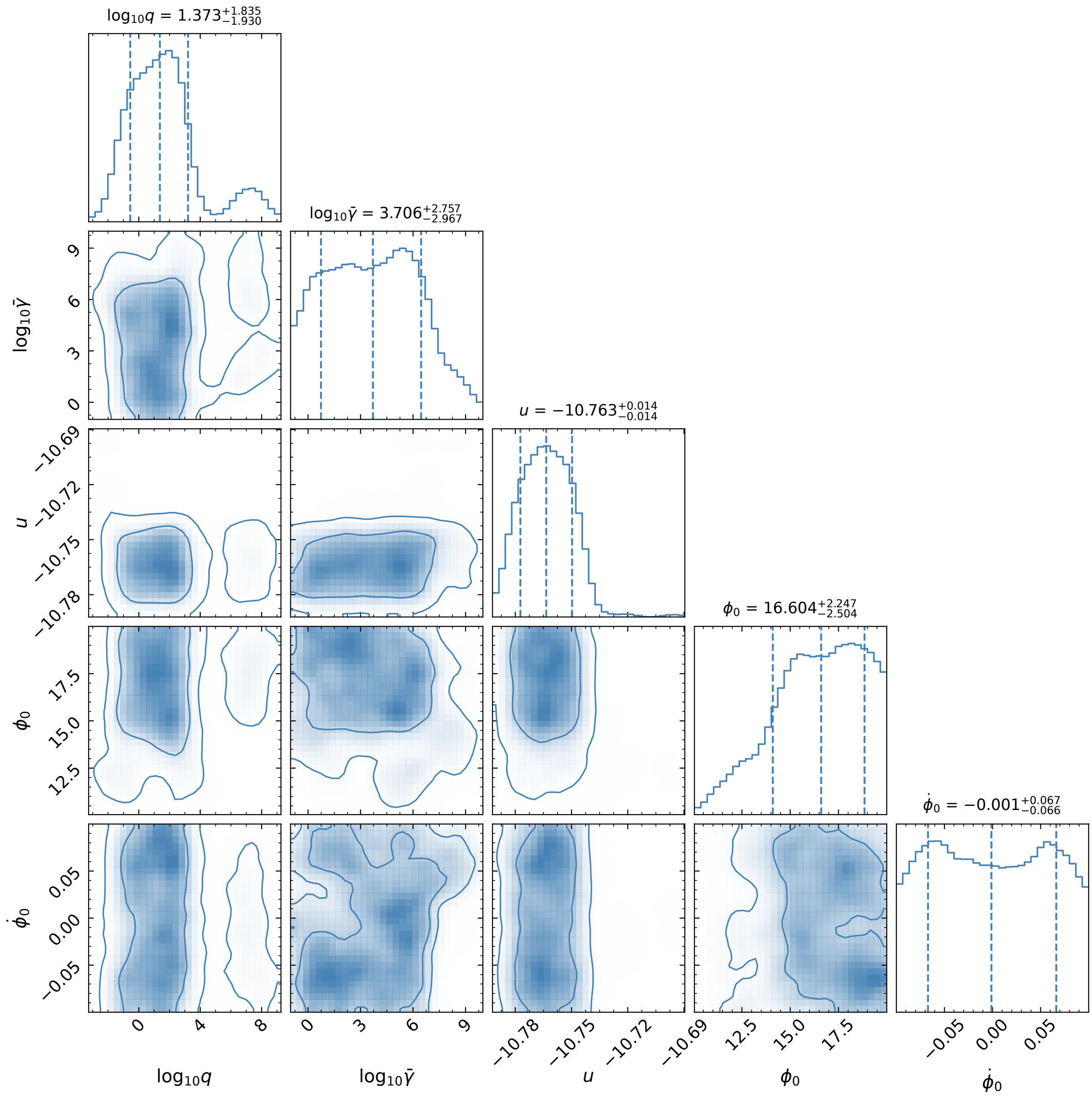


Figure 4.10: Posterior distribution of the sampled parameters in the real-$\bar{\gamma}$ branch.

A possible interpretation is that the MCMC has moved away from regions that are less favored by the combined likelihood and viability filters. The subsequent random test, therefore, probes the neighborhood selected by the posterior rather than the full prior tail, and once the remaining parameters are placed in the posterior-supported region, varying $\bar{\gamma}$ over the tested interval does not typically destroy background admissibility.

Rather than a quixotic pursuit of the entire parameter space, this analysis should be thought of as an expedition to the most viable regions. As is the way with finite things, the resolution of this analysis is constrained by the number of exact background evaluations used to train the emulator and by the length of the posterior sampling.

Moreover, further refinement would require additional exact Mathematica evaluations, especially near the boundaries between successful and failed trajectories. A proposal would be to use an adaptive sampling strategy to target regions where the emulator uncertainty is largest or where the posterior transitions sharply between viable and non-viable behavior.

4.4 Further Discussion

4.4.1 Alternative first-order quantization

Departing from the established path, we now turn our attention to a parallel possible interpretation of the tensor sector. This is achieved by keeping B_{ij} as part of the tensor phase space and quantizing the first-order Hamiltonian system directly. This is motivated by the Ashtekar-variable treatment of inflationary tensor perturbations, where the connection is retained as a canonical variable. This approach, developed in [71], demonstrates that quantization leads to extra degrees of freedom, which are then projected out by imposing the necessary reality conditions. The occurring novelty is due to the fact that the Barbero–Immirzi parameter, having dropped out of the classical field equations, will reappear in the quantum theory through the choice of canonical variables, the Hamiltonian ordering, and the implementation of the reality conditions. The resulting effect is that although the physical spectrum still contains the standard graviton degrees of freedom, the vacuum energy and vacuum fluctuations become chiral and this is potentially measurable through the associated production of correlations between temperature anisotropies and B-mode polarization of the CMB [72]. Ultimately, this will introduce a chiral asymmetry between the primordial power spectra of left and right helicity gravitons, to a degree determined by the imaginary part of $\bar{\gamma}$.

As the discerning reader will have undoubtedly noticed, careful consideration is required to generalize the results of this development to the present model, where $\bar{\gamma}$ is not a constant of the theory; its temporal variation derives from its dependence on the inflaton field. Indeed, starting from the effective expression 4.56 and using

$$\zeta - \eta = \frac{iA}{\bar{\gamma}}, \qquad A = \frac{1}{\phi_0^2} + \frac{q}{M_{\rm Pl}^2}, \tag{4.92}$$

we can reconstruct the effective Barbero–Immirzi parameter in Planck units as

$$\gamma(\phi) = \bar{\gamma}\frac{1 + q\phi^2}{\left(\frac{1}{\phi_0^2} + q\right)\phi^2}. \tag{4.93}$$

such that the scanned value $\bar{\gamma}$ is the effective value at the initial field position.

$$\gamma(\phi_0) = \bar{\gamma}, \tag{4.94}$$

Consequently, it becomes imperative to address whether this field dependence is dynamically important during the inflationary evolution. To quantify this, we define the fractional departure

$$\Delta_\gamma(\phi) = \left|\frac{\gamma(\phi)}{\bar{\gamma}}\right| - 1. \tag{4.95}$$

As it turns out, for a representative viable trajectory, this quantity remains below

$$\Delta_\gamma \lesssim 2.5 \times 10^{-7}. \tag{4.96}$$

Thus, in a fortuitous turn of events that routinely eludes empirical science, variation of $\bar{\gamma}$ along this trajectory is negligible for the background evolution. This conclusion opens the path for further investigation of the model under the assumptions of [71].

4.4.2 Reheating

The preceding analysis was centered around the inflationary phase and the extraction of the corresponding slow-roll observables. A complete post-inflationary treatment, however, would require an additional discussion of the reheating phase, namely the stage during which the energy stored in the inflaton sector is transferred into relativistic degrees of freedom.

Inflation is usually taken to end when at least one of the slow-roll parameters becomes of order unity. Denoting by ρ_{end} the energy density of the inflaton sector at this time, the simplest estimate of the reheating temperature assumes instantaneous conversion of this energy into radiation. In that approximation, we may write,

$$\rho_{\mathrm{end}} \simeq \rho_{\mathrm{rad}} = \frac{\pi^2}{30} g_*(T_{\mathrm{re}}) T_{\mathrm{re}}^4, \tag{4.97}$$

where $g_*(T_{\mathrm{re}})$ is the effective number of relativistic degrees of freedom at the reheating temperature. This estimate is useful as an upper-limit intuition, but it does not describe the time-dependent energy transfer between the oscillating inflaton background and the produced radiation bath.

A more refined, effective description is adopted in [43], where the post-inflationary system is treated as a pair of coupled fluids. In this description, the inflaton component, with effective equation-of-state parameter w, loses energy to radiation through a constant dissipation rate Γ. The evolution is then modeled by equations of the form

$$\dot{\rho}_\phi + 3H(1+w)\rho_\phi = -\Gamma\rho_\phi, \tag{4.98}$$

$$\dot{\rho}_{\mathrm{rad}} + 4H\rho_{\mathrm{rad}} = \Gamma\rho_\phi, \tag{4.99}$$

$$3M_{\mathrm{Pl}}^2 H^2 = \rho_\phi + \rho_{\mathrm{rad}}. \tag{4.100}$$

The initial conditions at the onset of reheating are naturally chosen as

$$\rho_\phi(t_{\rm end}) = \rho_{\rm end}, \qquad \rho_{\rm rad}(t_{\rm end}) \simeq 0. \tag{4.101}$$

The reheating temperature is then identified with the radiation temperature at the end of reheating, conventionally defined by the crossing condition

$$\rho_{\rm rad}(t_{\rm re}) = \rho_\phi(t_{\rm re}). \tag{4.102}$$

At that time,

$$\rho_{\rm rad}(t_{\rm re}) = \frac{\pi^2}{30} g_*(T_{\rm re}) T_{\rm re}^4. \tag{4.103}$$

In this framework, for finite Γ, reheating lasts for a finite interval, and the value of $T_{\rm re}$ depends on both the energy density at the end of inflation and the effective decay rate. The microscopic value of Γ should then be derived from the interactions of the inflaton with the matter sector. The inflationary analysis considered in this chapter yields quantities such as the end point of inflation and the corresponding energy density $\rho_{\rm end}$. Thus, to embed the model into the effective reheating framework described above, we would still need to identify the relevant decay channels and compute the corresponding effective dissipation rate Γ. If, for instance, the inflaton sector couples to neutrino-like degrees of freedom, the required information would be contained in the Einstein-frame interaction terms involving the inflaton and those fields. Such couplings may arise after the conformal transformation and field redefinitions used to bring the gravitational sector into Einstein-frame form.

Chapter 5

Empirical Analysis: Weak Lensing and Stochastic Gravitational Wave Background

"Because of the ancient principle of WYGIWYGAINGW[1]"

— Terry Pratchett

How seldom we pause to consider the fortune of gazing from the vantage of a three-dimensional expanse! Yet to our chagrin, we find ourselves shackled by our technological abilities, wherein the nature of our observational evidence is such that the ensuing narrative must adhere inextricably to the contours of a flattened domain.

In fact, within the scope of the subsequent analysis, the empirical datasets suited to our purpose are confined to projected sky maps that are segregated into distinct redshift intervals. It is, however, important to stress that, for instance, spectroscopic surveys would provide detailed redshifts and therefore make it possible to create a three-dimensional map of the sky, although at the price of obtaining orders of magnitude fewer objects.

In the following sections, we will explore two distinct observables of gravitational models that are related by their tomographic configuration, i.e., the subdivision of the corresponding empirical data into redshift bins, allowing the spatial and temporal evolution to be reconstructed along the line of sight. The chapter is thus divided into two parts. The first section discusses weak lensing by photometric density ridges, where we aim to measure the tangential shear of background galaxies around these ridge structures, whose definitions will be provided shortly. The second section will then turn to the stochastic gravitational wave background. There, the relevant quantity to examine is

[1]If you insist: "What You Get Is What You're Given And It's No Good Whining."

the large-scale angular cross-correlation between stochastic gravitational wave intensity maps and tomographic galaxy samples.

5.1 Weak Lensing by Photometric Density Ridges

The weak-lensing formalism introduced in Chapter 3 provides a statistical method for probing the projected matter distribution through the coherent distortion of background galaxy images. Analyses are usually formulated in terms of two-point statistics, either in real space or in Fourier space. In the standard applications of this framework, the foreground lensing structures are either individual galaxies, galaxy clusters, or the full projected matter field reconstructed through cosmic shear. Cosmic shear correlates the shapes of source galaxies with one another, often across tomographic redshift bins, and thereby probes the projected matter distribution responsible for lensing along the line of sight. Galaxy–galaxy lensing, by contrast, uses an additional foreground sample of lens galaxies, typically chosen to trace massive dark-matter haloes, and measures the coherent tangential distortion of background sources around those lenses.

The work discussed in this section develops a different foreground object: density ridges identified directly from the photometric galaxy samples of the Dark Energy Survey Year 3 (DES-Y3) data [35]. These ridges are interpreted as two-dimensional projected tracers of the filamentary cosmic web and are used as lenses around which the tangential shear is measured. Later sections will examine how this approach relates to the previous formalism as an extension.

5.1.1 Introduction to the Problem

The cosmic web is one of the most direct manifestations of gravitational structure formation. On megaparsec scales, it defines an intricate, multiscale, interconnected network that is no longer smooth and homogeneous [73]. It arises from two central ingredients of the standard picture of structure formation both of which have been encountered previously in Chapter 3. The first is the statistical character of the primordial density field, which is usually modeled as a Gaussian random field and specified by its power spectrum [74]. The second is the subsequent gravitational evolution of these initial perturbations [75]. Through gravitational instability, initially overdense regions accrete matter, become denser, and contract relative to the background expansion, while underdense regions evacuate matter and expand into the large voids that occupy most of the cosmic volume. As the evolution leaves the linear regime, gravitational clustering generates increasingly intricate spatial features in the matter distribution, giving rise to the filamentary and node-like structure central to this study.

The question addressed in this work is whether photometrically identified density ridges can be used as foreground lenses in a weak-lensing measurement. A density ridge may be understood as the curve-like locus of locally enhanced projected galaxy density. More precisely, given a smoothed density field, a ridge is a set of points satisfying the following conditions [76]

1. the Hessian matrix of the surface has $D1$ negative eigenvalues.

2. the projection of the surface gradient onto the associated $D1$ eigenvectors is zero.

Ridges can be visually recognized quite straightforwardly, but robustly and efficiently locating them from discrete samples and/or noisy data can prove challenging. In this work, we use the subspace-constrained mean shift (SCMS) algorithm. Details of the algorithm, as well as the relevant parameter settings, may be found in [12]. If these structures contain excess mass relative to their surroundings, they should produce a coherent tangential shear pattern in the shapes of background galaxies. The observable of interest is thus the mean shear profile obtained by stacking many ridge segments. The tangential shear around a ridge point may be written schematically as

$$\gamma_+ = -\gamma_1 \cos(2\phi) - \gamma_2 \sin(2\phi), \tag{5.1}$$

where ϕ is the polar angle of the source position relative to the nearest local ridge point[2]. The corresponding cross-component is given by

$$\gamma_\times = \gamma_1 \sin(2\phi) - \gamma_2 \cos(2\phi). \tag{5.2}$$

As seen previously, this acts as a useful null diagnostic: a physical lensing signal should appear primarily in the tangential component, while the cross component should remain consistent with zero.

Additionally, we strive to investigate whether this signal contains cosmological information, in particular through its response to the amplitude of structure growth, commonly parametrized by

$$S_8 = \sigma_8 \left(\frac{\Omega_m}{0.3}\right)^{1/2}. \tag{5.3}$$

Here σ_8 measures the amplitude of matter fluctuations on scales of $8\,h^{-1}$Mpc, and, Ω_m denotes the present matter density fraction [77].

5.1.2 Methodology

Having performed the simulations as in [12], the analysis proceeds in three main stages: ridge identification, ridge selection and segmentation, and weak-lensing measurement.

[2]The discerning reader will observe in γ an antanaclasis, not to be confused with the one encountered in the previous chapter.

Ridge Points to Filaments

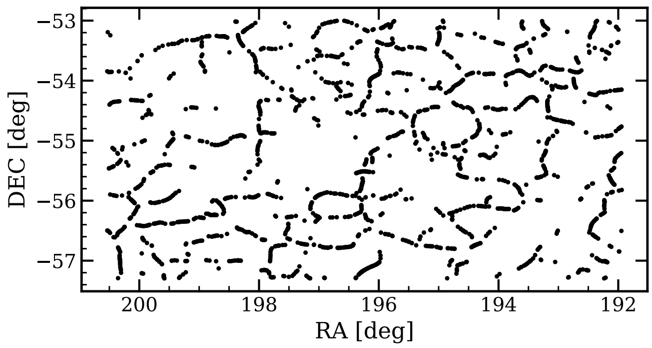
We aim to divide the ridge points identified through the SCMS method into distinct clusters that correspond to separate ridges. This task is quite ambiguous for us: there are several plausible approaches to separating ridges, although it is anticipated that the majority of sensible alternatives will produce a comparable lensing signal.

We achieve this through: 1. Forming a Minimum Spanning Tree (MST) [78] from the ridge point set by utilizing nearest-neighbour distances. This effectively links all the ridge points with connecting edges. 2. Detecting branch points in the MST as nodes with a degree greater than 2, and segmenting the tree at these points. This action separates the ridges at nodes where they converge. 3. Implementing the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [79] clustering algorithm in each MST segment to divide it into distinct ridges and to allocate labels to ridge points.

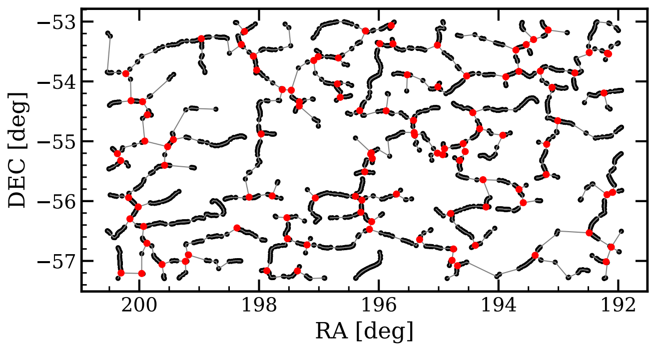
We start with a discrete collection of ridge coordinates that outline areas of increased projected density, aiming to extract geometrically coherent filament-like structures. In this context, the Minimum Spanning Tree (MST) represents the overall topological connectivity of the point distribution. The MST connects all points while minimizing the total length of edges and avoiding closed loops, thus offering a unique representation of the underlying filamentary framework. Each point initially determines its k-nearest neighbors through a KDTree search [80], creating a sparse weighted graph where edge weights reflect the pairwise spatial distances of connected neighbors.

Real cosmic filaments show junctions and bifurcations that create nodes linking multiple subfilaments. To isolate individual, physically coherent branches and minimize projection-induced blending in lensing studies, we identify branch points as nodes with degrees exceeding two and eliminate them from the MST. The resulting disjoint connected components depict separate filament segments. This is done because from a weak lensing standpoint, treating all branches as a single filament would amplify the projected overlap and artificially blend the shear signal from various directions; therefore, the segmentation step guarantees that each recovered structure yields a distinct, interpretable lensing signature.

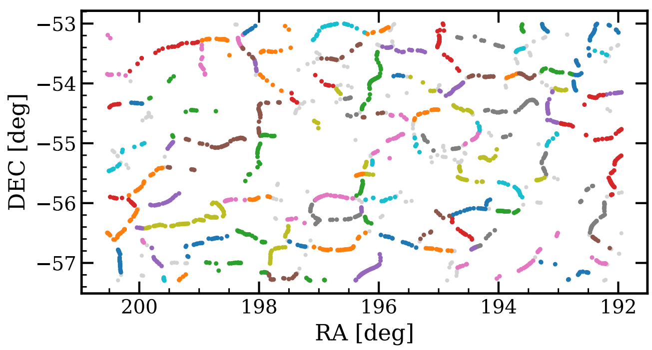
The next phase executes density-based segmentation on each MST component using the DBSCAN algorithm. It clusters points based on local spatial density without imposing any assumptions regarding cluster shape or topology. Within each MST segment, this method merges locally connected ridges into continuous filamentary chains while discarding erroneous or sparsely sampled points deemed noise. Figure 5.1 illustrates an overview of these stages of ridge segmentation and labeling.



(a) Ridge Points



(b) MST + Branch Construction.



(c) DBSCAN Filaments

Figure 5.1: Schematic overview of the ridge-to-filament identification pipeline.

Finally, the stacked tangential shear profile is computed by averaging the rotated source shears in logarithmic angular bins. Each source is paired with the nearest point on the corresponding filament segment, and the estimator in an angular bin centred at θ is

$$\hat{\gamma}_+(\theta) = \frac{\sum_{(r,s)\in\theta} w_s\, \gamma_{+,s}(r)}{\sum_{(r,s)\in\theta} w_s}, \tag{5.4}$$

where s labels source galaxies, r denotes the nearest ridge point associated with the source, w_s is the source weight, and $\gamma_{+,s}(r)$ is the tangential component of the source shear measured with respect to that ridge point. The cross component $\hat{\gamma}_\times(\theta)$ is computed analogously.

5.1.3 Data, Simulations, and Parameter Tests

Before applying the ridge-lensing estimator to the real DES data, the analysis is first tested on simulated DES-like catalogs, which allows the measurement pipeline to be studied in a controlled setting. The underlying cosmology, galaxy distribution, shear field, and redshift information are known. The simulations are generated using the GLASS package [81], which produces log-normal realizations of the matter density field from input matter power spectra. From these density fields, the corresponding convergence and shear fields are computed, and mock foreground lens and background source catalogs are generated with survey-like number densities and shape noise. In the DES-Y3 implementation, the foreground ridge catalog is constructed from the MagLim lens sample, while the shear signal is measured using the DES-Y3 source catalog, whose galaxy shapes are calibrated with the Metacalibration method [82].

As we have briefly mentioned, measuring distances is an astronomer's Gordian knot. The photometric redshifts, in contrast to those of spectra, have to be inferred from the fluxes measured in a sky image. This is challenging for several reasons. For instance, multiple galaxy types can have similar colors at different redshifts. More specifically, photometry in a small number of bands fails to pick out individual spectral lines that yield secure redshifts. Therefore, we have to content ourselves with extrapolating from the general shape of the spectrum. This issue can be further amplified by not having a good knowledge of the galaxy population and ultimately not being able to accurately estimate the redshift prior. Consequently, photometric surveys possess relatively large redshift uncertainties.

In the present study, this complication arises within the context of distinguishing between foreground lenses and background sources. Since lensing occurs only when the mass distribution lies between the observer and the source galaxy, configurations in which a putative ridge lies behind the source do not contribute to the signal and are instead considered noise. For this reason, we performed a redshift split where galaxies with redshift $z < 0.4$ are used as the foreground sample from which density ridges are identified, while galaxies with $z > 0.4$ are used as background sources for the shear measurement. However, in DES-like catalogs, we are faced with broad redshift distributions of galaxies, and therefore this approach entails a compromise; it reduces noise but degrades the lens structure. Figure 5.2 shows the redshift distribution of lenses and background sources

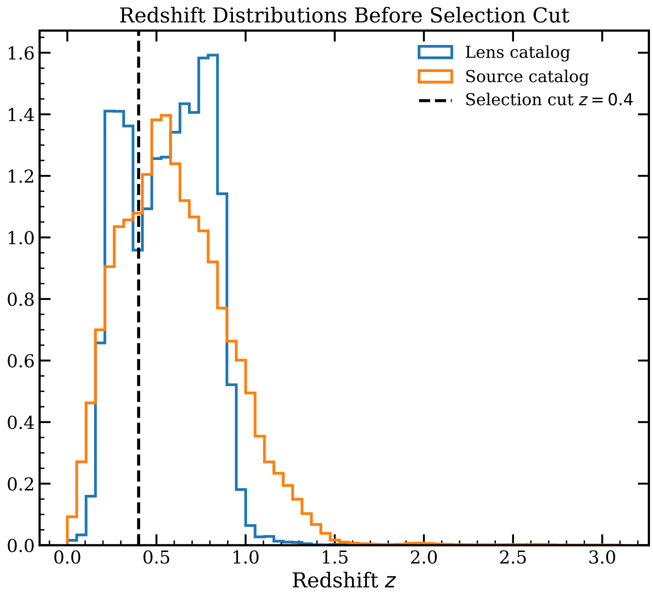


Figure 5.2: The input redshift distributions for the simulation foreground lenses (on whose density the ridges are defined) and background sources (whose orientation around the ridges we measure)

Having selected our lens and source samples, we begin our investigation on simulations containing no shape noise to test the convergence of algorithm parameters and examine the behavior of the underlying shear signal with respect to cosmology.

Several critical parameters bear significant physical relevance as they directly affect selection criteria pertaining to ridges. Key parameters in this regard include, in particular, the density threshold and the bandwidth.

Density threshold

After ridge points have been identified, the local projected galaxy density is evaluated at their positions using a kernel density estimate. Ridge points lying below a chosen density percentile are then removed. This procedure selects ridges associated with sufficiently overdense regions of the foreground galaxy field.

As the recurrent recalcitrance of empirical evidence demonstrates decisively, data manipulation constitutes an exercise in perpetual concessions. In essence, higher-density ridges are expected to trace more massive projected structures. They should therefore generate a stronger lensing signal that is less likely to be spurious features produced by sampling fluctuations. Increasing the threshold increases the purity of the ridge sample. However, it also reduces the number of retained ridge points, thereby reducing the number of lens–source pairs contributing to the signal.

In the noiseless simulations, increasing the density threshold systematically enhances the amplitude of the tangential shear profile, while preserving its broad morphology. The smaller-scale negative part of the signal is less sensitive to the threshold. We will repeat this particular analysis for the sample containing shape noise shortly, which will further clarify the behavior with respect to the density threshold.

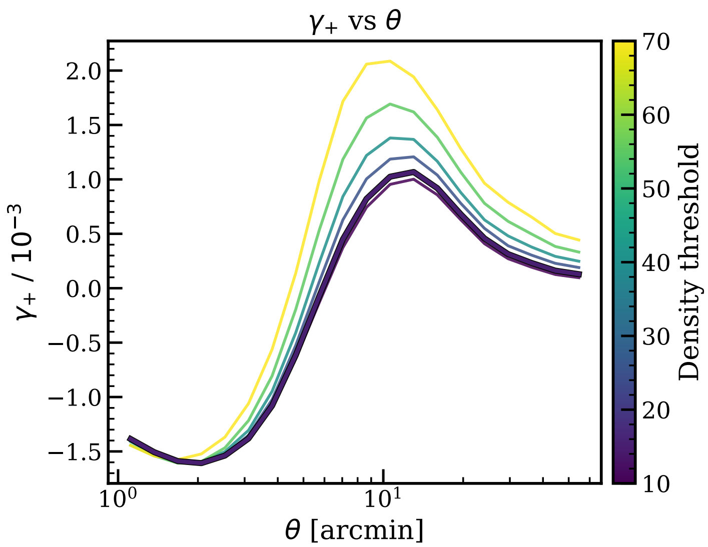


Figure 5.3: Variation of the signal-only tangential ridge shear signal with respect to a percentile used as density threshold; ridges with density below the threshold are excluded.

Bandwidth

The bandwidth is the most physically consequential SCMS parameter. It sets the smoothing scale used to construct the projected density field and enters the Gaussian kernel weighting used in the SCMS update. Thus, it defines the angular scale on which the galaxy distribution is interpreted as a continuous density field and determines whether the ridge-like structures are assumed to be physically meaningful.

If the bandwidth is selected to be too small, the reconstruction becomes sensitive to individual galaxies and their host haloes. In this limit, the distinction between ridge lensing and galaxy–galaxy lensing becomes weak, because the recovered structures are dominated by very short features around local overdensities. By contrast, if the bandwidth is too large, the reconstruction suppresses genuine smaller-scale filamentary structure and produces broader, more diffuse ridges. The bandwidth therefore controls the transition between a halo-dominated foreground selection and a more extended filamentary selection.

Practically, increasing the bandwidth suppresses the positive part of the signal, but has a more complicated behavior in the small-scale negative part. The location of the peak also shifts towards larger angular separations; this shift is expected as larger bandwidths effectively produce more diffuse ridges and, thus, the angular scale at which the signal is maximized moves outward.

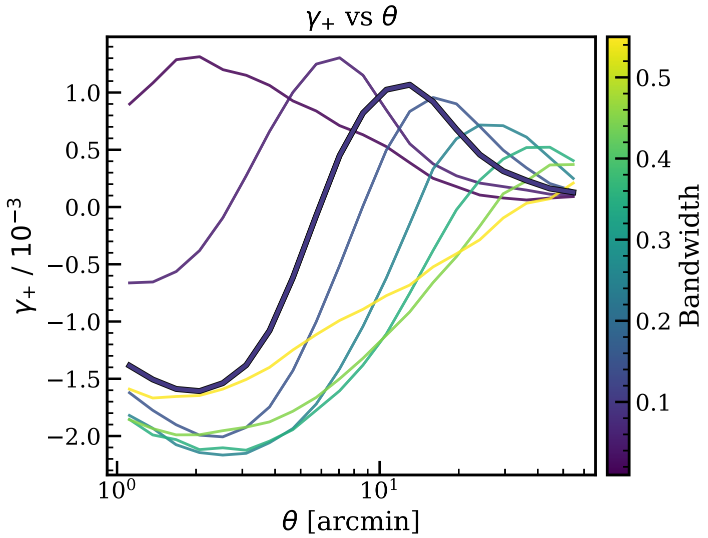


Figure 5.4: Variation of the signal-only tangential ridge shear signal with respect to the bandwidth (smoothing) parameter. For small bandwidths, the signal resembles the galaxy–galaxy lensing signal.[3]

5.1.4 Cosmological signal

In this section, we seek to identify the response of the ridge lensing observable to changes in cosmology. The choice of cosmological parameter directions was guided by the presentation of weak-lensing constraints in the (Ω_m, S_8) plane. In DES Year 3, the comparison between MagLim constraints, DES Year 1 redMaGiC constraints, and Planck 2018 shows the degeneracy direction between the matter density parameter and the clustering amplitude [83]. Thus, the fiducial cosmology is chosen to be

$$\Omega_{m,\text{fid}} = 0.3, \qquad S_{8,\text{fid}} = 0.8, \qquad \sigma_{8,\text{fid}} = 0.8. \tag{5.5}$$

Around this point, four one-dimensional sequences of cosmologies are generated with the middle point fixed to the fiducial model. These four sequences are designed to isolate different physical responses of the ridge-lensing signal.

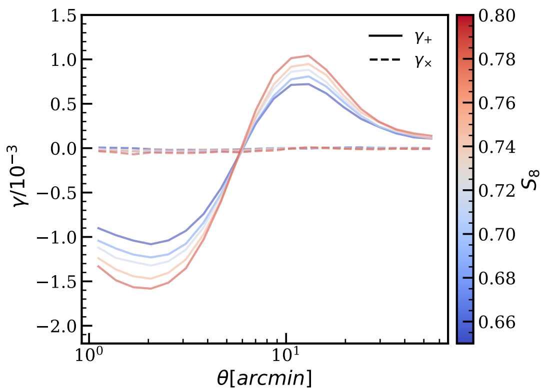
The first sequence varies σ_8 while keeping Ω_m fixed. Since σ_8 controls the normalization of matter fluctuations on $8\,h^{-1}$Mpc scales, this direction primarily tests whether ridge lensing responds to the amplitude of density fluctuations at fixed matter density. The second sequence varies Ω_m while keeping σ_8 fixed. This direction changes the matter density while holding fixed the fluctuation normalization. The third sequence varies the standard weak-lensing amplitude combination S_8. Along this trajectory, the cosmology moves in the direction to which weak-lensing observables are expected to be most sensitive. The final sequence performs the complementary test: S_8 is kept fixed while $S_{8,\perp}$

[3]This is demonstrated in [12].

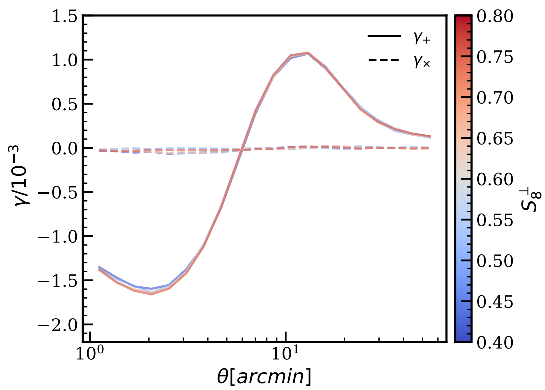
is varied. This direction is designed to test whether the ridge-lensing statistic contains information beyond the dominant weak-lensing amplitude combination. If the signal changes strongly in this direction, then the observable may be sensitive to cosmological information not captured by S_8 alone. Operationally, the two transformed coordinates used to construct these directions are

$$S_8 = \sigma_8 \left(\frac{\Omega_m}{0.3}\right)^{1/2}, \qquad S_{8,\perp} = \sigma_8 \left(\frac{\Omega_m}{0.3}\right)^{-2}. \tag{5.6}$$

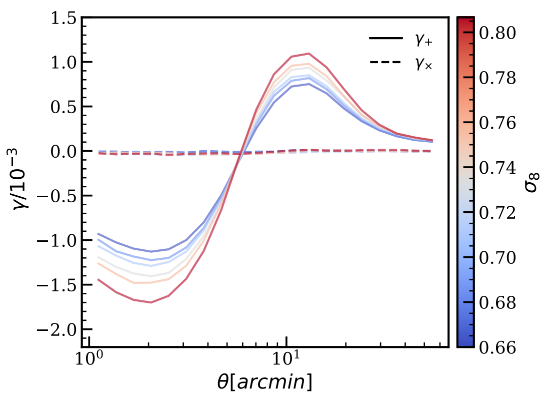
The S_8 response captured in Figure 5.5, appears primarily as an amplitude change rather than as a strong shift in the angular scale of the profile, and it happens to change the signal's profile significantly. Higher clustering amplitude produces more massive and more strongly lensing ridge environments. By contrast, the sequence varying $S_{8,\perp}$ at fixed S_8 produces a much weaker change in the profile. The two panels for Ω_m and fixed σ_8 show intermediate behavior between the two, consistent with the above interpretations.



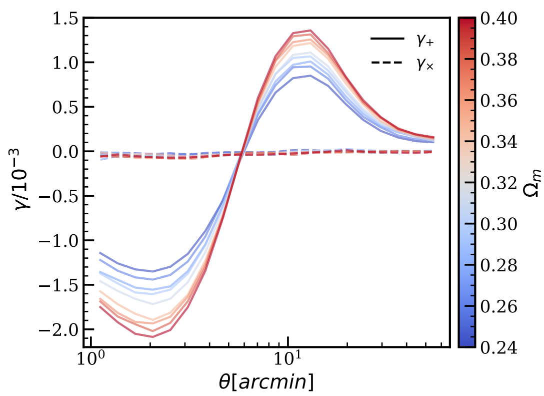
(a) Varying S_8.



(b) Varying $S_{8,\perp}$.



(c) Fixed Ω_m, varying σ_8.



(d) Fixed σ_8, varying Ω_m.

Figure 5.5: Behavior of shear profiles γ_+ and $\gamma_\times$ with respect to changing cosmological parameters. Top: S_8 and $S_{8,\perp}$. Bottom: fixed Ω_m and fixed σ_8.

5.1.5 Noisy simulations and DES-Y3 data

We may now move to the observationally relevant case, where the background source galaxy shapes include intrinsic ellipticity noise. In this case, shape noise is added to the source catalog at a level matched to the DES Year 3 Metacalibration sample [82].

We see that the lensing signal remains statistically significant after shape noise is included. Additionally, we perform the same shear measurement for the real DES-Y3 data. The results are broadly consistent in both amplitude and profile shape. In the fiducial noisy simulation, the tangential signal is detected with signal-to-noise $S/N \simeq 57$, while the DES-Y3 measurement gives a comparable detection at $S/N \simeq 47$. The cross component is consistent with zero, indicating that the measured signal is not dominated by an obvious parity-odd systematic.

A further diagnostic is provided by varying the density threshold in the noisy simulations. Increasing the threshold preferentially selects higher-density ridge points, which are expected to correspond to more massive projected structures and therefore to a stronger tangential shear signal. At the same time, a stricter threshold removes ridge points and reduces the number of available lens–source pairs. The detection significance increases from the lower threshold values and reaches its maximum around the 55th density percentile before declining for a still more restrictive cut.

Finally, the stability of the algorithm with respect to the random initialization of the SCMS mesh was tested by repeating the reconstruction with different random seeds. The resulting scatter in the shear profiles is small compared with the shape-noise covariance. Specifically, the typical fractional RMS is approximately 1%, with a maximum of 8.7% in the lowest signal-to-noise bin.

While our simple simulations are sufficient to recover the general signal strength, to be able to use them for precision parameter estimation we will need to make several improvements. In particular, we must apply more realistic treatment of redshift cuts to improve the simulations, and incorporate lens sample weights into the ridge estimation to account for varying depth across the field. With such improvements, we can consider emulation or simulation-based inference approaches to using ridge lensing for cosmological parameter estimates.

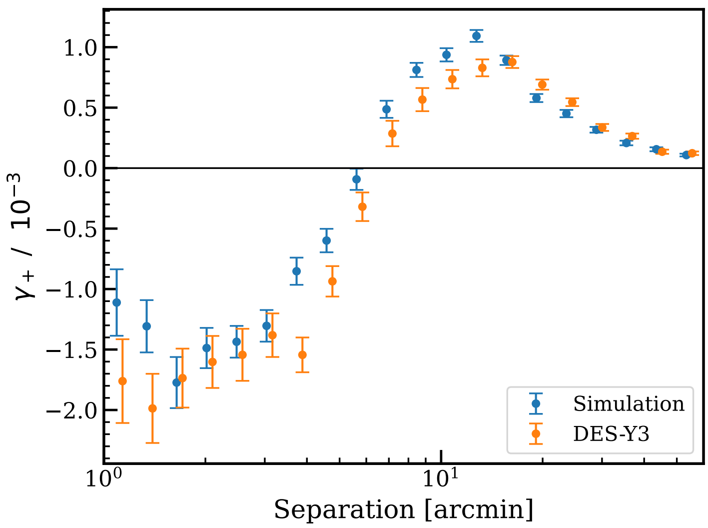


Figure 5.6: Ridge lensing tangential shear for Dark Energy Survey Year 3 data and a comparable noisy simulation with $\Omega_m = 0.3$ and $\sigma_8 = 0.8$

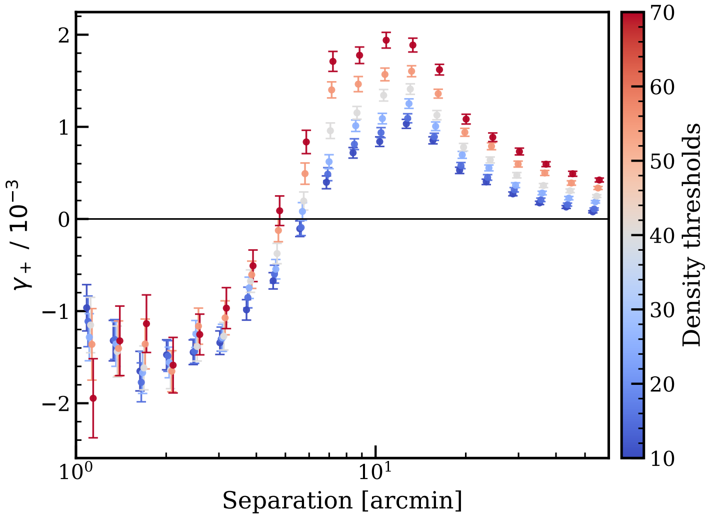


Figure 5.7: Ridge lensing tangential shear for increasing value of the threshold cut in noisy simulations

5.2 Tomographic Constraints on the Astrophysical Stochastic Gravitational Wave Background

In Chapter 3, we have briefly reviewed the gravitational wave energy content and some detection principles. The gravitational wave background may contain contributions from the early stages of the universe as well as many populations of sources not visible to standard astronomy [84]. In this section, we explore a different observable: the cross-correlation between maps of the stochastic gravitational wave background and tracers of the large-scale structure. Here, as well, the central idea is to use the matter distribution as a statistical template. In this case, the large-scale structure traces the astrophysical environments in which unresolved gravitational wave sources are expected to reside [13]. This makes cross-correlation with galaxy catalogs a clever way to search for an anisotropic astrophysical gravitational wave background. We carry out this analysis using tomographic data, which can be thought of as a projection of the Universe in time slices, i.e., we investigate different galaxy maps at various redshifts. To the best of our knowledge this provides the first upper-bound constraint on the production rate of gravitational waves as a function of cosmic time.

5.2.1 Astrophysical stochastic gravitational wave background

The Stochastic Gravitational Wave Background (SGWB) discussed here is the incoherent superposition of many unresolved gravitational wave sources generated by late-time astrophysical scenarios such as compact binary coalescences. Since these sources live in galaxies and trace the large-scale matter distribution, it would be reasonable to believe that their unresolved background might possess an anisotropic component correlated with galaxy overdensity. The SGWB intensity is quantified by the fractional gravitational wave energy density per logarithmic frequency interval and per unit solid angle and normalized to the critical density

$$\Omega_{\mathrm{GW}}(\hat{\mathbf{n}}, f) = \frac{1}{\rho_{c,0}} \frac{\mathrm{d}E_{\mathrm{GW}}}{\mathrm{d}\ln f \, \mathrm{d}V_0 \, \mathrm{d}\Omega_0}, \tag{5.7}$$

where $\rho_{c,0}$ is the critical density today, f is the observed frequency, and $\hat{\mathbf{n}}$ denotes a direction on the sky. For astrophysical sources distributed along the line of sight, the observed fluctuation in the SGWB can be written schematically as [13]

$$\Delta\Omega_{\mathrm{GW}}(\hat{\mathbf{n}}) = \int \mathrm{d}\chi \, q_{\mathrm{GW}}(\chi) \, \langle b\dot{\Omega}_{\mathrm{GW}}\rangle(z) \, \delta(\chi\hat{\mathbf{n}}, z), \tag{5.8}$$

where χ is comoving distance, δ is the matter overdensity, and $q_{\mathrm{GW}}(\chi)$ is the radial kernel that encodes the redshift weighting of the gravitational wave emissivity. This measurement will act as a constraint on the production rate weighted by the clustering bias of the sources written as

$$\langle b\dot{\Omega}_{\mathrm{GW}}\rangle(z). \tag{5.9}$$

Physically, this quantity measures how efficiently gravitational wave energy is produced by astrophysical sources at a given cosmic time, weighted by how strongly those sources trace the underlying matter field.

The fractional energy-density production rate may be related to the comoving gravitational wave emissivity [8] by

$$\dot{\Omega}_{\mathrm{GW}} = \frac{1}{\rho_{c,e}} \frac{\mathrm{d}E_e}{\mathrm{d}t_e \, \mathrm{d}\ln f_e \, \mathrm{d}V_e} = \frac{f_e}{\rho_{c,e}} \dot{j}_{\mathrm{GW}}, \tag{5.10}$$

where all quantities carrying the subscript e are evaluated in the source frame and $æ_{GW}$ represents gravitational wave emissivity. This expression makes explicit that $\dot{\Omega}_{\rm GW}$ is a fractional luminosity density which is the gravitational wave energy emitted per unit time, logarithmic frequency interval, and comoving volume, normalized to the critical density at emission.

5.2.2 Galaxy tomography and projected overdensity fields

The galaxy catalogs enter the analysis as tomographic tracers of the matter distribution. For a given galaxy sample, the projected overdensity may be written as

$$\Delta_g(\hat{\mathbf{n}}) = \int \mathrm{d}\chi\, q_g(\chi)\, b_g(z)\, \delta(\chi\hat{\mathbf{n}}, z), \tag{5.11}$$

where b_g is the galaxy bias and $q_g(\chi)$ is the radial selection kernel of the sample, where we have assumed a linear bias relation. A useful definition is

$$q_g(\chi) = H(z)\, p_g(z), \tag{5.12}$$

where $p_g(z)$ is the normalized redshift distribution of the catalog. This makes the galaxy field a weighted projection of the three-dimensional matter overdensity.

Each catalog probes a different redshift range such that a signal in a low-redshift galaxy sample would constrain the local gravitational wave production rate, while a signal in a higher-redshift sample would constrain the production rate at earlier cosmic times.

The four galaxy samples used in the analysis are 2MPZ, WI×SC, DECaLS, and Quaia [13]. They are treated as independent tomographic tracers of the large-scale structure. We use the galaxy auto-spectra to calibrate the effective galaxy bias and the galaxy–SGWB cross-spectra to constrain $\langle b\dot{\Omega}_{\rm GW}\rangle$.

5.2.3 Angular cross-power spectra

The central observable is the angular cross-power spectrum between the galaxy overdensity map and the SGWB intensity map,

$$C_\ell^{g\rm GW} = \frac{2}{\pi}\int \mathrm{d}k\, k^2 \int \mathrm{d}\chi_1\, \mathrm{d}\chi_2\, b_g\, \langle b\dot{\Omega}_{\rm GW}\rangle\, q_g(\chi_1)\, q_{\rm GW}(\chi_2)\, j_\ell(k\chi_1)\, j_\ell(k\chi_2)\, P(k; z_1, z_2), \tag{5.13}$$

where j_ℓ are spherical Bessel functions and $P(k; z_1, z_2)$ is the matter power spectrum approximated by

$$P(k; z_1, z_2) \simeq \left[P(k, z_1)P(k, z_2)\right]^{1/2}. \tag{5.14}$$

on the large scales relevant to the measurement. We note that this approximation assumes perfect correlation between the overdensity field at different redshifts.

For intuition, assuming a reasonably compact sample of galaxies, the galaxy kernel q_g is sharply peaked at the mean redshift of the sample, we may define the template spectra T_ℓ such that

$$C_\ell^{gg} = b_g^2 \, T_\ell^{gg}, \qquad C_\ell^{g\mathrm{GW}} = b_g \, \langle b\dot{\Omega}_{\mathrm{GW}} \rangle \, T_\ell^{g\mathrm{GW}}. \tag{5.15}$$

The templates T_ℓ^{gg} and $T_\ell^{g\mathrm{GW}}$ depend on the cosmological model. The data therefore constrain the galaxy bias b_g through the galaxy auto-spectrum and constrain the gravitational wave production amplitude through the cross-spectrum.

Furthermore, the SGWB map depends on the assumed frequency dependence of the background. The analysis adopts a power-law scaling with respect to the spectral index α at a reference frequency $f_{\mathrm{ref}} = 25\,\mathrm{Hz}$. Three values are then considered

$$\alpha = 0, \qquad \alpha = \frac{2}{3}, \qquad \alpha = 3. \tag{5.16}$$

The case $\alpha = 2/3$ is the standard expectation for an astrophysical background dominated by compact-binary inspirals. The case $\alpha = 0$ corresponds to the scale-invariant spectrum used for cosmological signals, while $\alpha = 3$ is included following previous Ligo–Virgo–KAGRA map analyses as a "best-fit" spectral index that traces the detector noise spectrum and yields the most sensitive result, or equivalently the most stringent upper limit.

5.2.4 Estimating the spectra

For the galaxy autocorrelation, the instrumental complications are relatively mild. At the angular scales of interest, the galaxy shot noise may be approximated as nearly white and homogeneous. The analysis uses bandpowers with width $\Delta\ell = 10$, restricting the measurement to scales for which the linear-bias description remains reliable.

The galaxy–SGWB cross-correlation requires more caution. The SGWB map noise is highly non-white and thus, a simple pseudo-C_ℓ estimator is sub-optimal. Instead, the analysis uses a quadratic minimum-variance estimator [85]. If the data vector is written schematically as

$$\mathbf{d} = \left(\mathbf{m}_{\mathrm{GW}}, \mathbf{m}_g\right), \tag{5.17}$$

where $\mathbf{m}_{\mathrm{GW}}$ and $\mathbf{m}_g$ are the SGWB and galaxy overdensity maps, the covariance takes the form

$$\mathbf{C} = \langle \mathbf{d}\,\mathbf{d}^T \rangle. \tag{5.18}$$

The parameters to be estimated are the relevant angular spectra,

$$\mathbf{p} = \left(C_\ell^{\mathrm{GWGW}}, C_\ell^{g\mathrm{GW}}, C_\ell^{gg}\right). \tag{5.19}$$

The quadratic estimator uses the covariance and its derivatives with respect to the power-spectrum parameters,

$$\mathbf{Q}_\ell = \frac{\partial \mathbf{C}}{\partial C_\ell}, \tag{5.20}$$

to form an inverse-variance-weighted estimate of the spectra.

As mentioned, measurement of $C_\ell^{g\mathrm{GW}}$ spectra tests whether directions on the sky with an excess of galaxies also show an excess of stochastic gravitational wave intensity, where a positive cross-correlation would indicate that part of the SGWB traces the large-scale structure, as expected for an astrophysical background sourced by unresolved compact binaries.

Regrettably, while the preceding body of work enjoys an unbroken symmetry of corroboration, this particular endeavor introduced a sobering dissonance, and the measured spectra yield a null result for all galaxy catalogs and for all three values of α. Indeed, harboring the anticipation of a detection would have been unwarranted as the current SGWB maps are still dominated by instrumental noise on the angular scales relevant for the analysis. Yet "sweet are the uses of adversity"! The null result does, in fact, prove informative. Once the measured spectra are consistent with zero, the same likelihood can be used to place upper limits on the amplitude parameter $\langle b\dot{\Omega}_{\mathrm{GW}}\rangle$.

5.2.5 Forecasts

Awaiting future progress in the field, we might ask how the same tomographic cross-correlation method will perform with upcoming gravitational wave detector networks. Since the galaxy maps are already much less noisy than the SGWB maps on the relevant scales, the limiting factor is the SGWB map noise. In the noise-dominated regime, the covariance of the cross-spectrum [86] is approximately controlled by

$$\mathrm{Cov}\left(C_\ell^{g\mathrm{GW}}, C_{\ell'}^{g\mathrm{GW}}\right) \simeq \delta_{\ell\ell'} \frac{N_\ell^{\mathrm{GW}} C_\ell^{gg}}{2\ell+1}, \tag{5.21}$$

where N_ℓ^{GW} is the SGWB map noise and C_ℓ^{gg} is the galaxy auto-spectrum. Consequently, improving the detector network reduces N_ℓ^{GW}, which directly tightens the uncertainty on $C_\ell^{g\mathrm{GW}}$ and hence on $\langle b\dot{\Omega}_{\mathrm{GW}}\rangle$.

We have considered future networks, including configurations with the Einstein Telescope and Cosmic Explorer. The strongest forecast case approaches the expected astrophysical signal level after a long observing time [13]. This does not guarantee a detection, because the true astrophysical amplitude and the residual systematics of future SGWB maps remain uncertain. It does, however, suggest that with third-generation detectors, galaxy–SGWB tomography may become a probe of the redshift evolution of gravitational wave source populations.

Conclusion

"These fragments I have shored against my ruins"

— T.S. Eliot, The Waste Land

The guiding thread unifying this work consisted of the attempt to understand gravity through several complementary descriptions, from a formal structure of gravitational dynamics to the observational probes of the late Universe.

The first core result considered was the Hamiltonian formulation of a class of parametrized gravitational theories that may be interpreted as spontaneously broken gauge theories of the complexified Lorentz group. In this framework, the gravitational field is described by a Lorentz-valued gauge connection together with a Higgs-like field whose covariant derivative reconstructs the tetrad field. The Hamiltonian analysis then provides the appropriate language in which to determine the dynamical content of the theory. Since the phase space is constrained, not all canonical variables correspond to physical degrees of freedom. The analysis shows the relative weighting of the self-dual and anti-self-dual sectors changes the constraint structure and therefore the number of propagating degrees of freedom. For a generic non-vanishing chiral asymmetry, the theory propagates three complex degrees of freedom, whereas in the symmetric case, the theory contains no local degrees of freedom. In the special chiral limits, an extension of General Relativity is recovered in a symmetry-broken regime.

Several directions naturally follow from these results. The first concerns the physical interpretation of the additional degree of freedom that appears in the general relativistic extension of the model. In the Hamiltonian analysis, this degree of freedom is associated with an effective gravitational contribution equivalent, at the level of the classical equations of motion, to a pressureless perfect fluid. This is noteworthy because dust-like degrees of freedom have previously been discussed in relation to the problem of time in quantum gravity [87]. The present framework therefore suggests a setting in which the relation between time, matter-like effective sources, and the canonical structure of gravity may be investigated further. A second direction concerns quantization. The chiral limits are closely related to Ashtekar-type formulations, where the Hamiltonian structure becomes polynomial but the imposition of appropriate reality conditions remains essential. Finally, the connection with chiral formulations of gravity motivates the study of tensor perturbations, in particular whether parity-sensitive primordial tensor spectra may arise in this class of models.

We then turned from the canonical structure of gravity to its cosmological implications. The left-right symmetric inflationary model studied here is motivated by the possibility that enlarged matter sectors and chiral gravitational couplings may leave imprints in the early Universe. In this setting, an extended Higgs sector permits non-minimal couplings to gravity, leading to effective inflationary dynamics after reduction to an FLRW background. The background analysis shows how the inflationary observables depend on the underlying model parameters and how their viable regions of parameter space may be identified through numerical exploration. Particular attention was given to the effective Barbero–Immirzi parameter, whose field dependence distinguishes this class of models from simpler single-field inflationary scenarios.

At the level of background dynamics, the model can reproduce viable inflationary behaviour in regions of parameter space compatible with observational constraints. The numerical analysis, therefore, establishes that the left-right symmetric structure can be embedded consistently into an inflationary setting. At the same time, the results imply that the principal future interest of the model may lie in the additional gravitational structure carried by the time-dependent chiral coupling. The tensor sector, in particular, upon quantization, offers the possibility of distinguishing this scenario from more conventional inflationary models, since parity-sensitive or chiral gravitational-wave signatures may arise when the effective Immirzi parameter varies on the cosmological background.

The future development of this project is therefore twofold. First, the inflationary analysis should be completed beyond the background level by deriving and solving the perturbation equations in a way that fully accounts for the first-order gravitational variables. This includes clarifying the tensor sector, the role of chiral polarization states, and the possible observational signatures in primordial gravitational waves. Second, the post-inflationary evolution must be treated with comparable care. Particularly, the relation between the end of inflation, the reheating temperature, and the subsequent radiation-dominated era affects the mapping between model parameters and observable scales.

The final part of the thesis departed from theoretical construction to empirical probes of gravitational phenomena. The weak-lensing project considered the lensing signal associated with photometric density ridges, which act as projected tracers of filamentary structures in the cosmic web. This work is motivated by the fact that much of the matter distribution in the Universe is organized into extended filaments connecting dense nodes and galaxy concentrations. Weak lensing provides a direct probe of the projected mass distribution, while density ridges provide a practical way to identify filament-like structures in photometric surveys.

The analysis demonstrates that ridge lensing can be detected in DES-Y3 data at high significance, though not independently of the galaxy–galaxy lensing contribution. This interpretation of the filament-lensing observable requires further methodological refinement. Simulations show that the signal depends on both cosmological and algorithmic choices, with the dominant cosmological sensitivity associated primarily with S_8.

The anticipated trajectory of this work consists of turning a detection-oriented method into a parameter-estimation tool. This requires a more careful separation of filamentary lensing from galaxy–galaxy lensing, improved modeling of projection effects, a better treatment of photometric redshift uncertainty, and a robust covariance framework. The method should also be tested on larger and deeper survey data, where the increased statistical power of surveys such as LSST and Euclid may allow ridge-based observables to become more informative. If these challenges are addressed, weak lensing by density ridges could offer a complementary route to studying the cosmic web.

The second empirical direction concerned the stochastic gravitational-wave background. Unlike individually resolved compact-binary events, the astrophysical SGWB is the accumulated contribution of many unresolved sources across cosmic history. Since these sources are expected to trace the large-scale structure, cross-correlating SGWB maps with galaxy surveys provides a way to search for anisotropies in the gravitational-wave background and to connect them with the distribution of matter. The tomographic approach is valuable because it allows the information to be separated into redshift ranges, thereby converting a projected cross-correlation into a constraint on gravitational-wave production.

The analysis considered in this work finds no statistically significant cross-correlation between the SGWB maps and the galaxy samples used. This null result provides upper bounds on the bias-weighted production rate density of gravitational waves by astrophysical sources as a function of cosmic time. The current constraints remain above the expected signal, but they establish a framework in which future gravitational-wave observing runs and improved galaxy maps may reach the sensitivity required for detection.

Parting words are never easy, for they must shoulder the cumulative *gravity* of an entire intellectual journey. But “we know what a masquerade all development is, and what effective shapes may be disguised in helpless embryos.—In fact, the world is full of hopeful analogies and handsome dubious eggs called possibilities.”[4]

[4]George Eliot, *Middlemarch*.

Appendix A

Extrinsic Curvature and ADM Constraints

Here we collect the geometric identities underlying the $3 + 1$ decomposition used in the main text. The purpose is to make explicit how the extrinsic geometry of a spatial hypersurface is related to the four-dimensional Einstein equations. The derivation also clarifies why the Hamiltonian and momentum constraints restrict initial data and how their preservation is tied to the contracted Bianchi identity.

Throughout this appendix, Greek indices $\mu, \nu, \rho, \ldots$ denote spacetime coordinate indices. Lowercase Latin indices $i, j, k, \ldots$ denote spatial coordinate indices on Σ_t. Abstract lowercase Latin indices $a, b, c, \ldots$ denote spacetime abstract indices. The spacetime metric has signature $(-+++)$. The future-directed unit normal to the hypersurface is denoted n^a, with

$$n^a n_a = -1. \tag{A.1}$$

The spatial metric, equivalently the projector onto Σ_t, is

$$h_{ab} = g_{ab} + n_a n_b. \tag{A.2}$$

It satisfies

$$h_{ab} n^b = 0, \qquad h_a{}^c h_c{}^b = h_a{}^b. \tag{A.3}$$

A.1 Foliation, Lapse, and Shift

Let the spacetime manifold be foliated by spacelike hypersurfaces Σ_t, where t is a coordinate parameter labelling the leaves. The coordinate time-flow vector decomposes into normal and tangential parts:

$$\left(\frac{\partial}{\partial t}\right)^a = N n^a + N^a, \tag{A.4}$$

where N is the lapse and N^a is the shift. The shift is tangent to the hypersurface:

$$n_a N^a = 0. \tag{A.5}$$

In adapted coordinates, $N^a = (0, N^i)$.

The lapse determines the proper time separation between neighbouring hypersurfaces along the normal direction:

$$d\tau = N\,dt. \tag{A.6}$$

The shift measures the tangential displacement of the spatial coordinate system from one hypersurface to the next. Thus, an observer advancing by coordinate time dt moves normally by Ndt and tangentially by $N^i dt$.

The ADM line element is

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt). \tag{A.7}$$

Hence

$$g_{\mu\nu} = \begin{pmatrix} -N^2 + h_{ij}N^i N^j & h_{ij}N^j \\ h_{ij}N^j & h_{ij} \end{pmatrix}, \tag{A.8}$$

and

$$g^{\mu\nu} = \begin{pmatrix} -\dfrac{1}{N^2} & \dfrac{N^i}{N^2} \\ \dfrac{N^i}{N^2} & h^{ij} - \dfrac{N^i N^j}{N^2} \end{pmatrix}. \tag{A.9}$$

The normal covector and vector are then

$$n_\mu = (-N, 0, 0, 0), \qquad n^\mu = \left(\frac{1}{N}, -\frac{N^i}{N}\right). \tag{A.10}$$

Where we may immediately verify that $n^\mu n_\mu = -1$.

A.2 Definition and Geometric Interpretation of the Extrinsic Curvature

The intrinsic metric h_{ij} describes distances measured within the hypersurface Σ_t. To extract the information regarding the embedding of the hypersurface, we need the extrinsic curvature.

$$K_{ab} = -h_a{}^c \nabla_c n_b. \tag{A.11}$$

Since the index a is projected with $h_a{}^c$, the first index of K_{ab} is tangent to the hypersurface.

$$K_{ab}n^b = -h_a{}^c n^b \nabla_c n_b = -\frac{1}{2} h_a{}^c \nabla_c (n^b n_b) = 0. \tag{A.12}$$

Thus, the second index also proves tangent and K_{ab} is purely spatial. This definition may be expanded using the projector Equation A.2

$$\begin{aligned} K_{ab} &= -({g_a}^c + n_a n^c)\nabla_c n_b \\ &= -\nabla_a n_b - n_a n^c \nabla_c n_b. \end{aligned} \tag{A.13}$$

Defining the acceleration of the normal vector by

$$a_b := n^c \nabla_c n_b, \tag{A.14}$$

we obtain

$$K_{ab} = -\nabla_a n_b - n_a a_b. \tag{A.15}$$

This identity makes explicit that K_{ab} is the hypersurface-projected derivative of the normal. Geometrically, it measures how the normal direction changes when parallel transported along the hypersurface.

A.3 Symmetry of the Extrinsic Curvature

Since the hypersurfaces are level sets of a scalar time function t, the normal covector is proportional to the gradient of t:

$$n_a = -N\nabla_a t. \tag{A.16}$$

Equivalently, one may write $n_a = -N\xi_a$, where $\xi_a = \nabla_a t$. Given the gradient of a scalar satisfies

$$\nabla_a \xi_b = \nabla_b \xi_a, \tag{A.17}$$

projecting both indices onto the hypersurface gives

$$\begin{aligned} K_{ab} &= -{h_a}^c {h_b}^d \nabla_c n_d \\ &= {h_a}^c {h_b}^d \nabla_c (N\xi_d) \\ &= {h_a}^c {h_b}^d \left[(\nabla_c N)\xi_d + N\nabla_c \xi_d \right]. \end{aligned} \tag{A.18}$$

The first term vanishes because ${h_b}^d \xi_d$ is proportional to ${h_b}^d n_d$. Thus

$$K_{ab} = N {h_a}^c {h_b}^d \nabla_c \xi_d. \tag{A.19}$$

Since $\nabla_c \xi_d = \nabla_d \xi_c$, it follows that

$$K_{ab} = K_{ba}. \tag{A.20}$$

A.4 Lie Derivative of the Spatial Metric

The relation between K_{ab} and the change of the spatial metric is obtained from the Lie derivative of h_{ab} along the normal vector:

$$\mathcal{L}_n h_{ab} = n^c \nabla_c h_{ab} + h_{ac} \nabla_b n^c + h_{bc} \nabla_a n^c. \tag{A.21}$$

Using $h_{ab} = g_{ab} + n_a n_b$ and metric compatibility $\nabla_c g_{ab} = 0$, we find

$$\begin{aligned} n^c \nabla_c h_{ab} &= n^c \nabla_c (n_a n_b) \\ &= (n^c \nabla_c n_a) n_b + (n^c \nabla_c n_b) n_a \\ &= a_a n_b + a_b n_a. \end{aligned} \tag{A.22}$$

Moreover,

$$\begin{aligned} h_{ac} \nabla_b n^c &= (g_{ac} + n_a n_c) \nabla_b n^c \\ &= \nabla_b n_a + n_a n_c \nabla_b n^c. \end{aligned} \tag{A.23}$$

The second term vanishes because

$$n_c \nabla_b n^c = \frac{1}{2} \nabla_b (n_c n^c) = 0. \tag{A.24}$$

Therefore

$$h_{ac} \nabla_b n^c = \nabla_b n_a, \qquad h_{bc} \nabla_a n^c = \nabla_a n_b. \tag{A.25}$$

Substitution into (A.21) gives

$$\mathcal{L}_n h_{ab} = a_a n_b + a_b n_a + \nabla_b n_a + \nabla_a n_b. \tag{A.26}$$

Using (A.15),

$$\nabla_a n_b + n_a a_b = -K_{ab}, \tag{A.27}$$

and the symmetry of K_{ab}, one obtains

$$\mathcal{L}_n h_{ab} = -2K_{ab}. \tag{A.28}$$

Thus the extrinsic curvature is the normal rate of change of the induced metric:

$$K_{ab} = -\frac{1}{2} \mathcal{L}_n h_{ab}. \tag{A.29}$$

Using the decomposition of the time-flow vector,

$$\partial_t^a = N n^a + N^a, \tag{A.30}$$

we have

$$\mathcal{L}_{\partial_t} h_{ij} = N \mathcal{L}_n h_{ij} + \mathcal{L}_{\vec{N}} h_{ij}. \tag{A.31}$$

Since h_{ij} is compatible with the spatial covariant derivative D_i,

$$\mathcal{L}_{\vec{N}} h_{ij} = D_i N_j + D_j N_i. \tag{A.32}$$

Therefore

$$\partial_t h_{ij} = -2NK_{ij} + D_i N_j + D_j N_i. \tag{A.33}$$

Equivalently,

$$K_{ij} = -\frac{1}{2N}\left(\partial_t h_{ij} - D_i N_j - D_j N_i\right). \tag{A.34}$$

Where we have used the identity

$$\mathcal{L}_n h_{ij} = \frac{1}{\phi}\mathcal{L}_{\phi n} h_{ij} \tag{A.35}$$

We have, therefore, shown the interpretation of K_{ij} as the velocity of the spatial metric corrected by the tangential drag generated by the shift.

A.5 Spatial Covariant Derivative

The spatial covariant derivative D_i, in the metric approach, is the Levi–Civita connection compatible with h_{ij}:

$$D_i h_{jk} = 0. \tag{A.36}$$

For a spacetime tensor $T^{b_1\cdots b_r}{}_{c_1\cdots c_s}$, its fully projected spatial covariant derivative is defined by projecting every free index and the derivative index:

$$D_a T^{b_1\cdots b_r}{}_{c_1\cdots c_s} = h_a{}^{a'} h^{b_1}{}_{d_1}\cdots h^{b_r}{}_{d_r} h_{c_1}{}^{e_1}\cdots h_{c_s}{}^{e_s} \nabla_{a'} T^{d_1\cdots d_r}{}_{e_1\cdots e_s}. \tag{A.37}$$

where this operation would be intrinsic to the hypersurface.

A.6 Gauss–Codazzi Equation

The Gauss–Codazzi equation relates the intrinsic curvature of Σ_t to the projection of the four-dimensional Riemann tensor. For vectors tangent to the hypersurface, the fully projected curvature satisfies

$$h_a{}^\alpha h_b{}^\beta h_c{}^\gamma h_d{}^\delta R_{\alpha\beta\gamma\delta} = {}^{(3)}R_{abcd} + K_{ac}K_{bd} - K_{ad}K_{bc}. \tag{A.38}$$

Here ${}^{(3)}R_{abcd}$ is the Riemann tensor associated with the spatial metric h_{ij}. Contracting (A.38) with $h^{ac}h^{bd}$, we obtain

$$h^{ac}h^{bd}R_{abcd} = {}^{(3)}R + K^2 - K_{ij}K^{ij}. \tag{A.39}$$

Using A.2, the left-hand side may be rewritten as

$$\begin{aligned} h^{ac}h^{bd}R_{abcd} &= (g^{ac} + n^a n^c)(g^{bd} + n^b n^d)R_{abcd} \\ &= R + 2n^a n^b R_{ab}. \end{aligned} \tag{A.40}$$

Since

$$G_{ab} = R_{ab} - \frac{1}{2}g_{ab}R, \tag{A.41}$$

and $n^a n^b g_{ab} = -1$, we find

$$2n^a n^b G_{ab} = R + 2n^a n^b R_{ab}. \tag{A.42}$$

Thus

$$2n^a n^b G_{ab} = {}^{(3)}R + K^2 - K_{ij}K^{ij}. \tag{A.43}$$

If the Einstein equations are written as

$$G_{ab} = 8\pi T_{ab}, \tag{A.44}$$

then, defining the energy density measured by normal observers as

$$\rho = n^a n^b T_{ab}, \tag{A.45}$$

equation (A.43) gives the Hamiltonian constraint:

$${}^{(3)}R + K^2 - K_{ij}K^{ij} = 16\pi\rho. \tag{A.46}$$

Notably, this equation contains no second time derivative of the spatial metric. It is a restriction on the data (h_{ij}, K_{ij}) on a single hypersurface.

A.7 Codazzi–Mainardi Equation

The Codazzi–Mainardi equation relates the mixed projection of the four-dimensional Riemann tensor to the spatial derivative of K_{ij}:

$$h_a{}^{\alpha} h_b{}^{\beta} h_c{}^{\gamma} n^{\delta} R_{\alpha\beta\gamma\delta} = D_b K_{ac} - D_a K_{bc}. \tag{A.47}$$

Contracting on the appropriate spatial indices gives

$$h^a{}_{\mu} n_{\nu} G^{\mu\nu} = D_j K^{ij} - D^i K. \tag{A.48}$$

With the matter momentum density

$$J^i = -h^i{}_{\mu} n_{\nu} T^{\mu\nu}, \tag{A.49}$$

the Einstein equations yield the momentum constraint

$$D_j (K^{ij} - h^{ij} K) = 8\pi J^i. \tag{A.50}$$

As in the Hamiltonian constraint, this equation restricts the initial data but it does not depend on a second time derivative. It states that the spatial variation of the extrinsic curvature must be compatible with the momentum density of matter.

A.8 Evolution Equations

The constraints account for four components of the Einstein equations. The remaining six components describe the evolution of K_{ij}. Let us now consider the following

$$n^a = \frac{1}{N} \left[\left(\frac{\partial}{\partial t} \right)^a - N^a \right]. \tag{A.51}$$

Hence the Lie derivative along the normal can be expressed as

$$\mathcal{L}_n K_{ij} = \frac{1}{N} \left(\mathcal{L}_{\partial_t} K_{ij} - \mathcal{L}_{\vec{N}} K_{ij} \right). \tag{A.52}$$

The spatial projection of the four-dimensional curvature gives

$$h_i{}^{\mu} h_j{}^{\nu} n^{\rho} n^{\sigma} R_{\rho\mu\sigma\nu} = \mathcal{L}_n K_{ij} + \frac{1}{N} D_i D_j N + K_{ik} K^k{}_j, \tag{A.53}$$

up to the sign convention chosen for K_{ij}. Equivalently, using the Gauss–Codazzi relation and the Einstein equations, this yields

$$\begin{aligned}\mathcal{L}_{\partial_t} K_{ij} - \mathcal{L}_{\vec{N}} K_{ij} = -D_i D_j N + N \left({}^{(3)}R_{ij} + K K_{ij} - 2K_{ik}K^k{}_j\right) \\ + 4\pi N \left[h_{ij}(S-\rho) - 2S_{ij}\right].\end{aligned} \tag{A.54}$$

Here

$$K = h^{ij}K_{ij}, \qquad \rho = n^\mu n^\nu T_{\mu\nu}, \tag{A.55}$$

and

$$S_{ij} = h_i{}^\mu h_j{}^\nu T_{\mu\nu}, \qquad S = h^{ij}S_{ij}. \tag{A.56}$$

Not to be confused with previous appearances of these quantity within this work. Since $\mathcal{L}_{\partial_t} K_{ij} = \partial_t K_{ij}$ in coordinates adapted to the foliation, this may also be written as

$$\begin{aligned}\partial_t K_{ij} = \mathcal{L}_{\vec{N}} K_{ij} - D_i D_j N + N \left({}^{(3)}R_{ij} + K K_{ij} - 2K_{ik}K^k{}_j\right) \\ + 4\pi N \left[h_{ij}(S-\rho) - 2S_{ij}\right].\end{aligned} \tag{A.57}$$

The Lie derivative along the shift is explicitly

$$\mathcal{L}_{\vec{N}} K_{ij} = N^k D_k K_{ij} + K_{ik} D_j N^k + K_{kj} D_i N^k. \tag{A.58}$$

Together, (A.33) and (A.57) form the ADM evolution system. The lapse N and shift N^i are not fixed by the constraints; they encode the freedom to choose how the hypersurfaces are stacked and how the spatial coordinates are transported between them.

The canonical momentum conjugate to h_{ij} is usually written as

$$\pi^{ij} = \sqrt{h}\left(K^{ij} - h^{ij}K\right). \tag{A.59}$$

In terms of this momentum, the Hamiltonian formulation expresses the dynamics as constrained evolution on phase space.

A.9 Constraint Preservation

The previous equations show that the Einstein equations divide into constraints and evolution equations. A consistent initial-value formulation requires the constraints to remain satisfied under time evolution. This property follows from the contracted Bianchi identity together with conservation of stress energy. We can denote

$$A_{ab} := G_{ab} - 8\pi T_{ab}. \tag{A.60}$$

The Einstein equations are equivalent to

$$A_{ab} = 0. \tag{A.61}$$

Therefore, the contracted Bianchi identity $\nabla^a G_{ab} = 0$, together with $\nabla^a T_{ab} = 0$, implies

$$\nabla^a A_{ab} = 0. \tag{A.62}$$

The tensor A_{ab} can then be decomposed into parts normal and tangential to the hypersurface as

$$A_{ab} = E_{ab} + n_a M_b + n_b M_a + n_a n_b H, \tag{A.63}$$

where

$$E_{ab} := h_a{}^c h_b{}^d A_{cd}, \tag{A.64}$$

$$M_a := -h_a{}^c n^d A_{cd}, \tag{A.65}$$

$$H := n^c n^d A_{cd}. \tag{A.66}$$

The equation $H = 0$ is the Hamiltonian constraint, $M_a = 0$ is the momentum constraint, and $E_{ab} = 0$ gives the spatially projected evolution equations.

If the evolution equations are imposed, then

$$E_{ab} = 0. \tag{A.67}$$

The question here is whether $H = 0$ and $M_a = 0$, imposed on one hypersurface, remain valid on subsequent hypersurfaces as we move forward in time. This is controlled by projecting (A.62) along the normal and tangent to the hypersurface. The normal projection is

$$n^b \nabla^a A_{ab} = 0. \tag{A.68}$$

Using (A.63), this equation gives the evolution of the Hamiltonian constraint in terms of the momentum constraint. In particular, when $E_{ab} = 0$, it has the schematic form

$$\mathcal{L}_n H = \text{spatial derivatives of } M_a + \text{terms proportional to } H \text{ and } M_a. \tag{A.69}$$

Therefore, if H and M_a vanish initially, the normal projection contains no source that would generate a violation of $H = 0$.

The spatial projection is

$$h^b{}_c \nabla^a A_{ab} = 0. \tag{A.70}$$

This equation governs the propagation of the momentum constraint. Once again, after imposing $E_{ab} = 0$, it has the schematic structure

$$\mathcal{L}_n M_c = \text{spatial derivatives of } H + \text{terms proportional to } H \text{ and } M_c. \tag{A.71}$$

Thus, the constraint is preserved by the evolution equations, provided it holds on the initial hypersurface. Equations (A.68) and (A.70) prove the fact that the constraints are not independent evolution equations; their consistency is tied to diffeomorphism invariance through the contracted Bianchi identity.

Appendix B

The Ensemble Average of Cosmological Density Fields

In Chapter 3, the statistical properties of perturbations are described through correlation functions. In particular, the two-point function introduced in Equation 3.93 contains brackets of the form $\langle \cdots \rangle$. These brackets signify an ensemble average. More precisely, if $f(\mathbf{x})$ is a stochastic field, then $\langle Q[f] \rangle$ denotes the expectation value of a functional $Q[f]$ with respect to a probability distribution over possible realizations of the field. Formally, we may write

$$\langle Q[f] \rangle = \int \mathcal{D}f \, Q[f] \, \mathcal{P}[f], \tag{B.1}$$

where $\mathcal{P}[f]$ is the probability functional assigning a statistical weight to each field configuration. In Fourier space, this may be viewed as an average over the Fourier coefficients $f_{\mathbf{k}}$. Thus, for a field expanded as

$$f(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{B.2}$$

the quantity

$$\langle f_{\mathbf{k}} f_{\mathbf{k}'} \rangle \tag{B.3}$$

measures the statistical correlation between the Fourier modes $\mathbf{k}$ and $\mathbf{k}'$. For a statistically homogeneous field, the two-point function in position space cannot depend on the absolute positions $\mathbf{x}$ and $\mathbf{y}$ separately. It can only depend on their separation:

$$\langle f(\mathbf{x}) f(\mathbf{y}) \rangle = \xi(\mathbf{x} - \mathbf{y}). \tag{B.4}$$

This expresses invariance under spatial translations. Indeed, after a common translation $\mathbf{x} \to \mathbf{x} + \boldsymbol{\alpha}$ and $\mathbf{y} \to \mathbf{y} + \boldsymbol{\alpha}$, homogeneity requires

$$\langle f(\mathbf{x} + \boldsymbol{\alpha}) f(\mathbf{y} + \boldsymbol{\alpha}) \rangle = \langle f(\mathbf{x}) f(\mathbf{y}) \rangle \,. \tag{B.5}$$

In Fourier space, the same condition implies that only opposite momenta can be correlated. Therefore, the two-point function must be proportional to a Dirac delta distribution enforcing momentum conservation:

$$\langle f_{\mathbf{k}} f_{\mathbf{k}'} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{k}') P_f(\mathbf{k}). \tag{B.6}$$

The appearance of $\delta^{(3)}(\mathbf{k} + \mathbf{k}')$ underlines statistical homogeneity in Fourier-space.

If the field is also statistically isotropic, then the correlation function cannot depend on the orientation of the separation vector, but only on its magnitude:

$$\xi(\mathbf{x} - \mathbf{y}) = \xi(|\mathbf{x} - \mathbf{y}|). \tag{B.7}$$

Equivalently, in Fourier space the power spectrum cannot depend on the direction of $\mathbf{k}$, but only on its magnitude:

$$P_f(\mathbf{k}) = P_f(k), \qquad k := |\mathbf{k}|. \tag{B.8}$$

Thus, for a statistically homogeneous and isotropic field, the two-point function takes the standard form

$$\langle f_{\mathbf{k}} f_{\mathbf{k}'} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{k}') P_f(k). \tag{B.9}$$

The power spectrum $P_f(k)$ therefore contains the variance of the field per Fourier mode. In position space, the corresponding two-point correlation function is obtained by Fourier transforming the power spectrum:

$$\xi(r) = \langle f(\mathbf{x}) f(\mathbf{y}) \rangle = \int \frac{d^3k}{(2\pi)^3} P_f(k) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}, \qquad r = |\mathbf{x} - \mathbf{y}|. \tag{B.10}$$

Using isotropy, the angular dependence can be integrated out, giving

$$\xi(r) = \frac{1}{2\pi^2} \int_0^\infty dk\, k^2 P_f(k) \frac{\sin(kr)}{kr}. \tag{B.11}$$

It is useful to introduce the dimensionless power spectrum

$$\Delta_f^2(k) := \frac{k^3}{2\pi^2} P_f(k), \tag{B.12}$$

so that the variance of the field can be written as an integral per logarithmic interval in k:

$$\left\langle f^2(\mathbf{x}) \right\rangle = \int_0^\infty \frac{dk}{k}\, \Delta_f^2(k). \tag{B.13}$$

A scale-invariant spectrum corresponds to an approximately constant $\Delta_f^2(k)$. Equivalently,

$$P_f(k) \propto \frac{1}{k^3}. \tag{B.14}$$

Thus, the brackets $\langle \cdots \rangle$ appearing in the two-point function represent an average over possible realizations of the stochastic field. Under the assumptions of statistical homogeneity and isotropy, this average is constrained to have a very specific form: in position space it depends only on the distance between two points, while in Fourier space it is diagonal in momentum and characterized by a single function $P_f(k)$.

Appendix C

Curvature Decompositions for Tensor Perturbations

In this appendix we collect several component decompositions used in the tensor perturbation calculation. The curvature of the spin connection is defined by

$$R^{IJ}{}_{\mu\nu} = 2\partial_{[\mu}\omega^{IJ}{}_{\nu]} + 2\omega^{I}{}_{K[\mu}\omega^{KJ}{}_{\nu]}. \tag{C.1}$$

For the $3+1$ split used in the perturbative calculation, the relevant components are

$$R^{0i}{}_{ta} = 2\partial_{[t}\omega^{0i}{}_{a]} + 2\omega^{0}{}_{j[t}\omega^{ji}{}_{a]}, \tag{C.2}$$

$$R^{0i}{}_{ab} = 2\partial_{[a}\omega^{0i}{}_{b]} + 2\omega^{0}{}_{j[a}\omega^{ji}{}_{b]}, \tag{C.3}$$

$$R^{ij}{}_{ta} = 2\partial_{[t}\omega^{ij}{}_{a]} + 2\omega^{i}{}_{k[t}\omega^{kj}{}_{a]} + 2\omega^{i}{}_{0[t}\omega^{0j}{}_{a]}, \tag{C.4}$$

$$R^{ij}{}_{ab} = 2\partial_{[a}\omega^{ij}{}_{b]} + 2\omega^{i}{}_{k[a}\omega^{kj}{}_{b]} + 2\omega^{i}{}_{0[a}\omega^{0j}{}_{b]}. \tag{C.5}$$

We may then define the following two contractions, which appear in the expansion of the action:

$$\mathcal{I}_{\epsilon} := \varepsilon^{\mu\nu\alpha\beta}\epsilon_{IJKL}e^{I}{}_{\mu}e^{J}{}_{\nu}R^{KL}{}_{\alpha\beta}, \tag{C.6}$$

$$\mathcal{I}_{\gamma} := \varepsilon^{\mu\nu\alpha\beta}\eta_{IK}\eta_{JL}e^{I}{}_{\mu}e^{J}{}_{\nu}R^{KL}{}_{\alpha\beta}, \tag{C.7}$$

where the decomposition of each term gives

$$\begin{aligned}\mathcal{I}_{\epsilon} = 2\varepsilon^{tabc}\epsilon_{0ijk}\Big(&e^{0}{}_{t}e^{i}{}_{a}R^{jk}{}_{bc} - e^{i}{}_{t}e^{0}{}_{a}R^{jk}{}_{bc} + 2e^{i}{}_{t}e^{j}{}_{a}R^{0k}{}_{bc}\\ &+ 2e^{0}{}_{a}e^{i}{}_{b}R^{jk}{}_{tc} + 2e^{i}{}_{a}e^{j}{}_{b}R^{0k}{}_{tc}\Big).\end{aligned} \tag{C.8}$$

$$\mathcal{I}_\gamma = 2\varepsilon^{tabc}\eta_{ij}\Big(- e^0{}_t e^i{}_a R^{0j}{}_{bc} + e^i{}_t e^0{}_a R^{0j}{}_{bc} + \eta_{kl} e^i{}_t e^k{}_a R^{jl}{}_{bc} \\ - 2e^0{}_a e^i{}_b R^{0j}{}_{tc} + \eta_{kl} e^i{}_a e^k{}_b R^{jl}{}_{tc}\Big). \tag{C.9}$$

C.1 Tensor Perturbation Ingredients

The background co-tetrad and spin connection are taken to be

$$\bar{e}^0 = N(t)\, dt, \tag{C.10}$$

$$\bar{e}^i = a(t)\delta^i_a\, dx^a, \tag{C.11}$$

$$\bar{\omega}^{0I} = B(t)\, \bar{e}^I, \tag{C.12}$$

$$\bar{\omega}^{IJ} = -C(t)\epsilon^{IJK}\bar{e}_K. \tag{C.13}$$

The corresponding components are therefore

$$\bar{e}^0{}_t = N, \qquad \bar{e}^0{}_a = 0, \tag{C.14}$$

$$\bar{e}^i{}_t = 0, \qquad \bar{e}^i{}_a = a\delta^i_a. \tag{C.15}$$

Then we can write the most general ansatz for perturbation as

$$\delta e^0 = \Psi\bar{e}^0 + L_I\bar{e}^I, \tag{C.16}$$

$$\delta e^I = N^I\bar{e}^0 + S^{IJ}\bar{e}_J, \tag{C.17}$$

$$\delta\omega^{0I} = \gamma^I\bar{e}^0 + \xi\bar{e}^I + \epsilon^{IJK}\mu_J\bar{e}_K + \tau^{IJ}\bar{e}_J, \tag{C.18}$$

$$\delta\omega^{IJ} = \epsilon^{IJ}{}_K\phi^K\bar{e}^0 + \zeta\epsilon^{IJ}{}_K\bar{e}^K + \nu^{[I}\bar{e}^{J]} + \epsilon^{IJ}{}_K\chi^{KL}\bar{e}_L. \tag{C.19}$$

where from the tetrad perturbations we obtains

$$\delta e^0{}_t = N\Psi, \tag{C.20}$$

$$\delta e^0{}_a = aL_a, \tag{C.21}$$

$$\delta e^i{}_t = NN^i, \tag{C.22}$$

$$\delta e^i{}_a = aS^i{}_a. \tag{C.23}$$

For tensor modes, we shall retain only the tensor perturbations of the tetrad and spin connection:

- $\delta e^i = S^{ij}\bar{e}_j$,
- $\delta\omega^{0i} = \tau^{ij}\bar{e}_j$,
- $\delta\omega^{ij} = \epsilon^{ij}{}_k\chi^{kl}\bar{e}_l$,

- $\delta\omega^{ij}{}_a = a\epsilon^{ij}{}_k\chi^k{}_a$.

Thus, in the tensor sector the relevant perturbative variables are

$$S^{ij}, \qquad \tau^{ij}, \qquad \chi^{ij}. \tag{C.24}$$

The curvature components required for the tensor perturbation calculation are then collected in Table C.1. Finally, the background contributions for each term of the action is given by

$$\mathcal{I}_{\epsilon BG} = 24Na^3B^2 - 24Na^3C^2 + 24a^3\dot{B} + 12a^2\dot{a}B, \tag{C.25}$$

$$\mathcal{I}_{\gamma BG} = -24Na^3B^2 - 12a^3\left(\dot{C} + C\frac{\dot{a}}{a}\right), \tag{C.26}$$

Table C.1: Background curvature components used in the tensor perturbation calculation.

Curvature component	**Expression**
$\bar{R}^{jk}{}_{bc}$	$a^2B^2(\delta^j_b\delta^k_c - \delta^j_c\delta^k_b) + a^2C^2(\delta^j_c\delta^k_b - \delta^j_b\delta^k_c)$
$\delta R^{jk}{}_{bc}$	$a\epsilon^{jk}{}_m(\partial_b\chi^m{}_c - \partial_c\chi^m{}_b) + a^2B(\tau^j{}_b\delta^k_c + \delta^j_b\tau^k{}_c - \tau^j{}_c\delta^k_b - \delta^j_c\tau^k{}_b) + a^2C(\delta^j_b\chi^k{}_c + \delta^k_b\chi^j{}_c - \delta^j_c\chi^k{}_b - \delta^k_c\chi^j{}_b)$
$\delta^2 R^{jk}{}_{bc}$	$a^2(\tau^j{}_c\tau^k{}_b - \tau^j{}_b\tau^k{}_c) + a^2(\chi^k{}_b\chi^j{}_c - \chi^k{}_c\chi^j{}_b)$
$\bar{R}^{0k}{}_{bc}$	$2a^2BC\epsilon^k{}_{bc}$
$\delta R^{0j}{}_{bc}$	$a(\partial_b\tau^j{}_c - \partial_c\tau^j{}_b) + a^2B(\epsilon^j{}_{cn}\chi^n{}_b - \epsilon^j{}_{bn}\chi^n{}_c) + a^2C(\tau_{mc}\epsilon^{mj}{}_b - \tau_{mb}\epsilon^{mj}{}_c)$
$\delta^2 R^{0j}{}_{bc}$	$a^2\tau^m{}_b\epsilon^{mj}{}_l\chi^l{}_c - a^2\tau^m{}_c\epsilon^{mj}{}_l\chi^l{}_b$
$\bar{R}^{jk}{}_{tc}$	$-a\epsilon^{jk}{}_c\left(\dot{C} + C\frac{\dot{a}}{a}\right)$
$\delta R^{jk}{}_{tc}$	$-a\epsilon^{jk}{}_l\left(\dot{\chi}^l{}_c + \chi^l{}_c\frac{\dot{a}}{a}\right)$
$\bar{R}^{0k}{}_{tc}$	$a\left(\dot{B} + B\frac{\dot{a}}{a}\right)\delta^k_c$
$\delta R^{0k}{}_{tc}$	$a\dot{\tau}^k{}_c + \dot{a}\tau^k{}_c$
$\delta^2 R^{0k}{}_{tc}$	0
$\bar{R}^{jl}{}_{tc}$	$-a\epsilon^{jl}{}_c\left(\dot{C} + C\frac{\dot{a}}{a}\right)$
$\delta R^{jl}{}_{tc}$	$a\epsilon^{jl}{}_k\left(\dot{\chi}^k{}_c + \chi^k{}_c\frac{\dot{a}}{a}\right)$
$\delta^2 R^{jl}{}_{tc}$	0

Appendix D

Scalar Perturbations and the Mukhanov–Sasaki Equation

This appendix derives the quadratic action for scalar perturbations in single-field inflation and shows how it leads to the Mukhanov–Sasaki equation in support of the discussion of scalar perturbations and the scalar power spectrum in Chapter 3. Throughout this appendix, we work in units $M_{\mathrm{Pl}}^2 = 1$

D.1 Scalar Field Coupled to Gravity

We begin from the Einstein–Hilbert action minimally coupled to a scalar field,

$$S = \frac{1}{2}\int d^4x\sqrt{-g}\left[R - g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - 2V(\phi)\right]. \tag{D.1}$$

The scalar field is decomposed into a homogeneous background and a perturbation,

$$\phi(t,\mathbf{x}) = \bar{\phi}(t) + \delta\phi(t,\mathbf{x}). \tag{D.2}$$

The goal is then to identify the true propagating scalar degree of freedom after the non-dynamical metric variables have been eliminated.

D.2 ADM Decomposition

The ADM decomposition writes the spacetime line element as

$$ds^2 = -N^2dt^2 + g_{ij}\left(dx^i + N^i dt\right)\left(dx^j + N^j dt\right), \tag{D.3}$$

where N is the lapse, N^i is the shift vector, and g_{ij} is the induced spatial metric on constant-time hypersurfaces. The inverse metric components are

$$g^{00} = -\frac{1}{N^2}, \qquad g^{0i} = \frac{N^i}{N^2}, \qquad g^{ij} = g^{ij} - \frac{N^i N^j}{N^2}, \tag{D.4}$$

where g^{ij} is the inverse of the spatial metric.

The extrinsic curvature is

$$K_{ij} = \frac{1}{2N}\left(\dot{g}_{ij} - \nabla_i N_j - \nabla_j N_i\right), \tag{D.5}$$

where ∇_i is the covariant derivative compatible with g_{ij}. It is convenient to define

$$E_{ij} := N K_{ij} = \frac{1}{2}\left(\dot{g}_{ij} - \nabla_i N_j - \nabla_j N_i\right), \qquad E = g^{ij} E_{ij}. \tag{D.6}$$

The quantity E_{ij} is the lapse-rescaled extrinsic curvature and measures how the spatial metric changes along the normal direction to the hypersurface.

Using the ADM decomposition, the four-dimensional Ricci scalar decomposes as

$$R = R^{(3)} + K_{ij}K^{ij} - K^2 + \text{boundary terms}, \tag{D.7}$$

where $R^{(3)}$ is the Ricci scalar of the spatial metric g_{ij}. Next we isolate how the scalar-field kinetic term is decomposed. The scalar part of the action contains

$$g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi. \tag{D.8}$$

In the ADM decomposition, the lapse and shift separate the time derivative of the scalar field into a part normal to the spatial hypersurface and a part tangent to it. Explicitly, we find

$$g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi = -N^{-2}\left(\dot{\phi} - N^i\partial_i\phi\right)^2 + g^{ij}\partial_i\phi\,\partial_j\phi. \tag{D.9}$$

The combination

$$\dot{\phi} - N^i\partial_i\phi$$

is the change of the scalar field along the direction normal to the spatial slice. The shift contribution subtracts the apparent time variation due only to motion along the spatial coordinates on the hypersurface. Consider the ADM line element in matrix form,

$$g^{\mu\nu} = \begin{pmatrix} -N^{-2} & N^j/N^2 \\ N^i/N^2 & g^{ij} - N^iN^j/N^2 \end{pmatrix}. \tag{D.10}$$

Substituting these components into the kinetic contraction demonstrates the result found above

$$
\begin{aligned}
g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi &= g^{00}\dot{\phi}^2 + 2g^{0i}\dot{\phi}\,\partial_i\phi + g^{ij}_{(4)}\partial_i\phi\,\partial_j\phi \\
&= -\frac{1}{N^2}\dot{\phi}^2 + \frac{2N^i}{N^2}\dot{\phi}\,\partial_i\phi + \left(g^{ij} - \frac{N^iN^j}{N^2}\right)\partial_i\phi\,\partial_j\phi \\
&= -\frac{1}{N^2}\left(\dot{\phi} - N^i\partial_i\phi\right)^2 + g^{ij}\partial_i\phi\,\partial_j\phi.
\end{aligned} \tag{D.11}
$$

Putting everything together and dropping all boundary terms, the action becomes

$$
\begin{aligned}
S = \frac{1}{2}\int dt\,d^3x\,\sqrt{g}\Big[&NR^{(3)} + N^{-1}\left(E_{ij}E^{ij} - E^2\right) \\
&+ N^{-1}\left(\dot{\phi} - N^i\partial_i\phi\right)^2 - Ng^{ij}\partial_i\phi\partial_j\phi - 2NV(\phi)\Big].
\end{aligned} \tag{D.12}
$$

The ADM variables N and N^i appear without time derivatives. They are therefore not propagating variables. Their equations of motion are constraints, and solving these constraints is the key step in isolating the scalar degree of freedom.

D.3 Gauge Choice

For scalar perturbations it is convenient to work in comoving gauge,

$$
\delta\phi = 0. \tag{D.13}
$$

In this gauge, the scalar field is unperturbed, and the scalar fluctuation is transferred to the spatial metric. We parametrize the spatial metric as

$$
g_{ij} = a^2(t)e^{2\mathcal{R}(t,\mathbf{x})}\delta_{ij}. \tag{D.14}
$$

Here we have set $C = 0$. This is because the term $\partial_i\partial_jC$, which is responsible for the scalar shear deformation of the spatial metric, can be removed by a scalar spatial gauge choice. We should note that in this appendix, we follow the common ADM convention in which the scalar curvature perturbation is introduced through $g_{ij} = a^2e^{2R}\delta_{ij}$. This differs by an overall sign from the metric convention used in Chapter 3. This sign convention does not affect the quadratic action, the Mukhanov–Sasaki equation, or the scalar power spectrum, derived below, which are invariant under $R \to -R$.

Equivalently, to first order,

$$
g_{ij} = a^2(t)\left(1 + 2\mathcal{R}\right)\delta_{ij} + \mathcal{O}(\mathcal{R}^2). \tag{D.15}
$$

The variable $\mathcal{R}$ is the comoving curvature perturbation. We also decompose the lapse and shift as

$$N = 1 + \alpha, \qquad N_i = \partial_i \sigma, \tag{D.16}$$

where α and σ are first-order scalar perturbations. The vector part of the shift is omitted because it does not contribute to scalar perturbations at linear order.

D.4 Constraint Equations

Varying the ADM action with respect to N^i gives the momentum constraint,

$$\nabla_i \left[N^{-1} \left(E^i{}_j - \delta^i_j E \right) \right] = N^{-1} \left(\dot{\phi} - N^i \partial_i \phi \right) \partial_j \phi. \tag{D.17}$$

In comoving gauge $\partial_i \phi = 0$, so the right-hand side vanishes:

$$\nabla_i \left[N^{-1} \left(E^i{}_j - \delta^i_j E \right) \right] = 0. \tag{D.18}$$

Varying with respect to N gives the Hamiltonian constraint,

$$R^{(3)} - N^{-2} \left(E_{ij} E^{ij} - E^2 \right) - N^{-2} \left(\dot{\phi} - N^i \partial_i \phi \right)^2 - g^{ij} \partial_i \phi \partial_j \phi - 2V = 0. \tag{D.19}$$

In comoving gauge this simplifies to

$$R^{(3)} - N^{-2} \left(E_{ij} E^{ij} - E^2 \right) - N^{-2} \dot{\bar{\phi}}^2 - 2V(\bar{\phi}) = 0. \tag{D.20}$$

D.5 Background Equations

At zeroth order, the spatial metric is

$$g_{ij} = a^2(t) \delta_{ij}, \qquad N = 1, \qquad N^i = 0. \tag{D.21}$$

The spatial curvature vanishes,

$$R^{(3)} = 0. \tag{D.22}$$

Setting $N_i = 0$, we compute E_{ij} as follows

$$E_{ij} = \frac{1}{2} \dot{g}_{ij} = \frac{1}{2} \frac{d}{dt} (a^2(t) \delta_{ij}) = a \dot{a} \delta_{ij} \tag{D.23}$$

and we may reformulate to get

$$E_{ij} = a^2 H \delta_{ij}, \qquad E^i{}_j = H \delta^i_j, \qquad E = 3H. \tag{D.24}$$

Therefore,

$$E_{ij}E^{ij} - E^2 = 3H^2 - 9H^2 = -6H^2. \tag{D.25}$$

The Hamiltonian constraint then gives

$$6H^2 - \dot{\bar{\phi}}^2 - 2V(\bar{\phi}) = 0, \tag{D.26}$$

or equivalently,

$$3H^2 = \frac{1}{2}\dot{\bar{\phi}}^2 + V(\bar{\phi}). \tag{D.27}$$

which is the standard Friedmann equation for a homogeneous scalar field. The scalar-field equation is then given by

$$\ddot{\bar{\phi}} + 3H\dot{\bar{\phi}} + V_{,\phi} = 0. \tag{D.28}$$

Together, Eqs. (D.27) and (D.28) will be used repeatedly to simplify the perturbative action.

D.6 Solving The Constraints

We now expand the momentum constraint to first order. Using

$$N = 1 + \alpha, \qquad N^{-1} = 1 - \alpha + \mathcal{O}(\alpha^2), \qquad N_i = \partial_i\sigma + \hat{N}_i \tag{D.29}$$

where $\partial_i\hat{N}_i = 0$. The spatial metric is

$$g_{ij} = a^2 e^{2\mathcal{R}}\delta_{ij}, \tag{D.30}$$

We find, to first order,

$$E_{ij} = a^2\left[\left(H + \dot{\mathcal{R}}\right)\delta_{ij} - \frac{1}{a^2}\partial_i\partial_j\sigma\right]. \tag{D.31}$$

Raising one index gives

$$E^i{}_j = \left(H + \dot{\mathcal{R}}\right)\delta^i_j - \frac{1}{a^2}\partial^i\partial_j\sigma + \mathcal{O}(2), \tag{D.32}$$

and

$$E = 3\left(H + \dot{\mathcal{R}}\right) - \frac{1}{a^2}\partial^2\sigma + \mathcal{O}(2). \tag{D.33}$$

The combination appearing in the momentum constraint is

$$E^i{}_j - \delta^i_j E = -2H\delta^i_j - 2\dot{\mathcal{R}}\delta^i_j - \frac{1}{a^2}\partial^i\partial_j\sigma + \frac{1}{a^2}\delta^i_j\partial^2\sigma + \mathcal{O}(2). \tag{D.34}$$

Multiplication by $N^{-1} = 1 - \alpha$ gives

$$N^{-1}\left(E^i{}_j - \delta^i_j E\right) = -2H\delta^i_j - 2\dot{\mathcal{R}}\delta^i_j + 2H\alpha\delta^i_j$$
$$- \frac{1}{a^2}\partial^i\partial_j\sigma + \frac{1}{a^2}\delta^i_j\partial^2\sigma + \mathcal{O}(2). \tag{D.35}$$

Taking the spatial divergence, the two shift terms cancel:

$$\partial_i\left(-\frac{1}{a^2}\partial^i\partial_j\sigma + \frac{1}{a^2}\delta^i_j\partial^2\sigma\right) = 0. \tag{D.36}$$

Therefore the first-order momentum constraint reduces to

$$\partial_j\left(H\alpha - \dot{\mathcal{R}}\right) = 0. \tag{D.37}$$

The spatially homogeneous part can be absorbed into the background solution, and we take

$$\alpha = \frac{\dot{\mathcal{R}}}{H}. \tag{D.38}$$

Thus, the lapse perturbation is fixed algebraically by the time derivative of the curvature perturbation. To obtain the scalar part of the shift, we must use the Hamiltonian constraint. Expanding the Hamiltonian constraint to first order and inserting the solution to the lapse perturbation gives

$$\epsilon\dot{\mathcal{R}} - \frac{1}{a^2}\partial^2\sigma - \frac{1}{a^2}\frac{\partial^2\mathcal{R}}{H} = 0. \tag{D.39}$$

where taking the ∇_j would provide only the scalar part. Also, we recall that in terms of the slow-roll parameter, a canonical single-field background becomes

$$\epsilon = -\frac{\dot{H}}{H^2} = \frac{\dot{\bar{\phi}}^2}{2H^2}, \tag{D.40}$$

where $M_{pl} = 1$. The equation above can also be written as

$$\partial^2\sigma = -\frac{1}{H}\partial^2\mathcal{R} + a^2\epsilon\dot{\mathcal{R}}. \tag{D.41}$$

Using the background relation

$$\dot{\bar{\phi}}^2 = 2\epsilon H^2, \tag{D.42}$$

the shift may be written in the form

$$\sigma = -\frac{\mathcal{R}}{H} + \chi, \qquad \partial^2\chi = a^2\epsilon\dot{\mathcal{R}}. \tag{D.43}$$

Thus,

$$N_i = \partial_i \sigma = \partial_i \left(-\frac{\mathcal{R}}{H} + \chi \right), \qquad \partial^2 \chi = a^2 \epsilon \dot{\mathcal{R}}. \tag{D.44}$$

scalar curvature of the spatial metric

It can be helpful to show the derivation of the scalar curvature of the spatial metric where the three-dimensional Ricci scalar is defined by

$$R^{(3)} = g^{ij} R^{(3)}_{ij}, \tag{D.45}$$

where g_{ij} is the spatial metric induced on the constant-time hypersurfaces. We have previously discussed the fact that the scalar perturbation is contained in the spatial metric and therefore, since $a = a(t)$ depends only on time, it does not contribute to spatial derivatives. Thus, for the purpose of computing the spatial Christoffel symbols, only the spatial dependence of $\mathcal{R}$ would be relevant. The three-dimensional Christoffel symbols are

$$^{(3)}\Gamma^k{}_{ij} = \frac{1}{2} g^{kl} \left(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij} \right). \tag{D.46}$$

Using Equation. (D.15), and keeping only terms linear in $\mathcal{R}$, we find

$$^{(3)}\Gamma^k{}_{ij} = \delta^k_j \partial_i \mathcal{R} + \delta^k_i \partial_j \mathcal{R} - \delta_{ij} \partial^k \mathcal{R}. \tag{D.47}$$

The spatial Ricci tensor is

$$R^{(3)}_{ij} = \partial_k {}^{(3)}\Gamma^k{}_{ij} - \partial_j {}^{(3)}\Gamma^k{}_{ik} + {}^{(3)}\Gamma^k{}_{kl} {}^{(3)}\Gamma^l{}_{ij} - {}^{(3)}\Gamma^k{}_{jl} {}^{(3)}\Gamma^l{}_{ik}. \tag{D.48}$$

At first order, the terms quadratic in the Christoffel symbols are second order in $\mathcal{R}$, and can therefore be dropped. Thus,

$$R^{(3)}_{ij} \simeq \partial_k {}^{(3)}\Gamma^k{}_{ij} - \partial_j {}^{(3)}\Gamma^k{}_{ik}. \tag{D.49}$$

We now compute the two terms separately. First,

$$\begin{aligned}
\partial_k {}^{(3)}\Gamma^k{}_{ij} &= \partial_k \left(\delta^k_j \partial_i \mathcal{R} + \delta^k_i \partial_j \mathcal{R} - \delta_{ij} \partial^k \mathcal{R} \right) \\
&= \partial_j \partial_i \mathcal{R} + \partial_i \partial_j \mathcal{R} - \delta_{ij} \partial^2 \mathcal{R} \\
&= 2 \partial_i \partial_j \mathcal{R} - \delta_{ij} \partial^2 \mathcal{R}.
\end{aligned} \tag{D.50}$$

Next,

$$\begin{aligned} {}^{(3)}\Gamma^{k}{}_{ik} &= \delta^k_k \partial_i \mathcal{R} + \delta^k_i \partial_k \mathcal{R} - \delta_{ik}\partial^k \mathcal{R} \\ &= 3\partial_i \mathcal{R} + \partial_i \mathcal{R} - \partial_i \mathcal{R} \\ &= 3\partial_i \mathcal{R}. \end{aligned} \tag{D.51}$$

Therefore,

$$\partial_j {}^{(3)}\Gamma^{k}{}_{ik} = 3\partial_i \partial_j \mathcal{R}. \tag{D.52}$$

Substituting Equations D.50 and D.52 into Equation D.49, we obtain

$$\begin{aligned} R^{(3)}_{ij} &= \left(2\partial_i\partial_j \mathcal{R} - \delta_{ij}\partial^2 \mathcal{R}\right) - 3\partial_i\partial_j \mathcal{R} \\ &= -\partial_i\partial_j \mathcal{R} - \delta_{ij}\partial^2 \mathcal{R}. \end{aligned} \tag{D.53}$$

The Ricci scalar is obtained by contracting with the inverse spatial metric. Since $R^{(3)}_{ij}$ is already first order, it is sufficient to use the background inverse metric,

$$g^{ij} = \frac{1}{a^2}\delta^{ij} + \mathcal{O}(\mathcal{R}). \tag{D.54}$$

Hence,

$$\begin{aligned} R^{(3)} &= g^{ij} R^{(3)}_{ij} \\ &= \frac{1}{a^2}\delta^{ij}\left(-\partial_i\partial_j \mathcal{R} - \delta_{ij}\partial^2 \mathcal{R}\right) \\ &= -\frac{1}{a^2}\partial^2 \mathcal{R} - \frac{3}{a^2}\partial^2 \mathcal{R} \\ &= -\frac{4}{a^2}\partial^2 \mathcal{R}. \end{aligned} \tag{D.55}$$

Thus, to first order,

$$R^{(3)} = -\frac{4}{a^2}\partial^2 \mathcal{R}. \tag{D.56}$$

It is interesting to note that the same result can be verified directly from the conformal transformation of the spatial curvature. Since $g_{ij} = a^2 e^{2\mathcal{R}}\delta_{ij}$, the spatial metric is conformally flat. We may write

$$g_{ij} = \Omega^2 \bar{g}_{ij}, \qquad \bar{g}_{ij} = \delta_{ij}, \qquad \Omega = a(t)e^{\mathcal{R}}. \tag{D.57}$$

The Ricci scalar of a d-dimensional metric transforms under a conformal rescaling as

$$R = \Omega^{-2}\left[\bar{R} - 2(d-1)\bar{\nabla}^2 \ln\Omega - (d-1)(d-2)\left(\bar{\nabla}\ln\Omega\right)^2\right], \tag{D.58}$$

where barred quantities are computed using the background metric $\bar{g}_{ij}$. In the present case,

$$d = 3, \qquad \bar{R} = 0, \qquad \bar{g}_{ij} = \delta_{ij}. \tag{D.59}$$

Therefore,

$$R^{(3)} = \Omega^{-2}\left[-4\bar{\nabla}^2 \ln\Omega - 2\left(\bar{\nabla}\ln\Omega\right)^2\right]. \tag{D.60}$$

The second term is quadratic in perturbations and may be dropped at linear order. Thus,

$$R^{(3)} \simeq -4\Omega^{-2}\bar{\nabla}^2 \ln\Omega. \tag{D.61}$$

This tells us that the space is conformally flat and the curvature contribution arises entirely from the variation of the conformal factor in space. We have

$$\ln\Omega = \ln a + \mathcal{R}. \tag{D.62}$$

Since $a = a(t)$ has no spatial dependence,

$$\bar{\nabla}^2 \ln\Omega = \bar{\nabla}^2\mathcal{R}. \tag{D.63}$$

Moreover, to first order in the curvature scalar it is sufficient to take

$$\Omega^{-2} = \frac{1}{a^2} + \mathcal{O}(\mathcal{R}), \tag{D.64}$$

because the correction to Ω^{-2} would multiply a first-order quantity and therefore contribute only at second order. Hence,

$$R^{(3)} = -\frac{4}{a^2}\partial^2\mathcal{R}, \tag{D.65}$$

in agreement with Eq. (D.55).

D.7 Partial Contributions

Using previous results, we now compute each of the contributions to the action.

Spatial curvature contribution

We now expand the action to second order in $\mathcal{R}$. The spatial metric is

$$g_{ij} = a^2 e^{2\mathcal{R}}\delta_{ij}. \tag{D.66}$$

The determinant is

$$\sqrt{g} = a^3 e^{3\mathcal{R}}, \tag{D.67}$$

and the three-dimensional Ricci scalar is

$$R^{(3)} = a^{-2} e^{-2\mathcal{R}} \left[-4\partial^2\mathcal{R} - 2(\partial\mathcal{R})^2\right]. \tag{D.68}$$

The contribution from $NR^{(3)}$ term is therefore

$$\begin{aligned}\frac{1}{2}\int dt\, d^3x\, \sqrt{g}\, NR^{(3)} &= \frac{1}{2}\int dt\, d^3x\, a^3 e^{3\mathcal{R}} N a^{-2} e^{-2\mathcal{R}} \left[-4\partial^2\mathcal{R} - 2(\partial\mathcal{R})^2\right] \\ &= \frac{1}{2}\int dt\, d^3x\, aNe^{\mathcal{R}} \left[-4\partial^2\mathcal{R} - 2(\partial\mathcal{R})^2\right].\end{aligned} \tag{D.69}$$

To second order, using

$$Ne^{\mathcal{R}} = 1 + \alpha + \mathcal{R} + \mathcal{O}(2), \tag{D.70}$$

we obtain

$$\sqrt{g}NR^{(3)} = a\left[-4\partial^2\mathcal{R} - 2(\partial\mathcal{R})^2 - 4(\alpha + \mathcal{R})\partial^2\mathcal{R}\right] + \mathcal{O}(3). \tag{D.71}$$

The term linear in $\partial^2\mathcal{R}$ is a total derivative and may be discarded. After integrating by parts,

$$\int d^3x\, \mathcal{R}\partial^2\mathcal{R} = -\int d^3x\, (\partial\mathcal{R})^2. \tag{D.72}$$

Thus the spatial-curvature part contributes

$$\frac{1}{2}\int dt\, d^3x\, a\left[2(\partial\mathcal{R})^2 - 4\alpha\partial^2\mathcal{R}\right], \tag{D.73}$$

up to boundary terms.

Extrinsic-curvature contribution

We next expand

$$N^{-1}\left(E_{ij}E^{ij} - E^2\right). \tag{D.74}$$

Using

$$E^i{}_j = H\delta^i_j + \dot{\mathcal{R}}\delta^i_j - \frac{1}{a^2}\partial^i\partial_j\psi + \mathcal{O}(2), \tag{D.75}$$

we obtain

$$\begin{aligned}E^i{}_j E^j{}_i = 3H^2 + 6H\dot{\mathcal{R}} - \frac{2H}{a^2}\partial^2\psi \\ + 3\dot{\mathcal{R}}^2 - \frac{2\dot{\mathcal{R}}}{a^2}\partial^2\psi + \frac{1}{a^4}\partial^i\partial_j\psi\partial^j\partial_i\psi + \mathcal{O}(3),\end{aligned} \tag{D.76}$$

and

$$E^2 = \left(3H + 3\dot{\mathcal{R}} - \frac{1}{a^2}\partial^2\psi\right)^2$$
$$= 9H^2 + 18H\dot{\mathcal{R}} - \frac{6H}{a^2}\partial^2\psi$$
$$+ 9\dot{\mathcal{R}}^2 - \frac{6\dot{\mathcal{R}}}{a^2}\partial^2\psi + \frac{1}{a^4}(\partial^2\psi)^2 + \mathcal{O}(3). \tag{D.77}$$

Therefore,

$$E^i{}_j E^j{}_i - E^2 = -6H^2 - 12H\dot{\mathcal{R}} + \frac{4H}{a^2}\partial^2\psi$$
$$- 6\dot{\mathcal{R}}^2 + \frac{4\dot{\mathcal{R}}}{a^2}\partial^2\psi + \frac{1}{a^4}\left[\partial^i\partial_j\psi\partial^j\partial_i\psi - (\partial^2\psi)^2\right] + \mathcal{O}(3). \tag{D.78}$$

The last bracket is a total derivative at quadratic order and does not contribute to the action after integration by parts:

$$\int d^3x \left[\partial^i\partial_j\psi\partial^j\partial_i\psi - (\partial^2\psi)^2\right] = 0. \tag{D.79}$$

Multiplying by

$$\sqrt{g}N^{-1} = a^3 e^{3\mathcal{R}}(1 - \alpha + \alpha^2 + \cdots), \tag{D.80}$$

and keeping terms up to quadratic order gives the extrinsic-curvature contribution. The terms linear in perturbations cancel against the linear terms from the scalar-field and potential sectors once the background equations are used.

D.8 Scalar-Field Contribution

In comoving gauge $\delta\phi = 0$, the scalar-field kinetic term is $N^{-1}\dot{\bar{\phi}}^2$. Expanding the lapse dependence gives

$$N^{-1}\dot{\bar{\phi}}^2 = \dot{\bar{\phi}}^2\left(1 - \alpha + \alpha^2\right) + \mathcal{O}(3). \tag{D.81}$$

Additionally, the potential contributes as

$$-2NV(\bar{\phi}) = -2V(\bar{\phi})\left(1 + \alpha\right) + \mathcal{O}(2). \tag{D.82}$$

Including the determinant,

$$\sqrt{g} = a^3\left(1 + 3\mathcal{R} + \frac{9}{2}\mathcal{R}^2 + \cdots\right), \tag{D.83}$$

we obtain the scalar-field and potential contributions to the quadratic action. As in the gravitational sector, all terms linear in perturbations vanish after using the background equations.

D.9 Quadratic Action

After collecting the gravitational and scalar-field contributions, substituting the constraint solutions

$$\alpha = \frac{\dot{\mathcal{R}}}{H}, \qquad \psi = -\frac{\mathcal{R}}{H} + \chi, \qquad \partial^2\chi = a^2\epsilon\dot{\mathcal{R}}, \tag{D.84}$$

and integrating by parts, all dependence on the non-dynamical variables cancels from the final quadratic action. One obtains

$$S^{(2)} = \frac{1}{2}\int dt\, d^3x\, \frac{\dot{\bar{\phi}}^2}{H^2}\left[a^3\dot{\mathcal{R}}^2 - a(\partial\mathcal{R})^2\right]. \tag{D.85}$$

Using

$$\epsilon = \frac{\dot{\bar{\phi}}^2}{2H^2}, \tag{D.86}$$

this can also be written as

$$S^{(2)} = \int dt\, d^3x\, a^3\epsilon\left[\dot{\mathcal{R}}^2 - \frac{1}{a^2}(\partial\mathcal{R})^2\right]. \tag{D.87}$$

Thus the scalar perturbation is given by the comoving curvature perturbation $\mathcal{R}$.

D.10 Conformal Time and the Mukhanov Variable

We now introduce conformal time τ, defined by

$$d\tau = \frac{dt}{a(t)}. \tag{D.88}$$

Derivatives with respect to τ will be denoted by primes:

$$\mathcal{R}' = \frac{d\mathcal{R}}{d\tau} = a\dot{\mathcal{R}}. \tag{D.89}$$

The quadratic action becomes

$$S^{(2)} = \frac{1}{2}\int d\tau\, d^3x\, z^2\left[(\mathcal{R}')^2 - (\partial\mathcal{R})^2\right], \tag{D.90}$$

where

$$z := a\frac{\dot{\bar{\phi}}}{H} = a\sqrt{2\epsilon}, \tag{D.91}$$

is the Mukhanov–Sasaki variable is defined by

$$\nu := z\mathcal{R}. \tag{D.92}$$

Equivalently,

$$\mathcal{R} = \frac{\nu}{z}. \tag{D.93}$$

Then

$$\mathcal{R}' = \frac{\nu'}{z} - \frac{z'}{z^2}\nu. \tag{D.94}$$

Substituting this into Eq. (D.90) gives

$$\begin{aligned} S^{(2)} &= \frac{1}{2}\int d\tau\, d^3x \left[\left(\nu' - \frac{z'}{z}\nu\right)^2 - (\partial\nu)^2\right] \\ &= \frac{1}{2}\int d\tau\, d^3x \left[(\nu')^2 - 2\frac{z'}{z}\nu\nu' + \left(\frac{z'}{z}\right)^2 \nu^2 - (\partial\nu)^2\right]. \end{aligned} \tag{D.95}$$

The mixed term can be integrated by parts:

$$\begin{aligned} -2\frac{z'}{z}\nu\nu' &= -\frac{z'}{z}(\nu^2)' \\ &\longrightarrow \left(\frac{z'}{z}\right)' \nu^2, \end{aligned} \tag{D.96}$$

up to a boundary term. Therefore,

$$\left(\frac{z'}{z}\right)' + \left(\frac{z'}{z}\right)^2 = \frac{z''}{z}. \tag{D.97}$$

The quadratic action becomes

$$S^{(2)} = \frac{1}{2}\int d\tau\, d^3x \left[(\nu')^2 - (\partial\nu)^2 + \frac{z''}{z}\nu^2\right]. \tag{D.98}$$

This is the action of a canonically normalized scalar variable with a time-dependent effective mass.

D.11 Mukhanov–Sasaki Equation

Varying Eq. (D.98) with respect to ν gives

$$\nu'' - \partial^2\nu - \frac{z''}{z}\nu = 0. \tag{D.99}$$

Expanding in Fourier modes,

$$\nu(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \nu_k(\tau) e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{D.100}$$

and using

$$\partial^2 \nu \to -k^2 \nu_k, \tag{D.101}$$

one obtains the Mukhanov–Sasaki equation

$$\nu_k'' + \left(k^2 - \frac{z''}{z}\right) \nu_k = 0. \tag{D.102}$$

This equation describes the evolution of scalar perturbations during inflation. Deep inside the Hubble radius, where $k^2 \gg z''/z$, the modes behave approximately as ordinary harmonic oscillators. On super-Hubble scales, where $k^2 \ll z''/z$, the time-dependent background controls the evolution, and the curvature perturbation $\mathcal{R} = \nu/z$ approaches a conserved mode for adiabatic single-field inflation.

D.12 Relation to the Gauge-Invariant Field Fluctuation

The gauge-invariant scalar-field fluctuation on spatially flat hypersurfaces is

$$Q = \delta\phi + \frac{\dot{\bar{\phi}}}{H} \Psi. \tag{D.103}$$

Here Ψ denotes the scalar curvature perturbation appearing in the general scalar metric decomposition. The comoving curvature perturbation is

$$\mathcal{R} = \Psi + H \frac{\delta\phi}{\dot{\bar{\phi}}}. \tag{D.104}$$

Combining the two definitions gives

$$Q = \frac{\dot{\bar{\phi}}}{H} \mathcal{R}. \tag{D.105}$$

Therefore

$$\nu = z\mathcal{R} = a\frac{\dot{\bar{\phi}}}{H} \mathcal{R} = aQ. \tag{D.106}$$

D.13 Connection with the Scalar Power Spectrum

The canonical action (D.98) provides the starting point for quantization. The mode functions ν_k are promoted to operators and normalized by the canonical commutation relations. The curvature perturbation is recovered from

$$\mathcal{R}_k = \frac{\nu_k}{z}. \tag{D.107}$$

The scalar power spectrum is defined by

$$\langle \mathcal{R}_{\mathbf{k}} \mathcal{R}_{\mathbf{k}'} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{k}') P_{\mathcal{R}}(k), \tag{D.108}$$

with dimensionless form

$$\Delta^2_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k). \tag{D.109}$$

In the slow-roll limit, evaluating the mode functions around horizon crossing gives

$$\Delta^2_{\mathcal{R}}(k) \simeq \frac{1}{8\pi^2} \frac{H^2}{\epsilon}\bigg|_{k=aH}, \tag{D.110}$$

in units $M_{\rm Pl} = 1$. Restoring the Planck mass gives

$$\Delta^2_{\mathcal{R}}(k) \simeq \frac{1}{8\pi^2 M_{\rm Pl}^2} \frac{H^2}{\epsilon}\bigg|_{k=aH}. \tag{D.111}$$

This is the result used in the main text when connecting single-field inflationary dynamics to the observed scalar spectrum.

Bibliography

[1] Tom WB Kibble. Lorentz invariance and the gravitational field. *Journal of mathematical physics*, 2(2):212–221, 1961.

[2] Élie Cartan. Sur une généralisation de la notion de courbure de riemann et les espaces à torsion. *Comptes Rendus, Ac. Sc. Paris*, 174:593–595, 1922.

[3] Hans F Westman and Tom G Zlosnik. Cartan gravity, matter fields, and the gauge principle. *Annals of Physics*, 334:157–197, 2013.

[4] Abhay Ashtekar. New hamiltonian formulation of general relativity. *Physical Review D*, 36(6):1587, 1987.

[5] Milutin Blagojević and Branislav Cvetković. General Poincaré gauge theory: Hamiltonian structure and particle spectrum. *Phys. Rev. D*, 98:024014, 2018. doi: 10.1103/PhysRevD.98.024014.

[6] Alan H Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23(2):347, 1981.

[7] Alexei A. Starobinsky. A New Type of Isotropic Cosmological Models Without Singularity. *Phys. Lett. B*, 91:99–102, 1980. doi: 10.1016/0370-2693(80)90670-X.

[8] Tristan L Smith and Robert R Caldwell. Lisa for cosmologists: calculating the signal-to-noise ratio for stochastic and deterministic sources. *Physical Review D*, 100 (10):104055, 2019.

[9] Matthias Bartelmann and Peter Schneider. Weak gravitational lensing. *Physics Reports*, 340(4-5):291–472, 2001.

[10] Mehraveh Nikjoo and Tom Zlosnik. Hamiltonian formulation of gravity as a spontaneously-broken gauge theory of the lorentz group. *Classical and Quantum Gravity*, 41(4):045005, 2024.

[11] Tomi S. Koivisto and Tom Zlosnik. Paths to gravitation via the gauging of parametrized field theories. *Phys. Rev. D*, 107(12):124013, 2023. doi: 10.1103/PhysRevD.107.124013.

[12] Mehraveh Nikjoo, Joe Zuntz, and Ben Moews. Weak lensing by photometric density ridges. *arXiv preprint arXiv:2603.04025*, 2026.

[13] David Alonso, Mehraveh Nikjoo, Arianna I Renzini, Emilio Bellini, and Pedro G Ferreira. Tomographic constraints on the production rate of gravitational waves from astrophysical sources. *Physical Review D*, 110(10):103544, 2024.

[14] Abhay Ashtekar. New variables for classical and quantum gravity. *Physical review letters*, 57(18):2244, 1986.

[15] Walther Gerlach and Otto Stern. Der experimentelle nachweis der richtungsquantelung im magnetfeld. In *Walther Gerlach (1889–1979) Eine Auswahl aus seinen Schriften und Briefen*, pages 26–29. Springer, 1989.

[16] Roger Penrose and Wolfgang Rindler. *Spinors and space-time: Volume 2, Spinor and twistor methods in space-time geometry*, volume 2. Cambridge university press, 1984.

[17] Pertti Lounesto. Clifford algebras and spinors. In *Clifford algebras and their applications in mathematical physics*, pages 25–37. Springer, 2001.

[18] Attilio Palatini. Deduzione invariantiva delle equazioni gravitazionali dal principio di hamilton. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 43(1):203–212, 1919.

[19] Theodore Frankel. *The geometry of physics: an introduction*. Cambridge university press, 2004.

[20] Sören Holst. Barbero's hamiltonian derived from a generalized hilbert-palatini action. *Physical Review D*, 53(10):5966, 1996.

[21] J Fernando Barbero. Real ashtekar variables for lorentzian signature space-times. *arXiv preprint gr-qc/9410014*, 1994.

[22] Samuel W MacDowell and Fereydoon Mansouri. Unified geometric theory of gravity and supergravity. *Physical Review Letters*, 38(14):739, 1977.

[23] Jacques Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*. Courier Corporation, 2014.

[24] Chen-Ning Yang and Robert L Mills. Conservation of isotopic spin and isotopic gauge invariance. *Physical review*, 96(1):191, 1954.

[25] Richard L Arnowitt, Stanley Deser, and Charles W Misner. The dynamics of general relativity. *Gen. Rel. Grav.*, 40(arXiv: gr-qc/0405109):1997–2027, 2008.

[26] Paul AM Dirac. *Lectures on quantum mechanics.* Courier Corporation, 2013.

[27] David W Hogg, Daniel J Eisenstein, Michael R Blanton, Neta A Bahcall, Jon Brinkmann, James E Gunn, and Donald P Schneider. Cosmic homogeneity demonstrated with luminous red galaxies. *The Astrophysical Journal*, 624(1):54–58, 2005.

[28] Alexander Friedmann. Über die krümmung des raumes. *Z. Phys.*, 10:377–386, 1922.

[29] Wuzheng Guo, Qiumin Wang, Shuo Cao, Marek Biesiada, Tonghua Liu, Yujie Lian, Xinyue Jiang, Chengsheng Mu, and Dadian Cheng. Newest measurements of hubble constant from desi 2024 baryon acoustic oscillation observations. *The Astrophysical Journal Letters*, 978(2):L33, 2025.

[30] Hubert Goenner. Weyl's contributions to cosmology. In *Hermann Weyl's Raum—Zeit—Materie and a General Introduction to His Scientific Work*, pages 105–137. Springer, 2001.

[31] N Aghanim et al. Planck 2018 results. vi. cosmological parameters. *Astron. Astrophys*, 641:A6, 2020.

[32] Michael S Turner. Dark matter and dark energy in the universe. *Physica Scripta*, 2000(T85):210–220, 2000.

[33] Daniel Moshe Scolnic, DO Jones, A Rest, YC Pan, R Chornock, RJ Foley, ME Huber, R Kessler, Gautham Narayan, AG Riess, et al. The complete light-curve sample of spectroscopically confirmed sne ia from pan-starrs1 and cosmological constraints from the combined pantheon sample. *The Astrophysical Journal*, 859(2):101, 2018.

[34] AG Adame, J Aguilar, S Ahlen, Se Alam, DM Alexander, M Alvarez, O Alves, A Anand, U Andrade, E Armengaud, et al. Desi 2024 vi: cosmological constraints from the measurements of baryon acoustic oscillations. *Journal of Cosmology and Astroparticle Physics*, 2025(02):021, 2025.

[35] Timothy MC Abbott, Michel Aguena, Alex Alarcon, S Allam, O Alves, A Amon, F Andrade-Oliveira, James Annis, S Avila, D Bacon, et al. Dark energy survey year 3 results: Cosmological constraints from galaxy clustering and weak lensing. *Physical Review D*, 105(2):023520, 2022.

[36] Stephen William Hawking and Roger Penrose. The singularities of gravitational collapse and cosmology. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 314(1519):529–548, 1970.

[37] Andrei D Sakharov. Violation of cp-invariance, c-asymmetry, and baryon asymmetry of the universe. In *In The Intermissions... Collected Works on Research into the*

Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16, pages 84–87. World Scientific, 1998.

[38] Yasumichi Aoki, G Endrődi, Zoltán Fodor, Sándor D Katz, and Kálmán K Szabó. The order of the quantum chromodynamics transition predicted by the standard model of particle physics. *Nature*, 443(7112):675–678, 2006.

[39] Scott Dodelson and Fabian Schmidt. *Modern cosmology*. Academic press, 2020.

[40] Richard H Cyburt, Brian D Fields, Keith A Olive, and Tsung-Han Yeh. Big bang nucleosynthesis: Present status. *Reviews of Modern Physics*, 88(1):015004, 2016.

[41] Viatcheslav Mukhanov. *Physical foundations of cosmology*. Cambridge university press, 2005.

[42] Alexei A Starobinsky. A new type of isotropic cosmological models without singularity. *Physics Letters B*, 91(1):99–102, 1980.

[43] Anish Ghoshal, Paweł Kozów, Marek Olechowski, and Stefan Pokorski. Cmb observables and reheat temperature as a window to models of inflation and freeze-in dark matter production. *arXiv preprint arXiv:2510.27587*, 2025.

[44] Miklos Långvik, Juha-Matti Ojanperä, Sami Raatikainen, and Syksy Räsänen. Higgs inflation with the holst and the nieh–yan term. *Physical Review D*, 103(8):083514, 2021.

[45] Daniel Baumann. Tasi lectures on inflation. *arXiv preprint arXiv:0907.5424*, 2009.

[46] Hannu Kurki-Suonio. Cosmological perturbation theory. *unpublished lecture notes available: http://theory. physics. helsinki. _/cmb/(referenced 23 Oct. 2005)*, 2005.

[47] James M Bardeen. Gauge-invariant cosmological perturbations. *Physical Review D*, 22(8):1882, 1980.

[48] Daniel Baumann. *Cosmology*. Cambridge University Press, 2022.

[49] Viatcheslav F Mukhanov. Quantum theory of gauge-invariant cosmological perturbations. *Zh. Eksp. Teor. Fiz*, 94(1), 1988.

[50] Philip JE Peebles and JT Yu. Primeval adiabatic perturbation in an expanding universe. *Astrophysical Journal, vol. 162, p. 815*, 162:815, 1970.

[51] Rashid A Sunyaev and Ya B Zeldovich. Small-scale fluctuations of relic radiation. *Astrophysics and Space Science*, 7(1):3–19, 1970.

[52] Joseph Silk. Cosmic black-body radiation and galaxy formation. *Astrophysical Journal, vol. 151, p. 459*, 151:459, 1968.

[53] Sergei Bashinsky and Uros Seljak. Neutrino perturbations in cmb anisotropy and matter clustering. *Phys. Rev. D*, 69(083002):165, 2004.

[54] Benjamin P Abbott, Richard Abbott, Thomas D Abbott, Matthew R Abernathy, Fausto Acernese, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, Rana X Adhikari, et al. Observation of gravitational waves from a binary black hole merger. *Physical review letters*, 116(6):061102, 2016.

[55] Charles W Misner, Kip S Thorne, and John Archibald Wheeler. *Gravitation*. Macmillan, 1973.

[56] Richard A Isaacson. Gravitational radiation in the limit of high frequency. ii. nonlinear terms and the effective stress tensor. *Physical Review*, 166(5):1272, 1968.

[57] James B Hartle. *Gravity: an introduction to Einstein's general relativity*. Cambridge University Press, 2021.

[58] Joseph H Taylor and Joel M Weisberg. A new test of general relativity-gravitational radiation and the binary pulsar psr 1913+ 16. *Astrophysical Journal, Part 1, vol. 253, Feb. 15, 1982, p. 908-920. NSF-supported research.*, 253:908–920, 1982.

[59] Joel M Weisberg and Joseph H Taylor. Relativistic binary pulsar b1913+ 16: Thirty years of observations and analysis. *arXiv preprint astro-ph/0407149*, 2004.

[60] Andrzej Trautman. Radiation and boundary conditions in the theory of gravitation. *arXiv preprint arXiv:1604.03145*, 2016.

[61] Judit Prat and David Bacon. Weak gravitational lensing. In *Encyclopedia of Astrophysics, Volume 5*, volume 5, pages 508–537, January 2026. doi: 10.1016/B978-0-443-21439-4.00071-7.

[62] Luigi Toffolatti, Planck Collaboration, et al. Planck 2013 results. xvi. cosmological parameters. *Astronomy and Astrophysics, 571, A16*, 2014.

[63] Jerome Martin, Christophe Ringeval, and Vincent Vennin. Encyclopædia inflationaris. *Physics of the Dark Universe*, 5:75–235, 2014.

[64] Florian Bauer and Durmuş A Demir. Inflation with non-minimal coupling: Metric vs. palatini formulations. *Physics Letters B*, 665(4):222–226, 2008.

[65] David I Kaiser. Conformal transformations with multiple scalar fields. *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, 81(8):084044, 2010.

[66] Tomo Takahashi and Tommi Tenkanen. Towards distinguishing variants of non-minimal inflation. *Journal of Cosmology and Astroparticle Physics*, 2019(04):035–035, 2019.

[67] Mikhail Shaposhnikov, Andrey Shkerin, Inar Timiryasov, and Sebastian Zell. Higgs inflation in einstein-cartan gravity. *Journal of Cosmology and Astroparticle Physics*, 2021(02):008–008, 2021.

[68] Peter AR Ade, Z Ahmed, M Amiri, D Barkats, R Basu Thakur, CA Bischoff, D Beck, JJ Bock, H Boenish, E Bullock, et al. Improved constraints on primordial gravitational waves using planck, wmap, and bicep/keck observations through the 2018 observing season. *Physical review letters*, 127(15):151301, 2021.

[69] Daniel Foreman-Mackey, David W Hogg, Dustin Lang, and Jonathan Goodman. emcee: the mcmc hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312, 2013.

[70] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[71] Laura Bethke and Joao Magueijo. Inflationary tensor fluctuations, as viewed by ashtekar variables and their generalizations. *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, 84(2):024014, 2011.

[72] Carlo R Contaldi, Joao Magueijo, and Lee Smolin. Anomalous cosmic-microwave-background polarization and gravitational chirality. *Physical review letters*, 101(14):141101, 2008.

[73] Nicholas Battaglia, J Richard Bond, Christoph Pfrommer, and JL Sievers. On the cluster physics of sunyaev–zel'dovich and x-ray surveys. ii. deconstructing the thermal sz power spectrum. *The Astrophysical Journal*, 758(2):75, 2012.

[74] James M Bardeen, JR Bond, Nick Kaiser, and AS Szalay. The statistics of peaks of gaussian random fields. *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 304, May 1, 1986, p. 15-61. SERC-supported research.*, 304:15–61, 1986.

[75] Phillip James Edwin Peebles. The large-scale structure of the universe. 2020.

[76] Christopher R Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Nonparametric ridge estimation. 2014.

[77] Planck Collaboration. Planck 2018 results. vi. cosmological parameters. *Astronomy & Astrophysics*, 641:A6, 2020.

[78] Joseph B Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1):48–50, 1956.

[79] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[80] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[81] Nicolas Tessore, Arthur Loureiro, Benjamin Joachimi, Maximilian von Wietersheim-Kramsta, and Niall Jeffrey. GLASS: Generator for Large Scale Structure. *The Open Journal of Astrophysics*, 6:11, March 2023. doi: 10.21105/astro.2302.01942.

[82] Erin S. Sheldon and Eric M. Huff. Practical Weak-lensing Shear Measurement with Metacalibration. , 841(1):24, May 2017. doi: 10.3847/1538-4357/aa704b.

[83] Anna Porredon, M Crocce, J Elvin-Poole, R Cawthon, G Giannini, J De Vicente, A Carnero Rosell, Ismael Ferrero, E Krause, X Fang, et al. Dark energy survey year 3 results: Cosmological constraints from galaxy clustering and galaxy-galaxy lensing using the maglim lens sample. *Physical Review D*, 106(10):103530, 2022.

[84] Nelson Christensen. Stochastic gravitational wave backgrounds. *Reports on Progress in Physics*, 82(1):016903, 2019.

[85] Max Tegmark. How to measure cmb power spectra without losing information. *Physical Review D*, 55(10):5895, 1997.

[86] Lloyd Knox. Cosmic microwave background anisotropy observing strategy assessment. *The Astrophysical Journal*, 480(1):72–78, 1997.

[87] Chris J Isham. Canonical quantum gravity and the problem of time. In *Integrable systems, quantum groups, and quantum field theories*, pages 157–287. Springer, 1993.