Universitat Autònoma de Barcelona

Study of the Relativistic Dynamics of Extreme-Mass-Ratio Inspirals

Marius Oltean

Study of the Relativistic Dynamics of Extreme-Mass-Ratio Inspirals

Marius Oltean

PhD Thesis, International Joint Doctorate
September 2019

Universitat Autònoma de Barcelona
Departament de Física, Programa de Doctorat RD 99/2011
Institut de Ciències de l'Espai (ICE, CSIC)
Institut d'Estudis Espacials de Catalunya (IEEC)

Université d'Orléans
Observatoire des Sciences de l'Univers en région Centre (OSUC)
Laboratoire de Physique et Chimie de l'Environnement et de l'Espace (LPC2E, CNRS)

Advisors:
Carlos Fernández Sopuerta (ICE, CSIC and IEEC)
Alessandro Spallicci (Université d'Orléans, OSUC and LPC2E, CNRS)

Tutor:
Diego Pavón Coloma (Universitat Autònoma de Barcelona)

Contents

General Thesis Summary vii

Resum General de la Tesi (translation in Catalan) viii

Résumé Général de la Thèse (translation in French) ix

List of Original Contributions xi

Notation and Conventions xiii

List of Figures xv

Acknowledgements xxi

Part I. Fundamentals of General Relativity:
Introduction, Canonical Formulation and Perturbation Theory 1

Chapter 1. Introduction 3
- 1.1. *Gravitatio mundi*, a brief historical prelude 6
- 1.2. The advent of relativity 11
- 1.3. Geometry, gravity and motion 16
- 1.4. Gravitational waves and extreme-mass-ratio inspirals 20
- 1.5. The self-force problem 22

Chapter 2. Canonical General Relativity 29
- 2.1. Introduction 32
- 2.2. Lagrangian formulation 34
- 2.3. Canonical formulation of general field theories 38
- 2.4. Canonical formulation of general relativity 54
- 2.5. Applications 61

Chapter 3. General Relativistic Perturbation Theory 71
- 3.1. Introduction 73
- 3.2. General formulation of perturbation theory 75

3.3. Perturbations of the Schwarzschild-Droste spacetime 83
3.4. Perturbations of the Kerr spacetime 88

Part II. Novel Contributions:
Entropy, Motion and Self-Force in General Relativity 93

Chapter 4. Entropy Theorems and the Two-Body Problem 95
4.1. Introduction 98
4.2. Entropy theorems in classical mechanics 102
4.3. Entropy theorems in general relativity 115
4.4. Entropy in the gravitational two-body problem 125
4.5. Conclusions 129

Chapter 5. The Motion of Localized Sources in General Relativity: Gravitational Self-Force from Quasilocal Conservation Laws 131
5.1. Introduction: the self-force problem via conservation laws 135
5.2. Setup: quasilocal conservation laws 139
5.3. General derivation of the gravitational self-force from quasilocal conservation laws 149
5.4. Application to the Gralla-Wald approach to the gravitational self-force 154
5.5. Discussion and conclusions 172
5.6. Appendix: conformal killing vectors and the two-sphere 178

Chapter 6. A Frequency-Domain Implementation of the Particle-without-Particle Approach to EMRIs 181
6.1. Introduction 183
6.2. The scalar self-force 184
6.3. The Particle-without-Particle method 187
6.4. Frequency domain analysis 188
6.5. Numerical implementation and results 190

Chapter 7. Conclusions 195

Appendix A. Topics in Differential Geometry: Maps on Manifolds, Lie Derivatives, Forms and Integration 203
A.1. Maps on manifolds 204
A.2. Lie derivatives 206
A.3. Differential forms 209
A.4. Integration on manifolds 211

Appendix B. Particle-without-Particle:
A Practical Pseudospectral Collocation Method for Linear Partial Differential Equations with Distributional Sources 215
B.1. Introduction 218
B.2. Setup 223
B.3. The Particle-without-Particle method 226
B.4. First order hyperbolic PDEs 234
B.5. Parabolic PDEs 240
B.6. Second order hyperbolic PDEs 245
B.7. Elliptic PDEs 251
B.8. Conclusions 252
B.9. Appendix: pseudospectral numerical schemes 253

Bibliography 261

General Thesis Summary

The principal subject of this thesis is the gravitational two-body problem in the extreme-mass-ratio regime—that is, where one mass is significantly smaller than the other—in the full context of our contemporary theory of gravity, general relativity. We divide this work into two broad parts: the first provides an overview of the theory of general relativity along with the basic mathematical methods underlying it, focusing on its canonical formulation and perturbation techniques; the second is dedicated to a presentation of our novel work in these areas, focusing on the problems of entropy, motion and the self-force in general relativity.

We begin in Part I, accordingly, by offering a historical introduction to general relativity as well as a discussion on current motivation from gravitational wave astronomy in Chapter 1. Then, in Chapter 2, we turn to a detailed technical exposition of this theory, focusing on its canonical (Hamiltonian) formulation. We end this part of the thesis with a rigorous development of perturbation methods in Chapter 3. For the convenience of the reader, we summarize some basic concepts in differential geometry needed for treating these topics in Appendix A.

In Part II, we begin with a study of entropy theorems in classical Hamiltonian systems in Chapter 4, and in particular, the issue of the second law of thermodynamics in classical mechanics and general relativity, with a focus on the gravitational two-body problem. Then in Chapter 5, we develop a general approach based on conservation laws for calculating the correction to the motion of a sufficiently small object due to gravitational perturbations in general relativity. When the perturbations are attributed to the small object itself, this effect is known as the gravitational self-force. It is what drives the orbital evolution of extreme-mass-ratio inspirals: compact binary systems where one mass is much smaller than—thus effectively orbiting and eventually spiralling into—the other, expected to be among the main sources for the future space-based gravitational wave detector LISA. In Chapter 6, we present some work on the numerical computation of the scalar self-force—a helpful testbed for the gravitational case—for circular orbits in the frequency domain, using a method for tackling distributional sources in the field equations called the Particle-without-Particle method. We include also, in Appendix B, some work

on the generalization of this method to general partial differential equations with distributional sources, including also applications to other areas of applied mathematics. We summarize our findings in this thesis and offer some concluding reflections in Chapter 7.

Resum General de la Tesi
(translation in Catalan)

El tema principal d'aquesta tesi és el problema gravitacional de dos cossos en el règim de raons de masses extremes - és a dir, on una massa és significativament més petita que l'altra - en el context complet de la nostra teoria contemporània de la gravetat, la relativitat general. Dividim aquest treball en dues grans parts: la primera proporciona una visió general de la teoria de la relativitat general juntament amb els mètodes bàsics matemàtics en què s'hi basa, centrant-se en la seva formulació canònica i les tècniques de pertorbació; la segona està dedicada a presentar la nostra contribució en aquests àmbits, centrada en els problemes de l'entropia, el moviment i la força pròpia en la relativitat general.

Comencem a la part I, en conseqüència, oferint una introducció històrica a la relativitat general, així com una discussió sobre la motivació actual a partir de l'astronomia d'ones gravitacionals al capítol 1. A continuació, al capítol 2, passem a una exposició tècnica detallada d'aquesta teoria, centrada sobre la seva formulació canònica (hamiltoniana). Acabem aquesta part de la tesi amb un desenvolupament rigorós de mètodes de pertorbació al capítol 3. Per a la comoditat del lector, resumim alguns conceptes bàsics en geometria diferencial necessaris per a tractar aquests temes a l'apèndix A.

A la part II, comencem amb un estudi dels teoremes d'entropia en sistemes clàssics hamiltonians al capítol 4, i en particular, la qüestió de la segona llei de la termodinàmica en la mecànica clàssica i la relativitat general, amb el focus en el problema gravitatori de dos cossos. Al capítol 5, desenvolupem una anàlisi general basada en lleis de conservació per a calcular la correcció en el moviment d'un objecte prou petit a causa de les pertorbacions gravitacionals de la relativitat general. Quan les pertorbacions s'atribueixen al propi objecte petit, aquest efecte es coneix com a força pròpia gravitacional. És el que impulsa l'evolució orbital de les caigudes en espiral amb raó de masses extrema: sistemes binaris compactes on una massa és molt menor que - i per tant, efectivament orbita i, finalment, fa espirals cap a - l'altre. Es preveu que siguin una de les principals fonts del futur detector d'ones gravitacionals LISA, situat en l'espai. Al capítol 6, es presenta un treball sobre el càlcul numèric de la força pròpia escalar - una prova útil per al cas gravitatori - per òrbites circulars en el domini de freqüència, utilitzant un mètode per abordar fonts de distribució en les equacions de camp anomenat el mètode Partícula-sense-Partícula. Incloem també, en l'apèndix B, alguns treballs sobre la generalització d'aquest mètode a equacions diferencials parcials generals amb fonts distribucionals, incloent també aplicacions a altres àrees

de matemàtiques aplicades. Resumim els nostres resultats en aquesta tesi i oferim algunes reflexions finals al capítol 7.

Résumé Général de la Thèse

(translation in French)

Le sujet principal de cette thèse est le problème gravitationnel à deux corps dans le régime des quotients extrêmes des masses - c'est-à-dire où une masse est nettement plus petite que l'autre - dans le contexte complet de notre théorie contemporaine de la gravité, la relativité générale. Nous divisons ce travail en deux grandes parties : la première fournit un aperçu de la théorie de la relativité générale ainsi que des méthodes mathématiques de base qui la sous-tendent, en mettant l'accent sur sa formulation canonique et les techniques de perturbation; la seconde est consacrée à une présentation de notre travail novateur dans ces domaines, en se concentrant sur les problèmes de l'entropie, du mouvement et de la force propre dans la relativité générale.

Nous commençons par la partie I en proposant une introduction historique à la relativité générale ainsi qu'une discussion sur la motivation actuelle à partir de l'astronomie des ondes gravitationnelles au chapitre 1. Ensuite, au chapitre 2, nous abordons un exposé technique détaillé de cette théorie, en nous concentrant sur sa formulation canonique (hamiltonienne). Nous terminons cette partie de la thèse par un développement rigoureux des méthodes de perturbation au chapitre 3. Pour la commodité du lecteur, nous résumons quelques concepts de base de la géométrie différentielle nécessaires pour traiter ces sujets dans l'annexe A.

Dans la partie II, nous commencerons par une étude des théorèmes de l'entropie dans les systèmes hamiltoniens classiques au chapitre 4, et en particulier par la question de la deuxième loi de la thermodynamique dans la mécanique classique et la relativité générale, en mettant l'accent sur le problème gravitationnel à deux corps. Ensuite, au chapitre 5, nous développons une analyse générale basée sur les lois de conservation pour calculer la correction au mouvement d'un objet suffisamment petit dues aux perturbations gravitationnelles dans la relativité générale. Lorsque les perturbations sont attribuées au petit objet lui-même, cet effet s'appelle la force propre gravitationnelle. C'est ce que détermine l'évolution orbitale des inspirals avec quotients extrêmes des masses : des systèmes binaires compacts dans lesquels une masse est beaucoup plus petite que - effectivement orbitant et finissant en faire des spirales dans - l'autre. On s'attend à ce qu'elles soient l'une des principales sources et parmi les plus intéressantes pour le futur détecteur spatial d'ondes gravitationnelles LISA. Au chapitre 6, nous présentons quelques travaux sur le calcul numérique de la force propre scalaire - un test utile pour le cas gravitationnel - pour les orbites circulaires dans le domaine fréquentiel, en utilisant une méthode pour traiter les

sources distributionnelles dans les équations de champ appelée la méthode Particule-sans-Particule. Nous incluons également, dans l'annexe B, des travaux sur la généralisation de cette méthode aux équations aux dérivées partielles générales avec sources distributionnelles, ainsi que des applications à d'autres domaines des mathématiques appliquées. Nous résumons nos résultats de cette thèse et proposons quelques réflexions finales au chapitre 7.

List of Original Contributions

Here we list the chapters of this thesis containing original contributions, along with their corresponding publication/preprint information as they appear also in the bibliography.

- **Chapter 4**: [Oltean, Bonetti, et al. 2016] M. Oltean, L. Bonetti, A. D. A. M. Spallicci, and C. F. Sopuerta, "Entropy theorems in classical mechanics, general relativity, and the gravitational two-body problem", Physical Review D **94**, 064049 (2016).

- **Chapter 5**: [Oltean, Epp, Sopuerta, et al. 2019] M. Oltean, R. J. Epp, C. F. Sopuerta, A. D. A. M. Spallicci, and R. B. Mann, "The motion of localized sources in general relativity: gravitational self-force from quasilocal conservation laws", arXiv:1907.03012 [astro-ph, physics:gr-qc, physics:hep-th] (2019). To be submitted to Physical Review D.

- **Chapter 6**: [Oltean, Sopuerta, et al. 2017] M. Oltean, C. F. Sopuerta, and A. D. A. M. Spallicci, "A frequency-domain implementation of the particle-without-particle approach to EMRIs", Journal of Physics: Conference Series **840**, 012056 (2017).

- **Appendix B**: [Oltean, Sopuerta, et al. 2019] M. Oltean, C. F. Sopuerta, and A. D. A. M. Spallicci, "Particle-without-Particle: A Practical Pseudospectral Collocation Method for Linear Partial Differential Equations with Distributional Sources", Journal of Scientific Computing **79**, 827 (2019).

Notation and Conventions

Here we summarize the basic notation and mathematical conventions used throughout this thesis. For more details, definitions and useful results, see Appendix A.

We use script upper-case letters ($\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, ...) typically for denoting mathematical spaces (manifolds, curves, etc.). The n-dimensional Euclidean space is denoted as usual by $\mathbb{R}^n$, the n-sphere of radius r by $\mathbb{S}^n_r$, and the unit n-sphere by $\mathbb{S}^n = \mathbb{S}^n_1$. For any two spaces $\mathscr{A}$ and $\mathscr{B}$ that are topologically equivalent (*i.e.* homeomorphic), we indicate this by writing $\mathscr{A} \simeq \mathscr{B}$.

The set of (k,l)-tensors (tensors with k covariant indices and l contravariant indices) on any manifold $\mathscr{U}$ is denoted by $\mathscr{T}^k{}_l(\mathscr{U})$. In particular, $T\mathscr{U} = \mathscr{T}^1{}_0(\mathscr{U})$ is the tangent bundle and $T^*\mathscr{U} = \mathscr{T}^0{}_1(\mathscr{U})$ the dual thereto (the cotangent bundle).

A spacetime is a $(3{+}1)$-dimensional Lorentzian manifold, typically denoted by $\mathscr{M}$. We work in the $(-,+,+,+)$ signature. Any (k,l)-tensor in $\mathscr{M}$ is equivalently denoted either using the (boldface) index-free notation $\boldsymbol{A} \in \mathscr{T}^k{}_l(\mathscr{M})$ following the practice of, e.g., [Misner et al. 1973; Hawking and Ellis 1975], or the abstract index notation $A^{a_1\cdots a_k}{}_{b_1\cdots b_l} \in \mathscr{T}^k{}_l(\mathscr{M})$ following that of, e.g., [Wald 1984]; that is, depending upon convenience, we equivalently write

$$\boldsymbol{A} = A^{a_1\cdots a_k}{}_{b_1\cdots b_l} \in \mathscr{T}^k{}_l(\mathscr{M})\,, \tag{0.0.1}$$

with Latin letters from the beginning of the alphabet (a, b, c, ...) being used for spacetime indices $(0,1,2,3)$. The components of $\boldsymbol{A}$ in a particular choice of coordinates $\{x^\alpha\}^3_{\alpha=0}$ are denoted by $A^{\alpha_1\cdots\alpha_k}{}_{\beta_1\cdots\beta_l}$, using Greek (rather than Latin) letters from the beginning of the alphabet (α, β, γ, ...). Spatial indices on an appropriately defined (three-dimensional Riemannian spacelike) constant time slice of $\mathscr{M}$ are denoted using Latin letters from the middle third of the alphabet in Roman font: in lower-case (i, j, k, ...) if they are abstract, and in upper-case (I, J, K, ...) if a particular choice of coordinates $\{x^I\}^3_{I=1}$ has been made.

More generally, discussing any n-dimensional manifold of interest, we may write this as a collection of objects $(\mathscr{U}, \boldsymbol{g}_{\mathscr{U}}, \boldsymbol{\nabla}_{\mathscr{U}})$, where $\mathscr{U}$ is the manifold itself, $\boldsymbol{g}_{\mathscr{U}}$ is a metric defined on it, and $\boldsymbol{\nabla}_{\mathscr{U}}$ the derivative operator compatible with this metric. Its natural volume form is by

$$\boldsymbol{\epsilon}_{\mathscr{U}} = \sqrt{\left|\det\left(\boldsymbol{g}_{\mathscr{U}}\right)\right|}\;\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n\,, \tag{0.0.2}$$

where $\wedge$ is the wedge product.

Let $\mathscr{S} \simeq \mathbb{S}^2$ be any (Riemannian) closed two-surface that is topologically a two-sphere. Latin letters from the middle third of the alphabet in Fraktur font ($\mathfrak{i}$, $\mathfrak{j}$, $\mathfrak{k}$, ...) are reserved for indices of tensors in $\mathscr{T}^k{}_l(\mathscr{S})$. In particular, for $\mathbb{S}^2$ itself, $\mathfrak{S}_{\mathfrak{ij}}$ is the metric, $\mathfrak{D}_{\mathfrak{i}}$ the associated derivative operator, and $\epsilon^{\mathbb{S}^2}_{\mathfrak{ij}}$ the volume form; in standard spherical coordinates $\{\theta, \phi\}$, the latter is simply given by

$$\boldsymbol{\epsilon}_{\mathbb{S}^2} = \sin\theta \, \mathrm{d}\theta \wedge \mathrm{d}\phi \,. \tag{0.0.3}$$

Contractions are indicated in the usual way in the abstract index notation: e.g., $U^a V_a$ is the contraction of $\boldsymbol{U}$ and $\boldsymbol{V}$. Equivalently, when applicable, we may simply use the "dot product" in the index-free notation, e.g. $U^a V_a = \boldsymbol{U} \cdot \boldsymbol{V}$, $A_{ab} B^{ab} = \boldsymbol{A} : \boldsymbol{B}$, etc. We must keep in mind that such contractions are to be performed using the metric of the space on which the relevant tensors are defined. Additionally, often we find it convenient to denote the component (projection) of a tensor in a certain direction by simply replacing its pertinent abstract index therewith: e.g., we equivalently write $U^a V_b = \boldsymbol{U} \cdot \boldsymbol{V} = U_{\boldsymbol{V}} = V_{\boldsymbol{U}}$, $A_{ab} U^a = A_{\boldsymbol{U} b}$, $A_{ab} U^a V^b = A_{\boldsymbol{UV}}$, etc. For any $(0, 2)$-tensor A_{ab}, we usually write its trace (in non-boldface) as $A = A_a{}^a = \mathrm{tr}(\boldsymbol{A})$.

Finally, let $\mathscr{U}$ and $\mathscr{V}$ be any two diffeomorphic manifolds and let $f : \mathscr{U} \to \mathscr{V}$ be a map between them. This naturally defines a map between tensors on the two manifolds, which we denote by $f_* : \mathscr{T}^k{}_l(\mathscr{U}) \to \mathscr{T}^k{}_l(\mathscr{V})$ and its inverse $(f^{-1})_* = f^* : \mathscr{T}^k{}_l(\mathscr{V}) \to \mathscr{T}^k{}_l(\mathscr{U})$. We generically refer to any map of this sort as a tensor transport [Felsager 2012]. It is simply the generalization to arbitrary tensors of the pushforward $f_* : T\mathscr{U} \to T\mathscr{V}$ and pullback $f^* : T^*\mathscr{V} \to T^*\mathscr{U}$, the action of which is defined in the standard way—see, e.g., Appendix C of Ref. [Wald 1984]. (Note that here the convention of sub-/super-scripting the star is the generally more common one used in geometry [Felsager 2012; Lee 2002]; it is sometimes opposite to and sometimes congruous with that used in the physics literature, e.g. [Wald 1984] and [Carroll 2003] respectively).

List of Figures

1.1 Detail of Kepler's model of Solar System motion based on Pythagorean solids, taken from [Koestler 1959] (adapted from *Mysterium Cosmographicum* [Kepler 1596]). A property of all Pythegorean solids, of which five exist, is that they can be exactly inscribed into—as well as circumscribed around—spheres. As only six planets were then known (from Mercury to Jupiter), this seemed to leave room for placing exactly these five perfect solids between their orbits (determined as an appropriate cross-section through the inscribing/circumscribing spheres). This figure shows the orbits of the planets up to Mars inclusive. 8

1.2 A figure of an ellipse (dotted oval) circumscribed by a circle from *Astronomia Nova* [Kepler 1609]. 8

1.3 Newton's solution of the two-body problem in the *Principia*. Extracted here are the figure used for his Proposition LVIII , as well as his Corollary 1 to this proposition [Newton 1687]: "Hence two bodies attracting each other with forces proportional to their distance, describe (by Prop. X), both round their common centre of gravity, and round each other, concentric ellipses; and, conversely, if such figures are described, the forces are proportional to the distances." [Chandrasekhar 2003] 11

1.4 Sketch of an extreme-mass-ratio inspiral (EMRI), a two-body system consisting of a stellar-mass compact object (SCO), usually a stellar-mass black hole, of mass $m \sim 10^{0-2} M_{\odot}$, orbiting and eventually spiralling into a (super-) massive black hole (MBH), of mass $M \sim 10^{4-7} M_{\odot}$, and emitting gravitational waves in the process. 21

1.5 The main approaches used in practice for the modeling of compact object binaries as a function of the mass ratio (increasing from 1) and the spacetime curvature involved. For low curvature (high separation between the bodies), post-Newtonian and post-Minkowskian methods are used. For high curvature (low separation) and low mass ratio, numerical relativity is used. For high curvature and extreme mass ratios, as the scale of a numerical grid would have to span orders of magnitude thus rendering it impracticable, perturbation theory must be used—in particular, self-force methods. 22

1.6 A depiction of the perturbative problem for the gravitational self-force (GSF). In particular, this represents one of the most popular conceptions of a so-called "self-consistent" approach [Gralla and Wald 2008]: at a given step (on a given Cauchy surface) in the time evolution problem, one computes the "correction to the motion" away from geodesic ($\mathring{\mathscr{C}}$) in the form of a deviation vector Z^a, determined by the GSF. Then, at the next time step, one begins on a new ("corrected") geodesic ($\mathring{\mathscr{C}}'$), computes the new deviation vector, and so on. 25

2.1 A compact region $\mathscr{V}$ in a spacetime $\mathscr{M}$ where the variation of physical fields are non-zero. 35

2.2 A depiction (in $(2+1)$ dimensions) of the foliation of a spacetime $(\mathscr{M}, \boldsymbol{g})$ into Cauchy surfaces $(\Sigma, \boldsymbol{h})$, where $\boldsymbol{h}$ is the metric induced on Σ by $\boldsymbol{g}$. These surfaces are defined by the constancy of a time function, $t(x^a) = \text{const.}$, which uniquely determine a normal vector $\boldsymbol{n}$. Additionally, one must define a time flow vector field $\boldsymbol{t}$ on $\mathscr{M}$ the integral curves of which intersect the"same spatial point" (with the same coordinates x^i) on different slices. 39

2.3 A depiction (in $(1+1)$ dimensions) of the spacetime $\mathscr{M}$ as constituted by a family of embedded submanifolds Σ_t obtained from an embedding map $i_t : \Sigma \to \mathscr{M}$. Three such submanifolds are shown at three different times, with the time flow vector field identifying the spatial coordinates between them. 40

2.4 An illustration of the transformation $f : T\mathscr{Q} \to T^*\mathscr{Q}$ from the configuration space tangent bundle into the phase space $\mathscr{P} = T^*\mathscr{Q}$. For example, suppose we have a two-dimensional configuration $\varphi = \{\varphi^1, \varphi^2\}$ (visually represented as one dimension), and correspondingly $\dot{\varphi} = \{\dot{\varphi}^1, \dot{\varphi}^2\}$. Suppose that here, $\dot{\varphi}^2$ is in the kernel of this map, *i.e.* it maps trivially to π_2 such that the only primary constraint is $0 = \pi_2 = \tilde{\zeta}$. The primary constraint surface $\tilde{\mathscr{C}}$ thus has coordinates $\{\varphi, \pi_1\}$. 49

2.5 An illustration of the configuration space tangent bundle $T\mathscr{Q}$, the phase space $\mathscr{P} = T^*\mathscr{Q}$, the primary constraint surface $\tilde{\mathscr{C}}$, and the full constraint surface $\mathscr{C}$ with their respective coordinates, and the maps/operations that respectively transform from one of these spaces to the next. 50

2.6 A visual representation of the "gauge freedoms" of GR. The embedding $i_t : \Sigma \to \mathscr{M}$ is shown in blue, along with the transformations on this embedding permitted by the constraints, shown in different colours. In particular, the primary constraints imply that we can change i_t to a new embedding $\tilde{i}_t$, shown in red, resulting from a change in the time flow vector field $\boldsymbol{t} \mapsto \tilde{\boldsymbol{t}}$ (or equivalently, $(N, \boldsymbol{N}) \mapsto (\tilde{N}, \tilde{\boldsymbol{N}})$). The embedded surface itself does not change, *i.e.* $\tilde{\Sigma}_t = \tilde{i}_t(\Sigma) = i_t(\Sigma) = \Sigma_t$, but the identification of spatial coordinates on sequential Cauchy surfaces in the family of embeddings does. On the other hand, the Hamiltonian constraint implies the freedom to change from i_t to $i'_{t'}$, shown in green, which is a change of foliation, or time function redefinition $t(x^a) \mapsto t'(x^a)$, such that $\Sigma'_{t'} = i'_{t'}(\Sigma)$ does not coincide with $\Sigma_t = i_t(\Sigma)$. Finally, the momentum constraint implies the freedom to map the spatial metric $\boldsymbol{h}$ in Σ by a diffeomorphism ϕ, $\boldsymbol{h} \mapsto \phi_* \boldsymbol{h}$. 59

3.1 Representation of a one-parameter family of spacetimes $\{\mathscr{M}_{(\lambda)}\}_{\lambda \geq 0}$ used for perturbation theory. Each of the $\mathscr{M}_{(\lambda)}$ are depicted visually in $(1+1)$ dimensions, as leaves of a (five-dimensional) product manifold $\mathscr{N} = \mathscr{M}_{(\lambda)} \times \mathbb{R}$, with the coordinate $\lambda \geq 0$ representing the perturbative expansion parameter. A choice of a map (or gauge) $\varphi_{(\lambda)} : \mathring{\mathscr{M}} \to \mathscr{M}_{(\lambda)}$, the flow of which is defined by the integral curves of a vector field $\mathbf{X} \in T\mathscr{N}$, gives us a way of identifying any point $p \in \mathring{\mathscr{M}} = \mathscr{M}_{(0)}$ on the "background" to one on some "perturbed" ($\lambda > 0$) spacetime, i.e. $p \mapsto \varphi_{(\lambda)}(p)$. 77

3.2 A gauge transformation consists in choosing a different vector field in $T\mathscr{N}$, or equivalently a different associated diffeomorphism, for identifying points between the background and the perturbed spacetimes. In this illustration, the point $p \in \mathring{\mathscr{M}}$ is mapped under the flow of $\mathbf{X}$ to the same point in $\mathscr{M}_{(\lambda)}$ as is $q \in \mathring{\mathscr{M}}$ under the flow of $\mathbf{Y}$ (for $p \neq q$ and $\mathbf{X} \neq \mathbf{Y}$). One thus has a gauge transformation on the background $q \mapsto \Psi_{(\lambda)}(q) = p$. 81

4.1 The idea of the proof for the Poincaré recurrence theorem. 103

4.2 The idea of the perturbative approach is to evaluate $\dot{S}$ along different directions in phase space away from equilibrium, and arrive at a contradiction with its strict positivity. 108

4.3 The topological approach relies on phase space compactness and Liouville's theorem, *i.e.* the fact that the Hamiltonian flow is volume-preserving. 114

5.1 A worldtube boundary $\mathscr{B}$ (topologically $\mathbb{R} \times \mathbb{S}^2$) in $\mathscr{M}$, with (outward-directed) unit normal n^a. The change in matter four-momentum between two constant time slices of this worldtube is given by the flux of the normal projection (in one index) of the matter stress-energy-momentum tensor T_{ab} through the portion of $\mathscr{B}$ bounded thereby. 137

5.2 A portion of a quasilocal frame $(\mathscr{B}; \boldsymbol{u})$ in a spacetime $\mathscr{M}$, bounded by constant t two-surfaces $\mathscr{S}_{\rm i}$ and $\mathscr{S}_{\rm f}$. In particular, $\mathscr{B} \simeq \mathbb{R} \times \mathbb{S}^2$ is the union of all integral curves (two-parameter family of timelike worldlines), depicted in the figure as dotted red lines, of the vector field $\boldsymbol{u} \in T\mathscr{B}$ which represents the unit four-velocity of quasilocal observers making up the congruence. The unit normal to $\mathscr{B}$ (in $\mathscr{M}$) is $\boldsymbol{n}$ and the normal to each constant t slice $\mathscr{S}$ of $\mathscr{B}$ is $\tilde{\boldsymbol{u}}$ (not necessarily coincidental with $\boldsymbol{u}$). Finally, $\mathscr{H}$ (with induced metric $\boldsymbol{\sigma}$) is the two-dimensional subspace of $T\mathscr{B}$ consisting of the "spatial" vectors orthogonal to $\boldsymbol{u}$. Note that unlike $\mathscr{S}$, $\mathscr{H}$ need not be integrable (indicated in the figure by the failure of $\mathscr{H}$ to make a closed two-surface). 142

5.3 Representation of a one-parameter family of QFs $\{(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})\}_{\lambda\geq 0}$ embedded correspondingly in a family of spacetimes $\{\mathscr{M}_{(\lambda)}\}_{\lambda\geq 0}$. 150

5.4 Representation of the Gralla-Wald family of spacetimes $\{\mathscr{M}_{(\lambda)}\}_{\lambda\geq 0}$. (This is an adaptation of Fig. 1 of [Gralla and Wald 2008].) The lined green region that "fills in" $\mathscr{M}_{(\lambda)}$ for $r \leq C\lambda$ is the "small" object which "scales down" to zero "size" and "mass" in the background $\mathring{\mathscr{M}}$. The solid black lines represent taking the "ordinary limit" (the "far away" view where the motion appears reduced to a worldline) and the dashed black lines the "scaled limit" (the "close by" view where the rest of the universe appears "pushed away" to infinity). The worldline $\mathring{\mathscr{C}}$, which can be proven to be a geodesic, is parametrized by $\mathring{z}^a(\mathring{\tau})$ and has four-velocity $\mathring{\boldsymbol{U}}$. The "deviation" vector $\boldsymbol{Z}$ on $\mathring{\mathscr{C}}$ is used for formulating the first-order correction to the motion. 157

5.5 A family of rigid quasilocal frames $\{(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})\}$ embedded in the Gralla-Wald family of spacetimes $\{\mathscr{M}_{(\lambda)}\}$ such that the inverse image of any such perturbed quasilocal frame in the background is inertial with the "point particle" approximation of the moving object, *i.e.* is centered on the geodesic $\mathring{\mathscr{C}}$. 166

5.6 An instantaneous rigid quasilocal frame $(\mathbb{S}^2_r, r^2\boldsymbol{\mathfrak{S}}, \boldsymbol{\mathfrak{D}})$ (where $\boldsymbol{\mathfrak{S}}$ and $\boldsymbol{\mathfrak{D}}$ respectively are the metric and derivative compatible with the unit two-sphere) inertial with the background "point particle". This means that the latter is located at the center of our Fermi normal coordinate system. 167

5.7 A family of rigid quasilocal frames $\{(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})\}$ embedded in the Gralla-Wald family of spacetimes $\{\mathscr{M}_{(\lambda)}\}$ inertial with the moving object in $\mathscr{M}_{(\lambda)}$. This means that $\mathscr{B}_{(\lambda)}$ is defined by the constancy of the Gralla-Wald r coordinate in $\mathscr{M}_{(\lambda)}$, for any $r > C\lambda$. Thus, the inverse image $\mathscr{B}$ of $\mathscr{B}_{(\lambda)}$ in the background $\mathring{\mathscr{M}}$ is centered, in general, *not* on the geodesic $\mathring{\mathscr{C}}$ followed by the "point particle" background approximation of the object, but on some timelike worldline $\mathscr{C} \neq \mathring{\mathscr{C}}$, with four-velocity $\boldsymbol{U} \neq \mathring{\boldsymbol{U}}$, which may be regarded as an approximation on $\mathring{\mathscr{M}}$ of the "true motion" of the object in $\mathscr{M}_{(\lambda)}$. Between $\mathring{\mathscr{C}}$ and $\mathscr{C}$ there is a deviation vector $\boldsymbol{Z}$, which can be identified with the deviation vector ("correction to the motion") in the Gralla-Wald approach. 170

5.8 An instantaneous rigid quasilocal frame $(\mathbb{S}^2_r, r^2\boldsymbol{\mathfrak{S}}, \boldsymbol{\mathfrak{D}})$ (where $\boldsymbol{\mathfrak{S}}$ and $\boldsymbol{\mathfrak{D}}$ respectively are the metric and derivative compatible with the unit two-sphere) inertial with the moving object in the perturbed spacetime. This means that the "point particle" approximation of this object in the background spacetime is *not* located at the center of our Fermi normal coordinate system. Instead, it is displaced in some direction ρ^I, which must be $\mathcal{O}(\lambda)$. 171

6.1 Schematic representation of the PwP formulation. The field equations with singular source terms become homogeneous equations at each side of the particle worldline together with a set of jump conditions to communicate their solutions. 187

A.1 An illustration of two manifolds $(\mathscr{M}, \boldsymbol{g})$ and $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ with a map $\phi : \mathscr{M} \to \tilde{\mathscr{M}}$ between them. This identifies any point $p \in \mathscr{M}$ with $\tilde{p} = \phi(p) \in \tilde{\mathscr{M}}$, and can be used, for example, to push-forward the vector $\boldsymbol{v} \in T_p\mathscr{M}$ to $\phi_* \boldsymbol{v} \in T_{\tilde{p}}\tilde{\mathscr{M}}$. If ϕ is a diffeomorphism, a general transport of tensors from one manifold to the other can be defined. Note that in this notation, the metric $\tilde{\boldsymbol{g}}$ of $\tilde{\mathscr{M}}$ is not necessarily the same as the metric transported from $\mathscr{M}$, i.e. $\phi_* \boldsymbol{g}$. If indeed $\phi_* \boldsymbol{g} = \tilde{\boldsymbol{g}}$, then ϕ is called an *isometry*—a symmetry of the metric. 204

A.2 An illustration of the meaning of the Lie derivative along a vector field $\boldsymbol{v}$ in $(\mathscr{M}, \boldsymbol{g})$ of a tensor $\boldsymbol{A}$, depicted here for ease of visualization in the case where $\boldsymbol{A}$ is a vector. In particular, one compares $\boldsymbol{A}_0 = (\boldsymbol{A})_{p_0}$ at the point p_0 corresponding to $t = 0$ in the flow $\phi_t^{(\boldsymbol{v})}$ with its value $\boldsymbol{A}_t = (\boldsymbol{A})_{p_t}$ at the point $p_t = \phi_t^{(\boldsymbol{v})}(p_0)$ for some $t > 0$ by transporting the latter back to p_0, i.e. one compares $\boldsymbol{A}_0$ with $(\phi_{-t}^{(\boldsymbol{v})})_* \boldsymbol{A}$ at p_0. Their difference divided by t, in the small t limit, is the Lie derivative. 208

B.1 Solution of the problem (B.4.1) with zero initial data. 237

B.2 Convergence of the numerical scheme for the problem (B.4.1). 238

B.3 Solution of (B.4.4) with (normalized) Gaussian initial data centered at $V = 0.3$. 239

B.4 Convergence of the numerical scheme for the problem (B.4.4). 240

B.5 Solution of the problem (B.5.1) with zero initial data. 242

B.6 Convergence of the numerical scheme for the problem (B.5.1). 243

B.7 Solution of the problem (B.5.5). 246

B.8 Price evolution and convergence of the numerical scheme for the problem (B.5.5). 247

B.9 Solution and convergence of the numerical scheme for the problem (B.6.1). 249

B.10 Solution (with $N = 80$) of the same problem as in Figure B.9 but using $\omega = 24$. 250

B.11 Convergence of the pseudospectral numerical scheme for the problem (B.7.1). 253

Acknowledgements

I am indebted most deeply to the guidance of my two advisors, Carlos F. Sopuerta and Alessandro Spallicci. Their patience and openness extend as far as any student could ask, and their knowledge of physics has always been a great pleasure from which to learn. I am especially grateful for the freedom they have allowed me in the topics pursued for this work. I thank them also for their invaluable help in its redaction at all stages, including Alessandro with the translations to French, and more generally in dealing with all the adventures of organizing a joint doctoral degree between Barcelona and Orléans.

A special thanks goes to my collaborator and former advisor, Robert Mann. Not that many years ago I took my first steps into gravitational physics research thanks to him, and I am deeply grateful for his continued involvement and guidance ever since.

I also greatly thank my collaborator Richard Epp, who was the first person to truly show me, as he has continued to do, that behind the arcane mathematics of relativity there is always an intuitive physical picture to tell the story. An extra thanks is due to my former colleague Paul McGrath, whose initial work on quasilocal frames during his own doctorate laid the foundation for a good part of ours here.

The places and people which have hosted me during the carrying out of this thesis have truly given me a sense of home. I thank immensely the colleagues and friends I have made over the past years here in Barcelona. I cannot imagine this time without Eric Brown, Angelo Gambino, Marcela Gus, Rafael Murrieta and Aurélien Severino. Graciès especialment al meu estimat amic Xavi Viader, inclòs amb les traduccions aquí al català. It has been a pleasure to work in the ICE-IEEC gravitational waves group with Pau Amaro, Lluís Gesa, Ferran Gibert, Jordina Ho Zhang, Ivan Lloro, Juan Pedro López, Víctor Martín, Miquel Nofrarias, Francisco Rivas, and Daniel Santos, to whom I owe a special thanks for initially helping me to settle. From the cosmology group here, I thank Enrique Gaztañaga as well as my good friend Esteban González. I also convey my very warm gratitude to my friends and colleagues at ESADE Business School, where I have been happy to alternate a bit from the completion of this work to some teaching; I am especially grateful to Núria Agell, Gerard Gràcia, Jordi Montserrat, Xari Rovira and Marc Torrens.

From my first year of this thesis spent in Orléans, I thank my former colleague and collaborator Luca Bonetti, and my friends Pratik Hardikar and Saketh Sistla. I would be remiss not to also add my friends and former colleagues from my earlier time in Canada prior to this thesis, especially Adam Bognat, Farley Charles, Heiko Santosh, Billy Tran and Anson Wong.

It has been a great privilege during this doctorate to have the occasion to travel and interact with many scientists outside of my home institutions. For interesting discussions and encouragements, I thank in particular Abraham Harte, Ulrich Sperhake, Helvi Witek and Miguel Zilhão. For many pleasant and always instructive visits for workshops and summer schools over the past years, I am particularly grateful to the hospitality of the Pedro Pascual Science Center in Benasque, Spain.

It is also a pleasure to thank my friend and former colleague Hossein Bazrafshan Moghaddam. Our conversations about physics always teach me something new. I thank him also for helping me organize a wonderful visit to present some of this work at IPM Tehran—additional thanks also to the hospitality there of Hassan Firouzjahi—and the University of Mashhad, Iran.

A dedication goes to the memory of Milton B. Zysman (1936-2019), polymath and friend, who taught me much about the history of science and whose influence towards inquisitive skepticism is, I hope, well alive in the pages that follow.

I am deeply thankful for the support of my family. Vă mulţumesc la toţi, în special bunicii, părinţii, şi sora mea Andreea. Nagyon köszönöm jó barátom, Radu Stănilă.

Financial support for the realization of this thesis was provided by the Natural Sciences and Engineering Research Council of Canada through a Postgraduate Scholarship - Doctoral, Application No. PGSD3 - 475015 - 2015; by Campus France through an Eiffel Bourse d'Excellence, Grant No. 840856D, awarded for carrying out an international joint doctorate; by LISA CNES funding; and by the Ministry of Economy and Business of Spain (MINECO) through contracts ESP2013-47637-P, ESP2015-67234-P and ESP2017-90084-P.

To clear the way leading from theory to experiment of unnecessary and artificial assumptions, to embrace an ever-wider region of facts, we must make the chain longer and longer. The simpler and more fundamental our assumptions become, the more intricate is our mathematical tool of reasoning; the way from theory to observation becomes longer, more subtle, and more complicated. Although it sounds paradoxical, we could say: Modern physics is simpler than the old physics and seems, therefore, more difficult and intricate.

[Einstein and Infeld 1938]

Part I

Fundamentals of General Relativity:

Introduction, Canonical Formulation and Perturbation Theory

CHAPTER 1

Introduction

Chapter summary. In this introduction, we present a brief history of the gravitational two-body problem and of the conception of gravitation in physics more generally, as well as a discussion of the current relevance of this problem—focusing on the extreme-mass-ratio-regime—in the era of gravitational wave astronomy.

We begin in Section 1.1 with a historical discussion of the gravitational two-body problem in pre-relativistic physics. Newton's work, especially the *Principia*, is undeniably regarded as constituting the first true solution to this problem. We discuss its relevance, including Newton's own views on gravity, as well as the path immediately leading to it, especially the work of Kepler.

In Section 1.2, we provide an account of the development of general relativity, our contemporary theory of gravity, including extracts from Einstein's own papers summarizing the essential content of the theory. With this occasion, we define and establish the notation we use in this thesis for the most basic mathematical objects.

In Section 1.3, we then discuss the interpretation of general relativity, and especially Einstein's views. Instead of the general idea of "gravity as geometry", an interpretation he seems to have found rather uninteresting due to its generality, he was much more fascinated with the connection between gravity and inertia, in particular, as established through the equation of motion for idealized particles, the geodesic equation.

This leads us, in Section 1.4, to a discussion of the current relevance of the problem of motion in general relativity thanks to the opportunities presented by the advent of gravitational-wave astronomy. In particular, we focus on systems called extreme-mass-ratio inspirals (EMRIs): these are compact binary systems where one object is much less massive than—thus effectively orbiting and eventually spiraling into—the other. Usually, the latter is a (super-) massive black hole at a galactic center, and the former is a stellar-mass black hole or a neutron star. It is anticipated that these will be one of the main sources for space-based gravitational wave detectors, specifically for the LISA mission expected to launch in the 2030s.

Finally, in Section 1.5, we enter into a bit of detail on the technical problem of modeling EMRIs. This involves calculating the correction to the motion, away from geodesic, caused by the backreaction of (the mass of) the orbiting object upon the gravitational field. This phenomenon is known as the gravitational self-force, and will be one of the major themes of this thesis.

Introducció (chapter summary translation in Catalan). En aquesta introducció, presentem una breu història del problema gravitatori de dos cossos i de la concepció de la gravitació en física més generalment, així com una discussió de la rellevància actual d'aquest problema - centrat en el règim de raons de masses extremes - en l'era de l'astronomia de les ones gravitacionals.

Comencem a la secció 1.1 amb una discussió històrica del problema gravitatori de dos cossos en física pre-relativista. L'obra de Newton, especialment els *Principia*, és considerada la primera veritable solució a aquest problema. Es discuteix la seva rellevància, incloent les opinions pròpies de Newton sobre la gravetat, així com el camí que hi dirigeix directament, especialment el treball de Kepler.

A la secció 1.2, exposem el desenvolupament de la relativitat general, la nostra teoria contemporània de la gravetat, inclosos extractes dels propis treballs d'Einstein que resumeixen el contingut essencial de la teoria. Amb aquesta ocasió, definim i establim la notació que fem servir en aquesta tesi per als objectes matemàtics més bàsics.

A la secció 1.3, es discuteix la interpretació de la relativitat general, i especialment les opinions d'Einstein. Enlloc de la idea general de la "gravetat com a geometria", una interpretació que sembla haver trobat poc interessant per la seva generalitat, estava molt més fascinat per la connexió entre la gravetat i la inèrcia, en particular, com es va establir mitjançant l'equació del moviment per partícules idealitzades, l'equació geodèsica.

Això ens porta, a la secció 1.4, a una discussió sobre la rellevància actual del problema del moviment en la relativitat general gràcies a les oportunitats que presenta l'arribada de l'astronomia d'ones gravitacionals. En particular, ens centrem en caigudes en espiral amb raó de masses extrema (*extreme-mass-ratio inspirals*, EMRIs): es tracta de sistemes binaris compactes on un objecte és molt menys massiu que - de manera que orbita i, finalment, fa espirals cap a - l'altre. Normalment, aquest últim és un forat negre (super) massiu en un centre galàctic, i el primer és un forat negre de massa estel·lar o una estrella de neutrons. Es preveu que aquesta sigui una de les principals fonts per als detectors d'ones gravitacionals basades en l'espai, en particular per a la missió LISA que es preveu llançar a la dècada dels 2030.

Finalment, a la secció 1.5, introduïm una mica de detall sobre el problema tècnic de modelar EMRIs. En particular, es tracta de calcular la correcció al moviment, allunyada del geodèsic, causada per la retroacció de (la massa de) l'objecte orbitant sobre el camp gravitatori. Aquest fenomen es coneix com la força pròpia gravitacional i serà un dels principals temes d'aquesta tesi.

Introduction (chapter summary translation in French). Dans cette introduction, nous présentons un bref historique du problème gravitationnel à deux corps et de la conception plus générale de la gravitation dans la physique, ainsi qu'une discussion sur la pertinence actuelle de ce problème - en se concentrant sur le régime des quotients extrêmes des masses - dans l'ère de l'astronomie des ondes gravitationnelles.

Nous commençons à la section 1.1 avec une discussion historique sur le problème gravitationnel à deux corps dans la physique pré-relativiste. Les travaux de Newton, en particulier le *Principia*, sont indéniablement considérés comme constituant la première véritable solution à ce problème. Nous discutons de sa pertinence, y compris de la propre vision de Newton sur la gravité, ainsi que du chemin qui y conduit immédiatement, en particulier les travaux de Kepler.

Dans la section 1.2, nous décrivons l'évolution de la relativité générale, notre théorie contemporaine de la gravité, avec des extraits d'articles d'Einstein résumant le contenu essentiel de la théorie. À cette occasion, nous définissons et établissons la notation que nous utilisons dans cette thèse pour les objets mathématiques les plus fondamentaux.

Dans la section 1.3, nous discutons ensuite de l'interprétation de la relativité générale et en particulier des points de vue d'Einstein. Au lieu de l'idée générale de la "gravité en tant que géométrie", une interprétation qu'il semble avoir trouvée pas assez inintéressante en raison de sa généralité, il était beaucoup plus fasciné par le lien entre la gravité et l'inertie, en particulier, établi par l'équation du mouvement de particules idéalisées, l'équation géodésique.

Ceci nous amène, dans la section 1.4, à une discussion sur la pertinence actuelle du problème du mouvement en relativité générale, grâce aux possibilités offertes par l'avènement de l'astronomie des ondes gravitationnelles. En particulier, nous nous concentrons sur les systèmes appelés inspirals avec quotients extrêmes des masses (*extreme-mass-ratio inspirals*, *EMRIs* en italique car mot non français dans le contexte ; encore une fois, voire si non déjà dit) : il s'agit de systèmes binaires compacts où un objet est beaucoup moins massif que - ce qui permet effectivement une orbite et au final en spirallant dans l'autre. Habituellement, le dernier est un trou noir (super) massif à un centre galactique et le premier est un trou noir à masse stellaire ou une étoile à neutrons. On s'attend à ce qu'ils soient l'une des principales sources de détecteurs d'ondes gravitationnelles situés dans l'espace, en particulier pour la mission LISA qui devrait être lancée dans les années 2030.

Enfin, dans la section 1.5, nous entrons dans les détails sur le problème technique de la modélisation des *EMRIs*. En particulier, il s'agit de calculer la correction du mouvement, loin de la géodésique, provoquée par la réaction en arrière (de la masse) de l'objet en orbite dans le champ gravitationnel. Ce phénomène est connu sous le nom de la force propre gravitationnelle et il constituera l'un des thèmes majeurs de cette thèse.

1.1. *Gravitatio mundi*, a brief historical prelude

The gravitational two-body problem, in its broadest form, has always occupied a role apart in the historical development of physics, astronomy, mathematics, and even philosophy: *How does one massive object move around another, and why that particular motion?* From labyrinthine epicycles, to Keplerian orbits, to the notion of a universal gravitational "force" and beyond, the centuries-old struggle to tackle this question directly precipitated—more so, arguably, than any other single physical problem—the emergence of modern scientific thought around the turn of the 18th century. Up to the present day, with vast opportunities currently presented by the revolutionary expansion of observational astronomy into the domain of gravitational waves, understanding and solving this problem has remained as galvanizing an incentive as ever for both technical as well as conceptual advances.

From our contemporary point of view, the two parts of the problem as formulated above—on the one hand, the empirical question of *how* motion occurs in a gravitational two-body system, and on the other hand, the theoretical question of *why* it is that (rather than any other conceivable) motion—are indisputably regarded as having reached their first true synthesis in the work of Newton, above all in the *Principia* [Newton 1687][1]. Certainly, hardly any of Newton's preeminent predecessors, from the ancient Greeks to the astronomers of the Renaissance, fell short of taking an avid interest in not only *how* the Moon and the planets moved, but *why* they moved so—or, perhaps offering a better sense of the epochal mindset, "*what*" moved them so. Still, pre-Newtonian "explanations" of heavenly mechanics generally appear to us today to rest rather closer to the realm of myth than to that of scientific theory.

The figure which stood at the point of highest inflection in the evolution of the intellectual mentality towards answering this latter, theoretical type of question was at the same time one of the greatest empiricists and mystics—Johannes Kepler (1571-1630). A restlessly contradictory character throughout his life, we can glean a brief sense of the dramatic psychological fluxes that marked it—and therethrough, ultimately, his entire era—by simply recalling Kepler's two most famous theoretical models for Solar System motion [Koestler 1959]. When he was in his mid-20s, he developed in a book called *Mysterium Cosmographicum* [Kepler 1596] a model in which the orbits of the planets around the Sun are determined by a particular embedding of Pythagorean solids[2] centered thereon, see Fig. 1.1. Then, a little over a decade later, in *Astronomia Nova* [Kepler 1609], he put forth an

[1] For an English translation with excellent accompanying commentary by Chandrasekhar for today's "common reader", see [Chandrasekhar 2003].

[2] Also known as Platonic solids, or perfect solids, these are the set of three-dimensional solids with identical faces (regular, convex polyhedra). It was shown by Euclid that only five such solids exist. They are [Koestler 1959]:

empirical model of elliptical orbits, based on the observations of Tycho Brahe, establishing what we nowadays refer to as Kepler's laws of planetary motion[3]. See Fig 1.2. What may be called the (neo-) Platonic basis of "explanation" underlying the former stands, to the modern reader, in radically sharp contrast with the manifestly quasi-mechanistic one at the basis the latter. This reasoning is brought by Kepler to its logical end in a letter to Herwart, which he wrote as *Astronomia Nova* was nearing completion (taken from [Koestler 1959]):

> My aim is to show that the heavenly machine is not a kind of divine, live being, but a kind of clockwork [...] insofar as nearly all the manifold motions are caused by a most simple [...] and material force, just as all motions of the clock are caused by a simple weight. And I also show how these physical causes are to be given numerical and geometrical expression.

One discerns in these lines an approach towards the sort of thinking that ultimately led to the paradigmatic Newtonian explanation of the elliptical shapes of the planetary orbits.

Arthur Koestler, in his authoritative history of pre-Newtonian cosmology *The Sleepwalkers* [Koestler 1959], to which we have referred so far a few times, traces out in great detail the work of Kepler and especially his "giving of the laws" of planetary motion. He summarizes their significance:

> Some of the greatest discoveries [...] consist mainly in the clearing away of psychological road-blocks which obstruct the approach to reality; which is why, *post factum*, they appear so obvious. In a letter to Longomontanus[33] Kepler qualified his own achievement as the "cleansing of the Augean stables".

(1) the tetrahedron (pyramid) bounded by four equilateral triangles; (2) the cube; (3) the octahedron (eight equilateral triangles); (4) the dodecahedron (twelve pentagons) and (5) the icosahedron (twenty equilateral triangles).

The Pythagoreans were fascinated with these, and associated four of them (1,2,3, and 5, in the above numbering) with the "elements" (fire, earth, air, and water, respectively) and the remaining one (4, the dodecahedron) with quintessence, the substance of heavenly bodies. The latter was considered dangerous, and so "[o]rdinary people were to be kept ignorant of the dodecahedron" [Sagan 1980].

[3] In fact, only the first two of what we today refer to as the three Keplerian laws of planetary motion were proposed in this work (the third he found a bit later): (1) the orbits of planets are ellipses with the Sun at a focus; (2) the planets move such that equal areas in the orbital plane are "swept out", by a straight line with the Sun, in equal time. It is interesting to remark that these were actually discovered in reverse order. For a detailed historical account, see Part Four, Chapter 6 of [Koestler 1959].

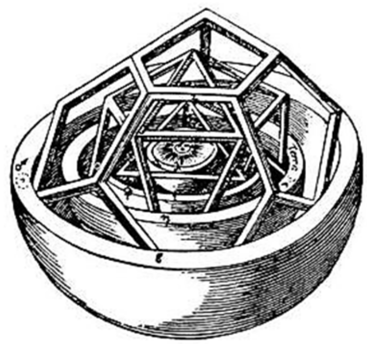

Figure 1.1. Detail of Kepler's model of Solar System motion based on Pythagorean solids, taken from [Koestler 1959] (adapted from *Mysterium Cosmographicum* [Kepler 1596]). A property of all Pythegorean solids, of which five exist, is that they can be exactly inscribed into—as well as circumscribed around—spheres. As only six planets were then known (from Mercury to Jupiter), this seemed to leave room for placing exactly these five perfect solids between their orbits (determined as an appropriate cross-section through the inscribing/circumscribing spheres). This figure shows the orbits of the planets up to Mars inclusive.

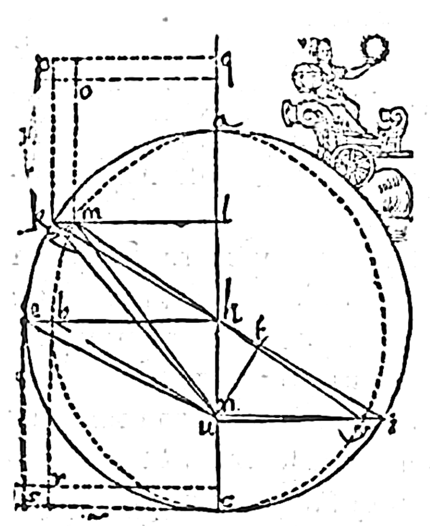

Figure 1.2. A figure of an ellipse (dotted oval) circumscribed by a circle from *Astronomia Nova* [Kepler 1609].

> But Kepler not only destroyed the antique edifice; he erected a new one in its place. His Laws are not of the type which appear self-evident, even in retrospect (as, say, the Law of Inertia appears to us); the elliptic orbits and the equations governing planetary velocities strike us as "constructions" rather than "discoveries". In fact, they make sense only in the light of Newtonian Mechanics. From Kepler's point of view, they did not make much sense; he saw no logical reason why the orbit should be an ellipse instead of an egg. Accordingly, he was more proud of his five perfect solids than of his Laws; and his contemporaries, including Galileo, were equally incapable of recognizing their significance. The Keplerian discoveries were not of the kind which are "in the air" of a period, and which are usually made by several people independently; they were quite exceptional one-man achievements. That is why the way he arrived at them is particularly interesting.

Nonetheless, the basic new concepts involved in articulating this new, clockwork-type worldview presented great conceptual difficulties. In the *Astronomia Nova*, for example, Kepler wrestled profusely with the concept of the "force" causing the motions in his imagined clockwork universe [Kepler 1609] (taken from [Koestler 1959]):

> This kind of force [...] cannot be regarded as something which expands into the space between its source and the movable body, but as something which the movable body receives out of the space which it occupies... It is propagated through the universe ... but it is nowhere received except where there is a movable body, such as a planet. The answer to this is: although the moving force has no substance, it is aimed at substance, i.e., at the planet-body to be moved...

Koestler remarks, interestingly, that Kepler's description above actually seems to be "closer to the modern notion of the gravitational or electro-magnetic *field* than to the classic Newtonian concept of *force*" [Koestler 1959].

With Newton's arrival on the scene, the vision of a mechanistic clockwork universe took definitive shape in the form of three laws of motion and the inverse-square law of universal gravitation—with Kepler's three laws recovered from these as particular consequences [Newton 1687]. Koestler contextualizes the relevance of this moment [Koestler 1959]:

> It is only by bringing into the open the inherent contradictions, and the metaphysical implications of Newtonian gravity, that one is able to realize the enormous courage – or sleepwalker's assurance – that was needed to use it as the basic concept of cosmology. In one of the most reckless and sweeping generalizations in the history of thought, Newton filled the entire space of the universe with interlocking forces of attraction, issuing from all particles of matter and acting on all particles of matter, across the boundless abysses of darkness.

> But in itself this replacement of the *anima mundi* by a *gravitatio mundi* would have remained a crank idea or a poet's cosmic dream; the crucial achievement was to express it in precise mathematical terms, and to demonstrate that the theory fitted the observed behaviour of the cosmic machinery – the moon's motion round the earth and the planets' motions round the sun.

Newton, of course, was famously aware of the "inherent contradictions" to which Koestler is referring. While comments to this effect appear in the *Principia* itself [Newton 1687], in a letter to Bentley just a few years later, he could not have been clearer *vis-à-vis* what he thought about his proposed theory—and in particular, the physical conception of gravitation offered by it (taken from [Koestler 1959]):

> It is inconceivable, that inanimate brute matter should, without the mediation of something else which is not material, operate upon and affect other matter without mutual contact, as it must be, if gravitation in the sense of Epicurus, be essential and inherent in it. And this is one reason why I desired you would not ascribe innate gravity to me. That gravity should be innate, inherent, and essential to matter, so that one body may act upon another at a distance through a vacuum, without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it. Gravity must be caused by an agent acting constantly according to certain laws; but whether this agent be material or immaterial, I have left open to the consideration of my readers.

No less difficult for the consideration of Newton's readers at that time was the new mathematics describing this metaphysically mysterious "agent". In fact, Newton notoriously avoided publishing his work on calculus—which he referred to as the "method of fluxions"—for decades, leading to the infamous controversy with Leibnitz over its discovery [Gleick 2004]. Meanwhile, the *Principia* [Newton 1687], though clearly bearing the basic elements of the infinitesimal analysis at the basis of calculus, was written essentially, one might say in "brute-force" style, in the technical language then commonly understood: Euclindean geometry. Newton presented his solution of the two-body problem—the proof of elliptical planetary motion as a consequence of his laws—in the *Principia*, Book I, Section XI, Propositions LVII-LXIII [Newton 1687]. See Fig. 1.3.

Soon afterward, the issue of perturbations to a two-body orbit from a third body, and more generally the question of the stability of the entire Solar System, quickly gained interest. Newton also raised this problem, and seems to have been doubtful about the possibility of long-term Solar System stability. Subsequent investigation into this issue went hand in hand with the development of perturbation theory, especially thanks to the work of Lagrange and Laplace. See [Laskar 2013] for a detailed review of the history and

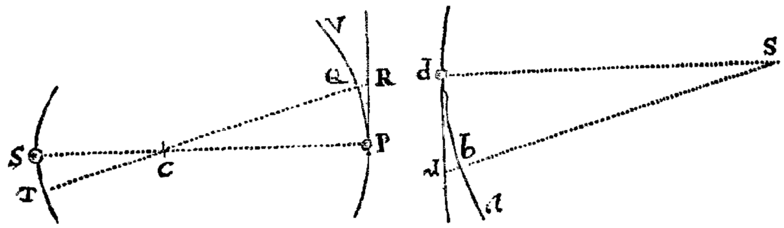
Corol. 1. Hinc corpora duo viribus diſtantiæ ſuæ proportionalibus ſe mutuo trahentia, deſcribunt (per Prop. X.) & circum commune gravitatis centrum, & circum ſe mutuo, Ellipſes concentricas: & vice verſa, ſi tales figuræ deſcribuntur, ſunt vires diſtantiæ proportionales.

FIGURE 1.3. Newton's solution of the two-body problem in the *Principia*. Extracted here are the figure used for his Proposition LVIII , as well as his Corollary 1 to this proposition [Newton 1687]: "Hence two bodies attracting each other with forces proportional to their distance, describe (by Prop. X), both round their common centre of gravity, and round each other, concentric ellipses; and, conversely, if such figures are described, the forces are proportional to the distances." [Chandrasekhar 2003]

the current status of this problem, including the discovery in the last few decades of chaos in Solar System dynamics.

1.2. The advent of relativity

While there certainly existed some known empirical discrepancies with Newton's theory by the end of the 19th century—among the most notable being, especially in view of the two-body problem, the perihelion precession of Mercury known since 1859—what primarily led to its overthrow had, at least in the vision of its chief perpetrator, much more to do with its eminently long-standing "inherent contradictions". Einstein, indeed, often regarded his development of relativity[4] as merely the proverbial cleansing of the Newtonian stables [Einstein 1954]:

> Let no one suppose [...] that the mighty work of Newton can really be superseded by [general relativity] or any other theory. His great and lucid ideas will retain their unique significance for all time as the foundation of our whole modern conceptual structure in the sphere of natural philosophy.

[4] In fact, Einstein wished to call it the "theory of invariance" (to highlight the invariance of the speed of light and that of physical laws in different reference frames), but the term "theory of relativity" coined by Max Planck and Max Abraham in 1906 quickly became, to Einstein's dissatisfaction, the more popular nomenclature and that used to this day [Galison et al. 2001].

There is a good deal of difference between the circumstances surrounding the emergence of general relativity compared with that of the Newtonian theory. While the latter went hand in hand with strong empirical contingencies—primary among these being, as we have seen, solving the two-body problem—the former was driven much more by basic conceptual and logical questions. Cornelius Lanczos, a mathematician contemporary with Einstein, comments [Lanczos 1949]:

> Einstein's Theory of General Relativity [...] was obtained by mathematical and philosophical speculation of the highest order. Here was a discovery made by a kind of reasoning that a positivist cannot fail to call "metaphysical," and yet it provided an insight into the heart of things that mere experimentation and sober registration of facts could never have revealed.

Viewed from such a standpoint, the local effects of special relativity—time dilation, length contraction and all the rest—as well as the globally curved (non-flat) geometry of the spacetime we inhabit can be regarded as following, essentially, as logical consequences from: *(i)* on the one hand, demanding consistency between the physical laws then known (in particular, as concerns the Maxwellian theory of electromagnetism), and *(ii)* on the other hand, dispensing with what appeared to be the most unnecessary assumptions causing the "inherent contradictions" of the Newtonian theory: in particular, the notion of absolute space, and connected with this, the formulation of physical laws in a privileged—the so-called inertial—class of coordinate reference frames. It is quite remarkable how what looks from this point of view as a sort of exercise in logic has ultimately produced such wonderfully diverse physical insights into the nature of gravity, and even—though this generally took longer to understand—the sorts of basic objects that can exist in our Universe, such as black holes and gravitational waves.

An issue that attracted much of Einstein's attention throughout his development of general relaitivity was that of the motion of an idealized "test" mass, that is, one provoking no backreaction in the field equations of the theory [Renn 2007; Lehmkuhl 2014]. Already in 1912, in a note added in proof to [Einstein 1912], he stated for the first time that the *geodesic equation*, that is, the extremization of curve length,

$$\delta \int \mathrm{d}s = 0\,, \tag{1.2.1}$$

is the equation of motion of point particles "not subject to external forces". In this case, $\mathrm{d}s$ is an infinitesimal distance element in any curved four-dimensional spacetime. By this point, Einstein understood that the basic mathematical methods for studying spacetime curvature, logically identified as gravity, were those of differential geometry pioneered during the previous century by Gauss, the Bolyais (Farkas and his son János), Lobachevsky, Riemann, Ricci and Levi-Civita, to name a few of the main players[5]. Thus the basic object, in a theory fundamentally concerned with length measurements (in the broadest sense), is the metric tensor, denoted $g_{\mu\nu}$ in Einstein's original notation. This object defines the notion of infinitesimal distance $\mathrm{d}s$, and hence also that of motion [Einstein 1913] (taken from [Lehmkuhl 2014]):

> A free mass point moves in a straight and uniform line according to [our Eq. (1.2.1)], where
>
> $$\mathrm{d}s^2 = \sum_{\mu\nu} g_{\mu\nu} \mathrm{d}x_\mu \mathrm{d}x_\nu \,. \tag{1.2.2}$$
>
> [...] In general, every gravitational field is going to be defined by ten components $g_{\mu\nu}$, which are functions of [local coordinates] x_1, x_2, x_3, x_4.

In 1914, Einstein actually used the word "geodesic" for the first time to refer to length-extremizing curves [Lehmkuhl 2014]. Then, in 1915, the theory was completed with the promulgation of the final form of the gravitational field equations governing his $g_{\mu\nu}$ [Einstein 1915] (English translation in [Einstein 1996a]). In a paper consolidating the theory the following year, Einstein summarizes the main ideas [Einstein 1916] (taken from [Einstein 1996b]):

> We make a distinction hereafter between "gravitational field" and "matter" in this way, that we denote everything but the gravitational field as "matter." Our use of the word therefore includes not only matter in the ordinary sense, but the electromagnetic field as well.
>
> [...] [We] require for the matter-free gravitational field that the symmetrical [Ricci tensor], derived from the [Riemann tensor], shall vanish. Thus we obtain ten equations for the ten quantities $g_{\mu\nu}$ [...].
>
> [...] It must be pointed out that there is only a minimum of arbitrariness in the choice of these equations. For besides [the Ricci tensor] there is no tensor of second rank which is formed from the $g_{\mu\nu}$ and its derivatives, contains no derivations higher than second, and is linear in these derivatives.*
>
> These equations, which proceed, by the method of pure mathematics, from the requirement of the general theory of relativity, give us, in combination with the equations of motion [our Eq. (1)], to a first approximation Newton's law of attraction, and to a second approximation the explanation of the motion of the perihelion of the planet Mercury discovered by Leverrier (as it remains after

[5] Much of the development of differential geometry had to do with attempts to prove Euclid's famous fifth postulate. Ever since the appearance of the *Elements*, which is based on five postulates, there had been skepticism regarding the necessity of the last of these. In its original form it is much more complicated to state than the first four, but it is equivalent to the statement that the sum of the three angles of a triangle is always equal to two right angles. The advent of differential ("non-Euclidean") geometry is essentially related to the relaxation of this condition, permitting the description of globally curved surfaces. See [Aczel 2000] for a brief history.

> corrections for perturbation have been made). These facts must, in my opinion, be taken as a convincing proof of the correctness of the theory.
>
> * Properly speaking, this can be affirmed only of [a linear combination of the Ricci tensor and the metric times the Ricci scalar]. [...]

We could have done no better ourselves to summarize the essential content of the theory of general relativity, modulo the inclusion of matter, which we address momentarily.

While so far we have been quoting Einstein directly along with his still widely used notation for spacetime indices—that is, with Greek letters—we shall, in our notation throughout this thesis, use Latin letters for spacetime indices instead, broadly following the conventions of [Wald 1984]. (In principle, these are to be understood as *abstract* indices. While this may be slightly abused sometimes if convenient and understood, we typically try to indicate when a particular choice of coordinates is employed by explicitly changing the indexing style, as further elaborated in the Notation and Conventions.) Thus, we denote by g_{ab} $(a, b = 0, 1, 2, 3)$ the spacetime metric, and we work in the $(-+++)$ signature. Using the summation convention that Einstein introduced not too long after Eq. (1.2.2), the square of the infinitesimal line element $\mathrm{d}s$ in terms of the metric components is given by:

$$\mathrm{d}s^2 = g_{ab}\mathrm{d}x^a\mathrm{d}x^b\,, \tag{1.2.3}$$

where $x^a = \{x^0, x^1, x^2, x^3\}$ are local coordinates.

We will often find it convenient to talk about tensors without always having to explicitly write their indices. For example, for referring to the mathematical object (metric tensor) g_{ab} we sometimes use, when more convenient, the completely equivalent (slanted boldface) index-free notation[6] $\boldsymbol{g}$, following the classic conventions of [Misner et al. 1973; Hawking and Ellis 1975]. This is the same idea as using either v^i $(i = 1, 2, 3)$ and $\vec{v}$ to represent abstractly the same object (in this case, a vector, *e.g.* a velocity) in classical physics. Interchanging between these two notations will make our expressions more compact and readable, and our language more fluid. For example, with a vector v^a and another w^a, index-free denoted $\boldsymbol{v}$ and $\boldsymbol{w}$ respectively, we write their inner product (with respect to the metric $\boldsymbol{g}$) as $\boldsymbol{v}\cdot\boldsymbol{w} = g_{ab}v^aw^b = v^aw_a$. We sometimes use similar notation for the "double" inner product, *i.e.* double contractions, *e.g.* for two rank-2 contravariant tensors A_{ab} and B_{ab}, we write $\boldsymbol{A} : \boldsymbol{B} = g^{ac}g^{bd}A_{ab}B_{cd} = A^{ab}B_{ab}$.

[6] We make this choice as often, at least in physics, the non-boldface g is used to refer to something else, in this case the determinant of the metric tensor; when not involving a metric tensor, but nonetheless still a rank-2 contravariant tensor, this notation is then typically reserved for referring instead to the trace.

Another helpful piece of notation we shall frequently employ is the use of index-free vectors (written in slanted bold) "as indices". This is meant to indicate projection in that index into the direction of the corresponding vector. For example, for any rank-2 contravariant tensor $\boldsymbol{A}$, we can write some of its projections in the directions of the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ as: $A_{\boldsymbol{v}b} = A_{ab}v^a$ (which is now a rank-1 contravariant tensor), $A_{\boldsymbol{v}\boldsymbol{w}} = A_{ab}v^a w^a$ (which is a scalar) etc.

These conventions naturally generalize to tensors of higher rank. For more details, see the Notation and Conventions as well as Appendix A.

The next basic object we need to define is the metric-compatible *derivative operator* or *connection* ∇_a (or $\boldsymbol{\nabla}$). It is the unique derivative operator the action of which on the metric makes the latter vanish, *i.e.* $\nabla_a g_{bc} = 0$ (or $\boldsymbol{\nabla g} = 0$). Equivalently, its action on an arbitrary vector field $\boldsymbol{v}$ is $\nabla_a v^b = \partial_a v^b + \Gamma^b{}_{ac} v^c$, with ∂_a denoting the usual partial derivative and $\Gamma^c{}_{ab} = \frac{1}{2} g^{cd}(\partial_a g_{bd} + \partial_b g_{ad} - \partial_d g_{ab})$ the Christoffel symbols. (If we had any other derivative operator on the RHS instead of the partial derivative, the latter are generally referred to as connection coefficients.) The action of $\boldsymbol{\nabla}$ on arbitrary tensors can be generalized from this.

We typically denote a spacetime, that is, a four-dimensional Lorentzian manifold, by $\mathscr{M}$. If a metric $\boldsymbol{g}$ and a compatible derivative $\boldsymbol{\nabla}$ are also defined on the manifold $\mathscr{M}$, then a spacetime more formally refers to the collection of objects

$$(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla}) \,, \tag{1.2.4}$$

such that $\boldsymbol{\nabla g} = 0$ in $\mathscr{M}$.

In our notation, the geodesic equation [Eq. (1.2.1)] is the same, and equivalent to the condition that the four-velocity u^a of the curve defined as a geodesic is parallel-transported therealong, *i.e.* it satisfies $u^a \nabla_a u^b = 0$, or in index-free notation,

$$\nabla_{\boldsymbol{u}} \boldsymbol{u} = 0 \,. \tag{1.2.5}$$

In local coordinates x^a, this in turn is equivalent to

$$\ddot{x}^a + \Gamma^a{}_{bc} \dot{x}^b \dot{x}^c = 0 \,, \tag{1.2.6}$$

where an overdot indicates a total derivative with respect to the (affine) parameter of the curve.

The notion of curvature is encoded in the *Riemann tensor* $R_{abc}{}^d$, defined from the derivative operator $\boldsymbol{\nabla}$ in the usual way: for any dual vector ω_a,

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)\, \omega_c = R_{abc}{}^d \omega_d \,. \tag{1.2.7}$$

Moreover, $R_{ac} = R_{abc}{}^{b}$ is the *Ricci tensor*[7] and, as usual, $R = \mathrm{tr}(\boldsymbol{R}) = g^{ab}R_{ab} = R_{a}{}^{a}$ is its trace, the *Ricci scalar*.

Defining the *Einstein tensor* as

$$\boldsymbol{G} = \boldsymbol{R} - \frac{1}{2}R\boldsymbol{g}\,, \tag{1.2.8}$$

the field equation of general relativity for the matter-free gravitational field, as Einstein introduced it above, is

$$\boldsymbol{G} = 0\,. \tag{1.2.9}$$

(This is equivalent to $\boldsymbol{R} = 0$.)

Matter, in the precise sense defined above by Einstein and the one to which we shall also adhere, is described by a stress-energy-momentum (symmetric rank-2 contravariant) tensor T_{ab}. If a matter action S_{M} is known (constructed from a Lagrangian yielding the correct matter field equations), $\boldsymbol{T}$ is simply defined, up to a factor, as the functional derivative of this action with respect to the metric, $\boldsymbol{T} = -\frac{2}{\sqrt{-g}}\frac{\delta}{\delta\boldsymbol{g}}S_{\mathrm{M}}$.

The *Einstein equation* in general states that this sources the Einstein tensor,

$$\boldsymbol{G} = \kappa\boldsymbol{T} \quad \text{in } \mathscr{M}\,, \tag{1.2.10}$$

where $\kappa = 8\pi G_{\mathrm{N}}/c^4$ is the Einstein constant, with G_{N} the Newton constant and c the speed of light. Note that we sometimes use the interchangeable nomenclature "Einstein equations" (in plural) to refer to the (ten) components of Eq. (1.2.10).

1.3. Geometry, gravity and motion

While Newton brazenly left "open to the consideration of [his] readers" the task of contemplating the nature his omnipresent gravitational "agent," Einstein had significantly more to say on this topic in the light and context of his own theory. A simplification of the main message of general relativity—reflected, at the most basic level, in the interpretation of the spacetime metric g_{ab} as the "gravitational field"—is that gravity ought to be conceived as nothing more than the manifestation of curvature in the geometry of spacetime. This is quite a generally accepted point of view today, and Einstein himself seems to have endorsed it at least at some "operational" level.

However, it seems that, to Einstein, the essence of the theory was more subtle than simply "reducing physics to geometry," a phrase to which he oftentimes attributed no, or otherwise completely tautological, meaning—insofar as the basic mathematical language of any theory of physics, at least in the post-Newtonian paradigm, lends itself to *some* level

[7] Note the very usual but notationally unfortunate use of the same symbol for denoting the these two tensors. In the index-free notation, we will reserve $\boldsymbol{R}$ usually to refer to the Ricci tensor (R_{ac}), and when we are talking about the Riemann tensor ($R_{abc}{}^{d}$) we shall make it clear.

of geometric representation by virtue of its ultimate association to our spatial experiences: from the field lines of Maxwell's theory that we have all seen plainly materialized, for example, in the orientation of iron shavings on a sheet of paper underneath a magnet, to more abstract constructs like the Hamiltonian phase space, described (as we shall see in ample detail in the next chapter of this thesis) by beautiful geometrical ideas of their own, in this case those of symplectic geometry. Indeed, Einstein explicitly complains about this in a letter to Lincoln Barnett towards the end of this life [Lehmkuhl 2014]:

> I do not agree with the idea that the general theory of relativity is geometrizing Physics or the gravitational field. The concepts of Physics have always been geometrical concepts and I cannot see why the g_{ik} field should be called more geometrical than f.[or] i.[nstance] the electromagnetic field or the distance of bodies in Newtonian Mechanics[8]. The notion comes probably from the fact that the mathematical origin of the g_{ik} field is the Gauss-Riemann theory of the metrical continuum which we wont look at as a part of geometry. I am convinced, however, that the distinction between geometrical and other kinds of fields is not logically found.

Instead, both during and after the development of general relativity, Einstein was much more fascinated with the connection implied by his theory between gravity and inertia—in particular, through the geodesic equation. The seeds of this lie in the equivalence principle, and specifically in his famous "fortunate thought" of 1907, which he recollects in 1920 (from [Lehmkuhl 2014]):

> Then I had the most fortunate thought of my life in the following form: The gravitational field only has a relative existence in a manner similar to the electric field generated by electro-magnetic induction. *Because for an observer in free-fall from the roof of a house, there is during the fall*—at least in his immediate vicinity—*no gravitational field*. Namely, if the observer lets go of any bodies, they remain, relative to him, in a state of rest or uniform motion, independent of their special chemical or physical nature.

In a series of lectures a year later, he elaborates upon his thoughts connecting these ideas [Einstein 1922] (from [Einstein 2002]):

> A material particle upon which no force acts moves, according to the principle of inertia, uniformly in a straight line. [...] The natural, that is, the simplest, generalization of the straight line which is meaningful in the system of concepts of the general (Riemannian) theory of invariants is that of the straightest, or geodesic, line. We shall accordingly have to assume, in the sense of the principle of equivalence, that the motion of a material particle, under the action only

[8] Note that this is reminiscent of some current ideas such as shape dynamics, pioneered by Barbour. See *e.g.* [Barbour 2012] for a review.

> of inertia and gravitation, is described by the equation,
>
> $$\frac{\mathrm{d}^2 x_\mu}{\mathrm{d}s^2} + \Gamma^\mu{}_{\alpha\beta} \frac{\mathrm{d}x_\alpha}{\mathrm{d}s} \frac{\mathrm{d}x_\beta}{\mathrm{d}s} = 0\,. \tag{1.3.1}$$
>
> In fact, this equation reduces to that of a straight line if all the components, $\Gamma^\mu{}_{\alpha\beta}$, of the gravitational field vanish.
>
> [...] [The above equations, our Eq. (1.3.1)] express the influence of inertia and gravitation upon the material particle. The unity of inertia and gravitation is formally expressed by the fact that the whole left-hand side of [our Eq. (1.3.1)] has the character of a tensor (with respect to any transformation of coordinates), but the two terms taken separately do not have tensor character. In analogy with Newton's equations, the first term would be regarded as the expression for inertia, and the second as the expression for the gravitational force.

It is worth underlining that the latter is only an analogy—and one that comes about only if one happened to be concerned with comparing general relativity to another specific theory, in this case the Newtonian one. To some extent, Einstein seems to have appreciated the unification of gravity and inertia in his theory, through the geodesic equation, similarly to that achieved between electricity and magnetism through Maxwell's equations. The interesting discrepancy, nevertheless, is that Einstein perceived his unification to lie not in the field equations of the theory itself (*i.e.* the Einstein equation), as had been the case with the electromagnetic unification, but rather in the equation of motion of idealized "test" particles.

Not surprisingly, perhaps, Einstein as well as others eventually became interested in the question of whether the geodesic equation could actually be obtained, under suitable conditions, as a consequence of the gravitational field equations—in *lieu* of postulating it as an independent, additional assumption. The first results in this direction began to arrive in the 1930s. In one of Einstein's first seminal papers specifically focused on this issue, co-written with Rosen, we see articulated the view that, indeed, the "field" concept ought to lie at the basis of motion [Einstein and Rosen 1935]:

> The main value of the considerations we are presenting consists in that they point the way to a satisfactory treatment of gravitational mechanics. One of the imperfections of the original relativistic theory of gravitation was that as a field theory it was not complete; it introduced the independent postulate that the law of motion of a particle is given by the equation of the geodesic. A complete field theory knows only fields and not the concepts of particle and motion. For these must not exist independently from the field but are to be treated as part of it. On the basis of the description of a particle without singularity, one has the possibility of a logically more satisfactory treatment of the combined problem: The problem of the field and that of the motion coincide.

Einstein continued working on this problem [Einstein, Infeld, and Hoffmann 1938], and by 1946 when he wrote the appendix to the third edition of his *Meaning of Relativity* [Einstein 2003], he was satisfied that it "has been shown that this law of motion—generalized to the case of arbitrarily large gravitating masses—can be derived from the field equations of empty space alone."

Since then, work in this direction has continued, and a variety of proofs have been put forward over the decades for the geodesic equation as the equation of motion of idealized "test" particles following from the field equations of general relativity. See [Geroch and Jang 1975; Ehlers and Geroch 2004] for some of the most famous such proofs. See also [Weatherall 2018] for a recent general review of the most widely used approaches, as well as an interesting novel proposal.

These proofs often—though certainly not always—involve, in some way, the modeling of the "test" particle as matter concentrated at a spatial point. In other words, one has a stress-energy-momentum tensor $T^{\rm PP}_{ab}$ for a point-particle, which is given by [Gralla and Wald 2008]:

$$T^{\rm PP}_{ab} = m\frac{u_a u_b}{\sqrt{-g}}\delta\left(x^i - z^i\right) , \tag{1.3.2}$$

where m is the mass, δ is in this case the three-dimensional Dirac delta function with x^i denoting spatial coordinates, z^i is the parametrization of the worldline in terms of proper time and u^a is the four-velocity (understood here as a function of proper time).

The next logical step from here is to pose the following problem: If a matter stress-energy-momentum tensor such as (1.3.2) is used to model the moving "small" mass m, this will source a similarly "small" correction to the spacetime metric g_{ab} through the Einstein equation. This, in turn, will induce a "small" correction to the (geodesic) motion. For historical reasons, mostly having to do with an analogous phenomenon in classical electromagnetism, this effect is referred to as the gravitational *self-force*. The analogy at the root of this nomenclature, of course, should be as clear as than that of Einstein when talking, *vis-à-vis* the Newtonian theory, of the Christoffel symbols expressing "the gravitational force" [Einstein 1922].

Until the last quarter century or so, the issue of the gravitational self-force had not been extensively studied. Concordantly, there had not been any truly compelling empirical opportunities available in astrophysics where self-force effects might be seen to play an important role. Now however, with the recent discovery of gravitational waves, a new window has been opened upon a wide variety of astrophysical phenomena, and especially binary systems in very strong gravitational regimes—including, prospectively, ones where the self-force plays a protagonistic role.

1.4. Gravitational waves and extreme-mass-ratio inspirals

The recent advent of gravitational wave astronomy—propelled by the ground-based direct detections achieved by the LIGO/Virgo collaboration (see [Abbott et al. 2018] for the detections during the O1 and O2 observing runs), the success of the LISA Pathfinder mission as a proof of principle for future space-based interferometric detectors [Armano et al. 2018, 2016], and the subsequent approval of the LISA mission for launch in the 2030s [Amaro-Seoane et al. 2017, 2013]—has generally brought a multitude of both practical and foundational problems to the foreground of gravitational physics today. While a plethora of possibilities for gravitational wave sources are actively being investigated theoretically and anticipated to become accessible observationally, both on the Earth as well as in space, the most ubiquitous class of such sources has manifestly been—and foreseeably will remain—the coalescence of compact object binaries [Celoria et al. 2018; Colpi and Sesana 2016]. These are two-body systems consisting of a pair of compact objects, say of masses M_1 and M_2, orbiting and eventually spiraling into each other. Each of these is, usually, either a black hole (BH) or a neutron star. There are also more general possibilities being investigated, including that of having a brown dwarf as one of the objects [Amaro-Seoane 2019].

The LIGO/Virgo detections during the first scientific runs [Abbott et al. 2018], O1 and O2, have all involved binaries of *stellar-mass compact objects* (SCOs) located in our local neighbourhood. These have comparable masses, of the order of a few tens of solar masses each ($M_1 \sim M_2 \sim 10^{0-2} M_\odot$). In addition second- and third- generation terrestrial detectors can also eventually see *intermediate-mass-ratio inspirals*, binaries consisting of an intermediate-mass BH, of $10^{2-4} M_\odot$, and an SCO. While there is as yet no direct evidence for the existence of the former sorts of objects, there are good reasons to anticipate their detection (through gravitational waves) most likely at the centers of globular clusters, and their study provides an essential link to the strongly perturbative regime of compact object binary dynamics.

It is even further in this direction that future space-based detectors such as LISA are anticipated to take us. In particular, LISA is expected to see *extreme-mass-ratio inspirals* (EMRIs) [Amaro-Seoane 2018], compact binaries where $M_1 \gg M_2$. An elementary sketch is depicted in Figure 1.4. The more massive object could be a (*super-*) *massive* black hole (MBH) of mass $M_1 = M \sim 10^{4-7}\, M_\odot$ located at a galactic center, with the significantly less massive object—effectively orbiting and eventually spiraling into the MBH—being an SCO: either a stellar-mass black hole or a neutron star, with $M_2 = m \sim 10^{0-2} M_\odot$.

Average estimates indicate that LISA will be able to see on the order of hundreds of EMRI events per year [Babak et al. 2017], with an expectation of observing, for each, thousands of orbital cycles over a period on the order of one year before the final plunge [Barack and Pound 2018]. The trajectories defining these cycles and the gravitational

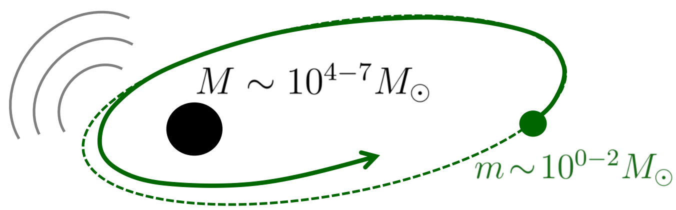

FIGURE 1.4. Sketch of an extreme-mass-ratio inspiral (EMRI), a two-body system consisting of a stellar-mass compact object (SCO), usually a stellar-mass black hole, of mass $m \sim 10^{0-2} M_\odot$, orbiting and eventually spiralling into a (super-) massive black hole (MBH), of mass $M \sim 10^{4-7} M_\odot$, and emitting gravitational waves in the process.

wave signals produced by them will generally look much more complex than the relatively generic signals from mergers of stellar-mass black holes of comparable masses as observed, for example, by LIGO/Virgo.

EMRIs will therefore offer an ideal experimental milieu for strong gravity: the complicated motion of the SCO around the MBH will effectively "map out" the geometry—that is, the gravitational field—around the MBH, thus presenting us with an unprecedented opportunity for studying gravity in the very strong regime [Babak et al. 2017; Berry et al. 2019]. In particular, among the possibilities offered by EMRIs are the measurement of the mass and spin of the MBH to very high accuracy, testing the validity of the Kerr metric as the correct description of BHs within general relativity (GR), and testing GR itself as the correct theory of gravity.

Yet, the richness of the observational opportunities presented by EMRIs comes with an inexorable cost: that is, a significant and as yet ongoing technical challenge in their theoretical modeling. This is all the more pressing as the EMRI signals expected from LISA are anticipated to be much weaker than the instrumental noise of the detector. Effectively, what this means is that extremely accurate models are necessary in order to produce the waveform templates that can be used to extract the relevant signals from the detector data stream. At the theoretical level, the problem of EMRI modeling cannot be tackled directly with numerical relativity (used for the LIGO/Virgo detections), simply due to the great discrepancy in (mass/length) scales; however, for the same reason, the approach that readily suggests itself is perturbation theory. See Figure 1.5 for a graphic depicting the main methods used for compact object binary modeling in the different regimes. In particular, modeling the strong gravity, extreme mass ratio regime turns out to be equivalent to a general and quite old problem which can be posed in any (not just gravitational) classical field theory: the so-called *self-force* problem.

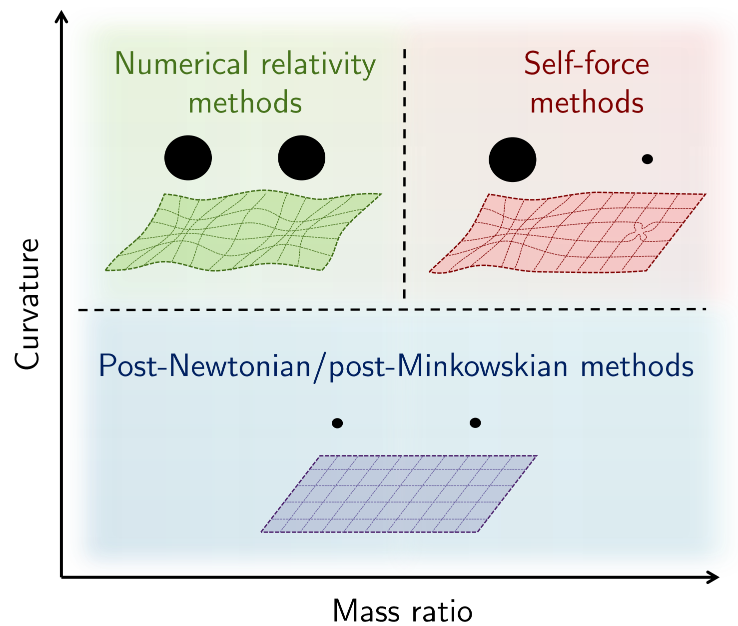

FIGURE 1.5. The main approaches used in practice for the modeling of compact object binaries as a function of the mass ratio (increasing from 1) and the spacetime curvature involved. For low curvature (high separation between the bodies), post-Newtonian and post-Minkowskian methods are used. For high curvature (low separation) and low mass ratio, numerical relativity is used. For high curvature and extreme mass ratios, as the scale of a numerical grid would have to span orders of magnitude thus rendering it impracticable, perturbation theory must be used—in particular, self-force methods.

1.5. The self-force problem

Suppose we are dealing with a theory for a field $\psi(x)$ in some spacetime. If the theory admits a Lagrangian formulation, we can usually assume that the field equations have the general form

$$L[\psi\left(x\right)] = S\left(x\right) \,, \tag{1.5.1}$$

where L is a (partial, possibly non-linear and typically second-order) differential operator, and we refer to S as the "source" of the field ψ. Broadly speaking, the problem of the self-force is to find solutions $\psi(x)$ satisfying (1.5.1) when S is "localized" in spacetime. Intuitively, it is the question of how the existence of a dynamical (field-generating) "small object" (a mass, a charge etc.) backreacts upon the total field ψ, and hence in turn upon its own future evolution subject to that field. Thus, an essential part of any detailed self-force analysis is a precise specification of what exactly it means for S to be localized. In standard approaches, one typically devises a perturbative procedure whereby S ends up being approximated as a distribution, usually a Dirac delta, compactly supported on a worldline—that is, the "background" (zeroth perturbative order) worldline of the small object. However, this already introduces a nontrivial mathematical issue: if L is non-linear (in the standard PDE sense), then the problem (1.5.1) with a distributional source S is mathematically ill-defined, at least within the classical theory of distributions [Schwartz 1957] where products of distributions do not make sense [Schwartz 1954][9].

One might therefore worry that nonlinear physical theories, such as GR, would a priori not admit solutions sourced by distributions, and we refer the interested reader to [Geroch and Traschen 1987] for a classic detailed discussion of this topic. The saving point is that, while the full field equation (in this case, the Einstein equation) may indeed be generally non-linear, if we devise a perturbative procedure (where the meaning of the "perturbation" is prescribed in such a way as to account for the presence of the small object itself), then the first-order field equation is, by construction, linear in the (first-order) perturbation $\delta\psi$ of ψ. Thus, assuming the "background" field is a known exact solution of the theory, it always makes sense to seek solutions $\delta\psi$ to

$$\delta L[\delta\psi\,(x)] = S\,(x)\ , \tag{1.5.2}$$

for a distributional source S, where δL indicates the first-order part of the operator L in the full field equation (1.5.1). As this only makes sense for the (linear) first-order problem, such an approach becomes again ill defined if we begin to ask about the (nonlinear) second- or any higher-order problem. Additional technical constructions are needed to deal with these, the most common of which for the gravitational self-force has been the so-called "puncture" (or "effective source") method [Barack and Golbourn 2007; Barack, Golbourn, and Sago 2007; Barack and Pound 2018; Vega and Detweiler 2008]; similar ideas have proven to be very useful also in numerical relativity [Baker et al. 2006; Campanelli et al. 2006]. For work on the second-order equation of motion for the gravitational self-force problem, see *e.g.* Refs. [Gralla 2012; Pound 2017, 2012]. For now, we assume that we are interested here in the first-order self-force problem (1.5.2) only.

Now concretely, in GR, our physical field ψ is simply the spacetime metric g_{ab} (where Latin letters from the beginning of the alphabet indicate spacetime indices), and following standard convention we denote a first-order perturbation thereof by $\delta g_{ab} = h_{ab}$. The

[9] Nonlinear theories of distributions are being actively investigated by mathematicians [Bottazzi 2017; Colombeau 2013; Li 2007], however at this point, to our knowledge, their potential applicability to the self-force problem has not been contemplated to any significant extent.

problem (1.5.2) is then just the first-order Einstein equation,

$$\delta G_{ab}[h_{cd}] = \kappa T^{\rm PP}_{ab}\,, \tag{1.5.3}$$

where G_{ab} is the Einstein tensor, $\kappa = 8\pi$ (in geometrized units $c = G = 1$) is the Einstein constant, and $T^{\rm PP}_{ab}$ the energy-momentum tensor of a "point particle" (PP) compactly supported on a given worldline $\mathring{\mathscr{C}}$. We will return later to discussing this more precisely, but in typical approaches, $\mathring{\mathscr{C}}$ turns out to be a geodesic—that is, the "background" motion of the small object, which is in this case a small mass[10]. Thus, simply solving (1.5.3) for h_{ab} assuming a fixed $\mathring{\mathscr{C}}$ for all time, though mathematically well-defined, is by itself physically meaningless: it would simply give us the metric perturbations caused by a small object eternally moving on the same geodesic. Instead what we would ultimately like is a way to take into account how h_{ab} modifies the motion of the small object itself. Thus in addition to the field equation (1.5.3), any self-force analysis must be supplemented by an *equation of motion* (EoM) telling us, essentially, how to move from a given background geodesic $\mathring{\mathscr{C}}$ at one step in the (ultimately numerical) time evolution problem to a new background geodesic $\mathring{\mathscr{C}}'$ at the next time step—with respect to which the field equation (1.5.3) is solved anew, and so on. This is sometimes referred to as a "self-consistent" approach. See Fig. 1.6 for a visual depiction.

The first proposal for an EoM for the *gravitational self-force* (GSF) problem was put forward in the late 1990s, since known as the MiSaTaQuWa equation after its authors [Mino et al. 1997; Quinn and Wald 1997]. On any $\mathring{\mathscr{C}}$, it reads:

$$\ddot{Z}^a = -\mathring{E}_b{}^a Z^b + F^a[h^{\rm tail}_{cd}; \mathring{U}^e]\,. \tag{1.5.4}$$

The LHS is a second (proper) time derivative of a *deviation vector* Z^a on $\mathring{\mathscr{C}}$ pointing in the direction of the "true motion" (away from $\mathring{\mathscr{C}}$), to be defined more precisely later. On the RHS, $\mathring{E}_{ab}$ is the electric part of the Weyl tensor on $\mathring{\mathscr{C}}$, such that the first term is a usual "geodesic deviation" term. The second term on the RHS is the one usually understood as being responsible for self-force effects: $F^a[\cdot;\cdot]$ is a four-vector functional of a symmetric rank-2 contravariant tensor and a vector, to which we refer in general (for any arguments) as the *GSF functional*. In any spacetime with a given metric $\mathring{g}_{ab}$ and compatible derivative operator $\mathring{\nabla}_a$, it is explicitly given by the following simple action of a first-order differential operator:

$$F^a[H_{bc}; V^d] = -\left(\mathring{g}^{ab} + V^a V^b\right)\left(\mathring{\nabla}_c H_{bd} - \frac{1}{2}\mathring{\nabla}_b H_{cd}\right) V^c V^d\,. \tag{1.5.5}$$

[10] We consider later in this thesis (Chapter 5) in detail one approach to the gravitational self-force which also proves geodesic motion as the "background" motion of point particles in GR.

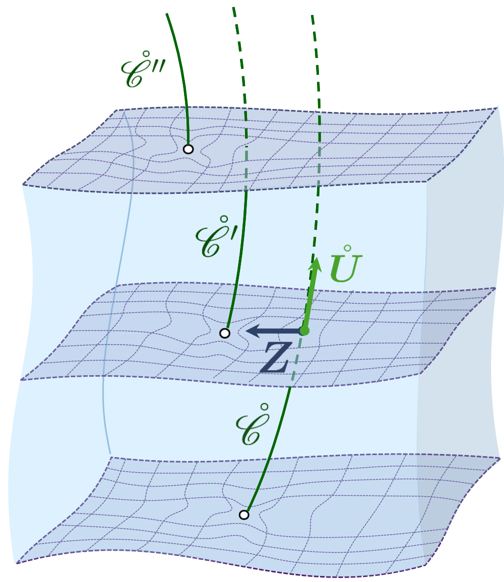

FIGURE 1.6. A depiction of the perturbative problem for the gravitational self-force (GSF). In particular, this represents one of the most popular conceptions of a so-called "self-consistent" approach [Gralla and Wald 2008]: at a given step (on a given Cauchy surface) in the time evolution problem, one computes the "correction to the motion" away from geodesic ($\mathring{\mathscr{C}}$) in the form of a deviation vector Z^a, determined by the GSF. Then, at the next time step, one begins on a new ("corrected") geodesic ($\mathring{\mathscr{C}}'$), computes the new deviation vector, and so on.

While this is easy enough to calculate once one knows the arguments, the main technical challenge in using the MiSaTaQuWa equation (1.5.4) lies precisely in the determination thereof: in particular, h^{tail}_{ab} is not the *full* metric perturbation h_{ab} which solves the field equation (1.5.3), but instead represents what is called the "tail" integral of the Green functions of h_{ab}. This quantity is well defined, but difficult to calculate in practice and usually requires the fixing of a perturbative gauge—typically the *Lorenz* gauge, $\mathring{\nabla}^b(h_{ab} - \frac{1}{2}h_{cd}\mathring{g}^{cd}\mathring{g}_{ab}) = 0$. Physically, h^{tail}_{ab} can be thought of as the part of the full perturbation $\boldsymbol{h}$ which is scattered back by the spacetime curvature. (In this way, $\boldsymbol{h}$ can be regarded as the sum of h^{tail}_{ab} and the remainder, what is sometimes called the "instantaneous" or "direct" part h^{direct}_{ab}, responsible for waves radiated to infinity [A. D. A. M. Spallicci, Ritter, and Aoudia 2014].)

An alternative, equivalent GSF EoM was proposed by Detweiler and Whiting in the early 2000s [Detweiler and Whiting 2003]. It relies upon a regularization procedure for the metric perturbations, *i.e.* a choice of a decomposition for h_{ab} (the full solution of the field equation (1.5.3)) into the sum of two parts: one which diverges—in fact, one which contains *all* divergent contributions—on $\mathring{\mathscr{C}}$, denoted $h^{\rm S}_{ab}$ (the so-called "singular" field, related to the "direct" part of the metric perturbation), and a remainder which is finite, $h^{\rm R}_{ab}$ (the so-called "regular" field, related to the "tail" part), so that one writes $h_{ab} = h^{\rm S}_{ab} + h^{\rm R}_{ab}$. An analogy with the self-force problem in electromagnetism gives some physical intuition behind how to interpret the meaning of this decomposition [Barack and Pound 2018], with $h^{\rm S}_{ab} \sim m/r$ having the heuristic form of a "Coulombian self-field." However, no procedure is known for obtaining the precise expression of $h^{\rm S}_{ab}$ in an arbitrary perturbative gauge, and moreover, once a gauge is fixed (again, usually the Lorenz gauge), this splitting is not unique [Barack and Pound 2018]. Nevertheless, if and when such an $h^{\rm S}_{ab}$ is obtained (from which we thus also get $h^{\rm R}_{ab} = h_{ab} - h^{\rm S}_{ab}$), the Detweiler-Whiting EoM for the GSF reads:

$$\ddot{Z}^a = -\mathring{E}_b{}^a Z^b + F^a[h^{\rm R}_{cd}; \mathring{U}^e] \,. \tag{1.5.6}$$

The EoMs (1.5.4) and (1.5.6) are equivalent in the Lorenz gauge and form the basis of the two most popular methods used today for the numerical computation of the GSF. For detailed reviews, see [Barack and Pound 2018; Poisson et al. 2011]. Yet, a great deal of additional technical machinery is required for handling gauge transformations. This is essential because, in the EMRI problem, the background spacetime metric—that of the MBH—is usually assumed to be Schwarzschild or Kerr. Perturbation theory for such spacetimes has been developed and is most easily carried out in, respectively, the so-called Regge-Wheeler and radiation gauges; in other words, in practice, it is often difficult (though not infeasible—see *e.g.* [Barack and Lousto 2005]) to compute h_{ab} directly in the Lorenz gauge for use in (1.5.4) or (1.5.6).

A proposal for an EoM for the GSF that is valid in a wider class of perturbative gauges was presented by Gralla in 2011 [Gralla 2011]. In particular, it is valid in what are called "parity-regular" gauges, *i.e.* gauges satisfying a certain parity condition. This condition ultimately has its origins in the Hamiltonian analysis of Regge and Teitleboim in the 1970s [Regge and Teitelboim 1974], wherein the authors introduce it in order to facilitate the vanishing of certain surface integrals and thus to render certain general-relativistic Hamiltonian notions, such as multipoles and "center of mass," well-defined mathematically. In parity-regular gauges (satisfying the Regge-Teitleboim parity condition), the Gralla EoM—mathematically equivalent, in the Lorenz gauge, to the MiSaTaQuWa and the Detweiler-Whiting EoMs—is:

$$\ddot{Z}^a = -\mathring{E}_b{}^a Z^b + \frac{1}{4\pi} \lim_{r\to 0} \int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2} \, F^a[h_{cd}; \mathring{U}^e] \,. \tag{1.5.7}$$

The GSF (last) term on the RHS is obtained in this approach by essentially relating the deviation vector (the evolution of which is expressed by the LHS) with a gauge transformation vector and then performing an "angle average" over an r-radius two-sphere $\mathbb{S}^2_r$, with volume form $\boldsymbol{\epsilon}_{\mathbb{S}^2}$, of the so-called "bare" GSF, $F^a[h_{bc}; \mathring{U}^d]$. The latter is just the GSF functional [Eq. (1.5.5)] evaluated directly using the *full* metric perturbatiuon h_{ab} (*i.e.* the "tail" plus "direct" parts, or equivalently, the "regular" plus "singular" parts), around (rather than at the location of) the distributional source. The point therefore is that this formula never requires the evaluation of h_{ab} on $\mathring{\mathscr{C}}$ itself, where it is divergent by construction; instead, away from $\mathring{\mathscr{C}}$ it is always finite, and (1.5.7) says that it suffices to compute the GSF functional (1.5.5) with h_{ab} directly in the argument (requiring no further transformations), and integrate it over a small sphere.

The manifest advantage of (1.5.7) relative to (1.5.4) or (1.5.6) is that no computations of tail integrals or regularizations of the metric perturbations are needed at all. Yet, to our knowledge, there has thus far been no attempted numerical computation of the GSF using this formula. One of the issues with this remains that of the perturbative gauge: the parity-regular class actually still does not include the Regge-Wheeler and radiation gauges, and thus further work is needed to use an h_{ab} computed in such gauges. Aside from the practical issues with a possible numerical implementation of this, there is also a conceptual issue: this formula results from an essentially mathematical argument—by a convenient "averaging" over the angles—so as to make it well-defined in a Hamiltonian setting via a relation to a canonical definition of the center of mass. Yet its general form as a closed two-surface integral suggestively hints at the possibility of interpreting it not merely as a convenient mathematical relation, but as a real physical flux of (some notion of) "gravitational momentum". We contend and will demonstrate in this thesis (specifically, in Chapter 5) that indeed an even more general version of (1.5.7) results from the consideration of momentum conservation laws in GR.

CHAPTER 2

Canonical General Relativity

Chapter summary. The aim of this chapter is to introduce the basic mathematical language and technical machinery of the theory of general relativity following variational methods. We focus especially on developing the canonical formulation of general relativity, also known as the Hamiltonian or (3 + 1) formulation. In essence, this provides a way of turning the second-order field equations of the theory for the spacetime metric into a first-order time-evolution problem for the induced spatial three-metric and its conjugate momentum. There are, however, also constraint equations in addition to these (constraining permissible initial conditions as well as their development subject to the dynamical equations), and their proper handling requires a great deal of subtlety. The existence of constraints in general relativity is in fact directly related to—and offers fruitful insight on—the gauge freedom available in the theory.

We begin in Section 2.1 with a brief introduction. We generally discuss four broad areas of application for canonical general relativity: mathematical relativity, numerical relativity, quantum gravity, and the issue of gravitational energy-momentum. We return to each of these in the final section of this chapter with specific examples, once we have developed the mathematical tools in detail.

In Section 2.2, we present the Lagrangian formulation of classical field theories in general, and then general relativity in particular, forming the typical starting point of any canonical analysis. We comment on the appearance of constraint equations already at the Lagrangian level, a proper explanation of which requires the canonical picture.

In Section 2.3, we develop the canonical formulation of field theories in general, with a careful accounting of the issue of constraints. To this end, we prescribe here the general recipe for foliating spacetime into constant-time spatial (Cauchy) three-surfaces, such that a notion of time evolution in spacetime can be defined. We also define the canonical phase space and the Hamiltonian equations of motion for general field theories, introducing the basic mathematical methods of symplectic geometry.

Then, in section 2.4, we proceed to apply this formalism to general relativity in order to formulate it as a canonical theory. In particular, the canonical variables are the induced three-metric on each spatial slice (associated with the choice of foliation), as well as the lapse function and the shift vector (associated with a choice of a time flow vector field), plus their respective conjugate momenta. The lapse and shift are not dynamical variables:

their associated equations are first-order in time, and their conjugate momenta vanish. These constitute the constraints of general relativity.

Finally, in Section 2.5, we return in greater detail to the four broad areas of application of canonical general relativity enumerated in the introductory section, and we provide illustrations with explicit examples.

Relativitat general canònica (chapter summary translation in Catalan). L'objectiu d'aquest capítol és introduir el llenguatge matemàtic bàsic i la maquinària tècnica de la teoria de la relativitat general seguint mètodes variacionals. Ens centrem especialment en desenvolupar la formulació canònica de la relativitat general, també coneguda com la formulació hamiltoniana o (3 + 1). En essència, això proporciona una manera de convertir les equacions de camp de segon ordre de la teoria per al tensor mètric de l'espai-temps en un problema d'evolució temporal de primer ordre per al tensor mètrica espacial tridimensional induït i el seu moment conjugat. Tanmateix, també hi ha equacions de restricció a més d'aquestes (restringint les condicions inicials admissibles, així com el seu desenvolupament subjecte a les equacions dinàmiques), i el seu correcte maneig requereix molta subtilesa. L'existència de restriccions en la relativitat general està directament relacionada amb - i ofereix una visió útil sobre - la llibertat de mesura disponible en la teoria.

Comencem a la Secció 2.1 amb una breu introducció. Generalment es discuteixen quatre àmplies àrees d'aplicació de la relativitat general canònica: la relativitat matemàtica, la relativitat numèrica, la gravetat quàntica i el tema de l'energia i la quantitat de moviment gravitatòria. Tornem a cadascun d'aquests a la secció final d'aquest capítol amb exemples específics, un cop desenvolupades les eines matemàtiques en detall.

A la secció 2.2, presentem la formulació lagrangiana de les teories de camps clàssics en general, i la relativitat general en particular, formant el punt de partida típic de qualsevol anàlisi canònica. Comentem l'aparició d'equacions de restricció ja a nivell lagrangià, una explicació adequada de la qual es requereix la formulació canònica.

A la secció 2.3, desenvolupem la formulació canònica de les teories de camps en general, amb una acurada explicació del problema de les restriccions. Amb aquesta finalitat, prescrivim aquí la recepta general per foliar l'espai-temps en superfícies espacials tridimensionals de temps constant (superfícies Cauchy), de manera que es pot definir una noció d'evolució en el temps en l'espai-temps. També definim l'espai de fase canònica i les equacions de moviment hamiltonianes per a les teories generals de camp, introduint els mètodes matemàtics bàsics de la geometria simplectica.

A continuació, a la secció 2.4, procedim a aplicar aquest formalisme a la relativitat general per tal de formular-lo com a teoria canònica. En particular, les variables canòniques són el tensor mètric tridimensional induït en cada llesca espacial (associat a l'elecció de la foliació), així com la funció de lapse i el vector de shift (associats amb una elecció d'un camp vectorial de flux de temps), més els seus moments conjugats respectius. El lapse i el shift no són variables dinàmiques: les seves equacions associades són de primer ordre en el temps i els seus moments conjugats desapareixen. Aquests constitueixen les restriccions de la relativitat general.

Finalment a la secció 2.5, tornem amb més detall a les quatre àmplies àrees d'aplicació de la relativitat general canònica enumerades a la secció introductòria, i proporcionem il·lustracions amb exemples explícits.

Relativité générale canonique (chapter summary translation in French). Le but de ce chapitre est de présenter le langage mathématique de base et la machinerie technique de la théorie de la relativité générale suivant les méthodes variationnelles. Nous nous concentrons particulièrement sur le développement de la formulation canonique de la relativité générale, également connue sous le nom de la formulation hamiltonienne ou (3 + 1). En substance, cela fournit un moyen de transformer les équations de champ de deuxième ordre de la théorie pour le tenseur métrique de l'espace-temps en un problème d'évolution temporelle de premier ordre pour le tenseur métrique spatial trois-dimensionnel induit et son moment conjugué. Cependant, il existe également des équations de contrainte (restreignant les conditions initiales admissibles ainsi que leur développement en fonction des équations dynamiques), et leur traitement correct nécessite beaucoup de subtilité. L'existence de contraintes dans la relativité générale est en fait directement liée à - et offre un éclairage utile sur - la liberté de jauge disponible dans la théorie.

Nous commençons à la section 2.1 avec une brève introduction. Nous traitons généralement quatre grands domaines d'application de la relativité générale canonique : la relativité mathématique, la relativité numérique, la gravitation quantique et la question de l'énergie et la quantité de mouvement gravitationnelles. Nous reviendrons sur chacun d'eux dans la dernière section de ce chapitre avec des exemples spécifiques, une fois que nous aurons développé les outils mathématiques en détail.

Dans la section 2.2, nous présentons la formulation lagrangienne des théories de champ classiques en général, puis de la relativité générale en particulier, constituant le point de départ typique de toute analyse canonique. Nous commentons sur l'apparition d'équations de contraintes déjà au niveau lagrangien, dont l'explication correcte nécessite la formulation canonique.

Dans la section 2.3, nous développons la formulation canonique des théories de champ en général, tenant en compte la question des contraintes. À cette fin, nous prescrivons ici la recette générale de feuilleter l'espace-temps en surfaces spatiales trois-dimensionnelles à temps constant (surfaces de Cauchy), de manière à pouvoir définir une notion d'évolution temporelle dans l'espace-temps. Nous définissons également l'espace des phases canonique et les équations hamiltoniennes du mouvement pour les théories générales du champ, en introduisant les méthodes mathématiques de base de la géométrie symplectique.

Ensuite, dans la section 2.4, nous appliquons ce formalisme à la relativité générale afin de la formuler en tant que théorie canonique. En particulier, les variables canoniques sont le tenseur métrique trois-dimensionnel induite sur chaque tranche spatiale (associée au choix de la foliation), et la fonction de déchéance (*lapse*) et le vecteur de décalage (*shift*) (associés au choix d'un champ de vecteurs de flux temporel), ainsi que leurs moments conjugués respectifs. La déchéance et le décalage ne sont pas de variables dynamiques : leurs équations associées sont du premier ordre dans le temps et leurs moments conjugués disparaissent. Celles-ci constituent les contraintes de la relativité générale.

Enfin, dans la section 2.5, nous reviendrons plus en détail sur les quatre grands domaines d'application de la relativité générale canonique énumérés dans la section introductive et nous fournissons des illustrations avec des exemples explicites.

2.1. Introduction

There are a number of diverse motivations for casting GR into a canonical form, and for our choice to introduce the topic in this way. We begin by enumerating four broad areas of interest, and comment more on each, focusing on specific examples of applications, in the final section of this chapter.

(1) Mathematically, canonical methods provide a very useful way to develop the sort of geometrical tools generally used for studying subsets of spacetimes, in particular by splitting them up into (usually, families of lower-dimensional) hypersurfaces via some established procedure. The classical Hamiltonian approach splits spacetime into spatial slices defined, for example, by the constancy of a time function, and is thus adapted to studying "the entire space" at different instants of time. Similar technical constructions can be employed for studying the dynamics of finite (bounded) systems within a spacetime throughout some span of time; in such a case, one could foliate spacetime, for instance, by the constancy of a radial function in order to study the dynamics of the resulting worldtubes. (We shall see more along these lines in Chapter 5.) Spacetime splittings of this sort, and especially $(3 + 1)$ splittings, have supplied the basic framework for many important results in the mathematics of GR.

(2) Practically, canonical methods form the basis of numerical relativity—that is, formulating the Einstein equation as a suitable set of time-dependent partial differential equations (PDEs) which, given some appropriate initial data, can be evolved on computers to obtain numerical solutions. Simulations of strongly dynamical gravitational systems rely critically on methods of this sort.

(3) In going beyond GR, in particular in seeking theories of quantum gravity, canonical methods are often regarded as a key connection between the languages of GR and quantum mechanics. Indeed, typical canonical quantization procedures follow some variant of transforming classical canonical variables into operators on a Hilbert space of quantum states. Loop quantum gravity, for example, is a candidate theory of quantum gravity which essentially attempts to do this for gravitational canonical variables defined in a suitable way.

(4) Last but not at all least, from a physical point of view, canonical methods form the traditional starting point for understanding the notion of gravitational energy-momentum. In particular, they can supply definitions of gravitational energy-momentum of an entire spacetime under some specific conditions, *e.g.* the Arnowitt-Deser-Misner (ADM) energy-momentum for asymptotically-flat spacetimes. On the other hand, canonical methods are not designed to be able to say much more than this, and in particular, anything about the gravitational energy-momentum of a *finite* spatial region within some spacetime. For the latter, methods such as the worldtube boundary splittings mentioned in (1) are better designed, yet to this day no general consensus exists among relativists on the "best" way to do this. We will comment more on this in later chapters of this thesis, but we end here by remarking that, nevertheless, any proposed definition for gravitational energy-momentum of a finite system is generally expected to agree with, *e.g.*, the ADM energy-momentum in the flat asymptotic limit.

Canonical GR encompasses a variety of possible formulations of GR in terms of some chosen canonical variables (configurations and their conjugate momenta). The first canonical formulation of GR was achieved in 1950, following a quantum gravity motivation, by [Pirani and Schild 1950, Pirani 1951], and independently the following year by [Anderson and Bergmann 1951].

Then, [Dirac 1958a,b] formulated the general framework for working with constrained Hamiltonian systems, a topic we systematically develop in this chapter before applying it to GR. Beginning in the following year, Arnowitt, Deser and Misner [Arnowitt et al. 1959] devised the first coordinate-independent canonical formulation of GR, since then known eponymously as the *ADM formulation*. Following a series of further papers in the ensuing years, the authors summarized their results in a 1962 review article [Arnowitt et al. 1962], republished more recently [Arnowitt et al. 2008]. Today, it remains undoubtedly the most famous basic canonical formulation of GR.

Over the decades, other formulations have been developed in response to the application needs—*e.g.*, the Ashtekar variable formulation used in quantum gravity, the

Baumgarte-Shapiro-Shibata-Nakamura (BSSN) and generalized harmonic formulations used widely in numerical relativity (upon which we will elaborate further in the final section), and numerous others.

Our presentation of canonical GR in this chapter is based, in its broadest outlines, on Chapter 10 and Appendix E of [Wald 1984], in combination with Chapter 3 of [Bojowald 2011] (especially for the general formulation canonical theories and constraints). See also Chapter 4 of [Poisson 2007] for many of the the step-by-step computations, largely omitted here *in lieu* of directly stating the main results.

For mathematical clarifications, we generally refer to [Lee 2002] for geometry (see also [Nakahara 2003], written with more of a view towards physics), and [Evans 1998] for PDE theory.

2.2. Lagrangian formulation

2.2.1. Lagrangian formulation of general field theories. Let $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ be any $(3+1)$-dimensional spacetime. Suppose that we are interested in a theory describing a collection of fields $\psi = \{\psi^A(x^a)\}_{A\in\mathscr{I}}$ in $\mathscr{M}$, where $A \in \mathscr{I}$ is a general (possibly multi-) index for the fields ψ^A (and will be accordingly omitted if understood), *i.e.* it can include tensor indices, field indices etc. For example, if we are considering only gravity, then $\psi = \boldsymbol{g}$, *i.e.* our collection of fields includes only the spacetime metric g_{ab}. If we are considering gravity coupled to a matter field, for example Maxwellian electromagnetism, then $\psi = \{\boldsymbol{g}, \boldsymbol{F}\}$, where F_{ab} is the Faraday tensor.

Let $\mathcal{S}[\psi]$ be a functional of ψ. Let $\{\psi_{(\lambda)}\}_{\lambda\in\mathbb{R}}$ be a smooth one-parameter family of field values and let $\delta\psi^A = (\partial_\lambda \psi^A_{(\lambda)})|_{\lambda=0}$. For all such families, suppose moreover that $(\partial_\lambda \mathcal{S}[\psi_{(\lambda)}])|_{\lambda=0}$ exists and also that there exists a smooth field χ_A dual to ψ^A (meaning that if $\psi^A \in \mathscr{T}^k{}_l(\mathscr{M})$, then $\chi_A \in \mathscr{T}^l{}_k(\mathscr{M})$), such that

$$\left(\frac{\partial \mathcal{S}}{\partial \lambda}\right)\bigg|_{\lambda=0} = \int_{\mathscr{M}} \boldsymbol{e}\, \chi_A \delta\psi^A \,. \tag{2.2.1}$$

Here, for reasons that will become transparent shortly, we choose to write the integral with respect to the Minkowski volume form[1],

$$\boldsymbol{e} = \mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 = \mathrm{d}^4 x \,. \tag{2.2.2}$$

The factor of $\sqrt{-g}$, with $g = \det(\boldsymbol{g})$, multiplying the above to yield the volume form of $\mathscr{M}$,

$$\boldsymbol{\epsilon}_{\mathscr{M}} = \sqrt{-g}\, \boldsymbol{e} \,, \tag{2.2.3}$$

[1] In fact, $\boldsymbol{e}$ can be chosen to be *any* Lorentzian volume form the non-vanishing components of which take the values ± 1. See [Wald 1984] for a more detailed discussion as to why the present construction is actually independent of the choice of a volume form satisfying this property.

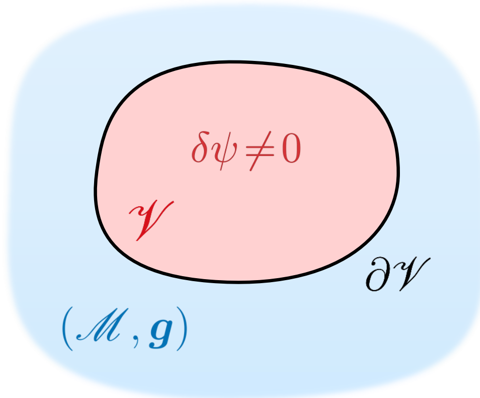

FIGURE 2.1. A compact region $\mathscr{V}$ in a spacetime $\mathscr{M}$ where the variation of physical fields are non-zero.

is absorbed into the definition of χ_A. Then, $\mathcal{S}[\psi]$ is said to be *functionally differentiable* at $\psi = \psi_{(0)}$ and its *functional derivative* is defined as

$$\left.\left(\frac{\delta\mathcal{S}}{\delta\psi^A}\right)\right|_{\psi_{(0)}} = \chi_A \,. \tag{2.2.4}$$

We will now focus our attention upon a certain class of such functionals $\mathcal{S}$. Let $\mathscr{V}$ be a compact region in $\mathscr{M}$ such that $\mathrm{supp}(\delta\psi^A) \subset \mathscr{V}$ (*i.e.* $\delta\psi^A$ takes non-zero values only in the interior of $\mathscr{V}$). See Fig. 2.1. We assume that $\mathcal{S}$ has the form

$$\mathcal{S}\left[\psi\right] = \int_{\mathscr{V}} \boldsymbol{e}\, \mathcal{L}\left[\psi\right] \,, \tag{2.2.5}$$

such that

$$\mathcal{L}\left[\psi\right] = \sqrt{-g} f\left(\psi^A, \boldsymbol{\nabla}\psi^A, \ldots, \boldsymbol{\nabla}\cdots\boldsymbol{\nabla}\psi^A\right) \tag{2.2.6}$$

where f is a local function of ψ^A and a finite number of its derivatives. If $\mathcal{S}$ is functionally differentiable and extremized at the field values ψ^A which are solutions to the field equations of the theory, then $\mathcal{S}$ is referred to as an *action*. We then refer to $\mathcal{L}$ as the Lagrangian density, and the specification of such an $\mathcal{L}$ is what is meant by a Lagrangian formulation of the theory. Note that we may have $\boldsymbol{g} \in \psi$ (*i.e.* the gravitational field may be included in the theory, as in GR), and this is the reason for which we have preferred to simply absorb the $\sqrt{-g}$ (in this case, ψ dependent) factor into $\mathcal{L}$, and thus to write $\mathcal{S}$ as an integral with respect to the flat volume form $\boldsymbol{e}$ instead of the natural spacetime volume form $\boldsymbol{\epsilon}_{\mathscr{M}}$.

All major theories of classical physics, including GR, admit a Lagrangian formulation. In other words, their field equations are equivalent to the extremization of an action $\mathcal{S}[\psi]$ with respect to their physical fields ψ, which in turn can be shown to be equivalent to a system of PDEs known as the *Euler-Lagrange* equations. For field theories of typical interest, including GR and Maxwellian electromagnetism (EM), $\mathcal{L}$ depends on ψ^A and its first derivatives only, *i.e.* $\mathcal{L} = \sqrt{-g} f(\psi^A, \boldsymbol{\nabla}\psi^A)$. In this case, these equations are

$$0 = \frac{\delta \mathcal{S}}{\delta \psi^A} \Leftrightarrow 0 = \frac{\partial \mathcal{L}}{\partial \psi^A} + \frac{1}{2} \frac{\partial \mathcal{L}}{\partial (\nabla_a \psi^A)} \nabla_a \ln(-g) - \nabla_a \left(\frac{\partial \mathcal{L}}{\partial (\nabla_a \psi^A)} \right) . \tag{2.2.7}$$

These are second-order PDEs for ψ^A.

Now fix a coordinate system $\{x^\alpha\} = \{t, x^i\}$. Clearly, all terms which are second order in the derivatives of ψ will emerge from the final term of the above equation: by implicit differentiation, this is

$$\nabla_\alpha \left(\frac{\partial \mathcal{L}}{\partial (\nabla_\alpha \psi^A)} \right) = \frac{\partial^2 \mathcal{L}}{\partial (\nabla_\alpha \psi^A) \partial \psi^B} \nabla_\alpha \psi^B + \frac{\partial^2 \mathcal{L}}{\partial (\nabla_\alpha \psi^A) \partial (\nabla_\beta \psi^B)} \nabla_\alpha \nabla_\beta \psi^B . \tag{2.2.8}$$

The coefficient of the second term on the RHS above is known as the *principal symbol* of the PDE; the $\alpha = t = \beta$ component of this term contains all the second time derivatives of the fields. Thus, the Euler-Lagrange equations have the form

$$0 = W_{AB} \partial_t^2 \psi^B + l_A , \tag{2.2.9}$$

where we have defined

$$W_{AB} = \frac{\partial^2 \mathcal{L}}{\partial (\nabla_t \psi^A) \partial (\nabla_t \psi^B)} , \tag{2.2.10}$$

and l_A indicates lower (*i.e.* first and zeroth) order terms in time derivatives. It is thus apparent that if and only if W_{AB} is non-degenerate in the indices $A, B \in \mathscr{I}$ will we be able to obtain a complete set of solutions to the coupled set of equations, *i.e.* a set of $n = \mathrm{card}(\mathscr{I})$ (the cardinality of $\mathscr{I}$) solutions. If that is the case, we would be able to invert W_{AB} and write the complete system explicitly as

$$0 = \partial_t^2 \psi^B + \left(W^{-1}\right)^{AB} l_A . \tag{2.2.11}$$

If W_{AB} is degenerate, however, then ψ^A and its derivatives up to first order in time and second order in space (and in particular, their initial conditions for the time evolution problem) cannot take arbitrary values. Specifically, they must yield an l_A which is in the image of W_{AB} seen as a linear mapping between vector spaces, the dimension of which is thus less than n. Equivalently, W_{AB} can be seen as a matrix (with indices in $\mathscr{I}$) that has $(n - \mathrm{rank}(W_{AB}))$ null eigenvectors v_{j}^A, with j (in serif font) used to label the set of these eigenvectors (each having components labelled by $A \in \mathscr{I}$). In other words, v_{j}^A are such that $v_{\mathsf{j}}^A W_{AB} = 0$. Multiplying the Euler-Lagrange equation by v_{j}^A on the left thus yields the independent equations,

$$0 = v_{\mathsf{j}}^A l_A . \tag{2.2.12}$$

These equations are known as constraints, since they do not involve second time derivatives of the fields and are thus not regarded as "dynamical" *i.e.* they do not prescribe the time evolution. We will gain a deeper appreciation for what this means once we pass to the Hamiltonian picture, but before we do, let us apply our ideas so far to GR.

2.2.2. Lagrangian formulation of GR. In vacuum, our only field is the gravitational field, $\psi = \boldsymbol{g}$. If, in addition to the requirement that $\mathrm{supp}(\delta \boldsymbol{g}) \subset \mathscr{V}$ we also assume that $\mathrm{supp}(\boldsymbol{\nabla}\delta \boldsymbol{g}) \subset \mathscr{V}$, then the Einstein equation can be recovered fully from an action of the form (2.2.5). In particular,

$$\mathcal{S}_{\mathrm{EH}}\left[\boldsymbol{g}\right] = \int_{\mathscr{V}} \boldsymbol{e}\, \mathcal{L}_{\mathrm{EH}}\left[\boldsymbol{g}\right] \,, \quad \mathcal{L}_{\mathrm{EH}} = \frac{1}{2\kappa}\sqrt{-g}R\,, \tag{2.2.13}$$

are the Einstein-Hilbert action and Lagrangian respectively, with κ denoting the Einstein constant, $\kappa = 8\pi G/c^4 = 8\pi$ in units of $G = 1 = c$. This formulation of GR was first proposed by [Hilbert 1915].

In modern Lagrangian formulations of GR, however, it is typical not to assume anything about the support of $\boldsymbol{\nabla}\delta \boldsymbol{g}$, and in particular its values on the bounday $\partial\mathscr{V}$. Equivalently, only the metric components (and not the derivatives thereof) are to be regarded as being "held fixed on the boundary" when one "varies the action". In this case, one must add a boundary term to (2.2.13) in order to cancel contributions involving $\boldsymbol{\nabla}\delta \boldsymbol{g}$. In particular, this is known as the *Gibbons-Hawking-York boundary term*, first proposed by [York 1972] and later developed by [Gibbons and Hawking 1977]. It is given by the integral of the trace of the extrinsic curvature $\boldsymbol{K}$ of $\partial\mathscr{V}$, $K = \mathrm{tr}(\boldsymbol{K})$. Including this term with the appropriate numerical factor, we will thus henceforth take

$$\mathcal{S}_{\mathrm{G}}\left[\boldsymbol{g}\right] = \mathcal{S}_{\mathrm{EH}} + \frac{1}{\kappa}\int_{\partial\mathscr{V}} \boldsymbol{\epsilon}_{\partial\mathscr{V}}\, K \tag{2.2.14}$$

to be the total gravitational action, *i.e.* the action of GR. For more on this, see also [J. D. Brown, Lau, et al. 2002] and Chapter 12 of [Padmanabhan 2010].

By direct computation, prior to imposing $\delta \boldsymbol{g}|_{\partial\mathscr{V}} = 0$, one finds that for a family $\boldsymbol{g}_{(\lambda)}$ in the sense of the previous subsection,

$$\left(\frac{\partial}{\partial\lambda}\mathcal{S}_{\mathrm{G}}\left[\boldsymbol{g}_{(\lambda)}\right]\right)\Big|_{\lambda=0} = \frac{1}{2\kappa}\left(\int_{\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{M}}\, \boldsymbol{G} : \delta\boldsymbol{g} + \int_{\partial\mathscr{V}} \boldsymbol{\epsilon}_{\partial\mathscr{V}}\, \boldsymbol{\Pi} : \delta\boldsymbol{\gamma}\right) . \tag{2.2.15}$$

Here, $\boldsymbol{G}$ is the Einstein tensor of $\boldsymbol{g}$, $\boldsymbol{\gamma} = \boldsymbol{g}|_{\partial\mathscr{V}}$ is the induced metric on $\partial\mathscr{V}$, and $\boldsymbol{\Pi}$ is its so-called canonical momentum (the nomenclature of which will become clearer when we pass to the Hamiltonian formulation in the next section), given in terms of the extrinsic curvature by $\boldsymbol{\Pi} = \boldsymbol{K} - K\boldsymbol{\gamma}$. Thus, from (2.2.15), the stationary action principle yields the vacuum Einstein equation,

$$\boldsymbol{G} = 0\,, \tag{2.2.16}$$

provided only that $\delta\boldsymbol{\gamma}|_{\partial\mathscr{V}} = 0$.

The appearance of constraints can already be seen manifestly at the level of the full Einstein equation (2.2.16). Choose a coordinate system $\{x^\alpha\} = \{t, x^i\}$. Then, by direct computation (writing the components of the Einstein tensor $G_{\alpha\beta}$ purely in terms of those of the metric $\boldsymbol{g}$), one finds that

$$G^t{}_t = l\,, \tag{2.2.17}$$

$$G^t{}_i = l_i\,, \tag{2.2.18}$$

$$G_{ij} = -\frac{1}{2}g^{tt}\partial_t^2 g_{ij} - \frac{1}{2}g_{ij}(g^{tk}g^{tl} - g^{tt}g^{kl})\partial_t^2 g_{kl} + l_{ij}\,, \tag{2.2.19}$$

where l, l_i and l_{ij} (the l_A in the previous subsection) indicate terms lower than second (*i.e.* first or zeroth) order in time derivatives of $g_{\alpha\beta}$. So we see that the "time-time" and "time-space" Einstein equations, (2.2.17) and (2.2.18) respectively, must be constraints, and the "space-space" Einstein equations (2.2.19) are the only true dynamical equations of motion. This already gives us a crude idea that it is somehow only the "spatial" part of the spacetime metric that plays a dynamical role when the Einstein equation is regarded as a time-dependent problem. To understand what is meant by this precisely, and to gain a proper appreciation for the meaning of the constraints, we must pass to the canonical (Hamiltonian) formulation of the theory.

2.3. Canonical formulation of general field theories

2.3.1. Spacetime foliation. A canonical formulation of a field theory in a spacetime $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ presupposes an assumption called *global hyperbolicity*. There exist a number of equivalent definitions for what this means, but for our purposes the most suggestive one to state is the following: $\mathscr{M}$ is said to be globally hyperbolic if and only if it admits the following topology:

$$\mathscr{M} \simeq \mathbb{R} \times \Sigma\,, \tag{2.3.1}$$

where Σ is a *Cauchy surface*—a closed set which does not intersect its chronological future, and the domain of dependence of which is the entire spacetime. As we wish to refrain from entering here into any further technicalities pertaining to general-relativistic causal structure, a broad and important subject in its own right, we refer the interested reader to [Hawking and Ellis 1975] and Chapter 8 of Ref. [Wald 1984] for precise topological definitions of these terms. Physically, Σ can be thought of as representing the entire (three-) space at a given instant of time.

Thus, given (2.3.1), a canonical formulation must begin with a specification of the meaning of "time" and "change in time". This means that one must specify a choice of a foliation of spacetime into "constant time" Riemannian three-slices ("instants of time"), as well as a "time direction" (indicating how one identifies spatial points on the slices at "different times"). Typically one does this by introducing, respectively, a *time function* $t(x^a)$ on $\mathscr{M}$ such that $\nabla^a t$ is everywhere timelike (which is always possible if $\mathscr{M}$ is globally hyperbolic), along with a *time flow vector field* t^a in $T\mathscr{M}$ such that $t^a \nabla_a t = 1$ (intuitively

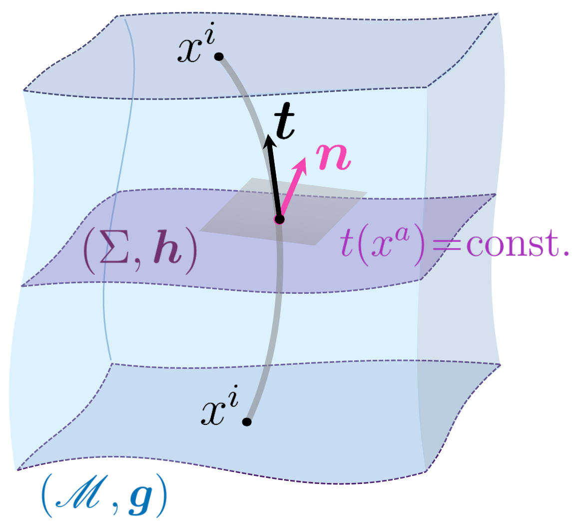

FIGURE 2.2. A depiction (in $(2+1)$ dimensions) of the foliation of a spacetime $(\mathscr{M}, \boldsymbol{g})$ into Cauchy surfaces $(\Sigma, \boldsymbol{h})$, where $\boldsymbol{h}$ is the metric induced on Σ by $\boldsymbol{g}$. These surfaces are defined by the constancy of a time function, $t(x^a) = \text{const.}$, which uniquely determine a normal vector $\boldsymbol{n}$. Additionally, one must define a time flow vector field $\boldsymbol{t}$ on $\mathscr{M}$ the integral curves of which intersect the"same spatial point" (with the same coordinates x^i) on different slices.

ensuring that the interpretation of "time" implied by these two objects is consistent). These are shown in Fig. 2.2.

In this way, the surfaces of constant t in $\mathscr{M}$ are spacelike Cauchy surfaces, and the integral curves of $\boldsymbol{t}$ define a mapping between spatial slices as follows: one identifies the intersections of any particular integral curve of $\boldsymbol{t}$ with all constant t slices as being the "same spatial point" (*i.e.* as corresponding to the same "spatial" coordinate x^i). The condition $\nabla_{\boldsymbol{t}} t = 1$ guarantees that any integral curve of $\boldsymbol{t}$ will intersect any constant t slice exactly once, making the identification well-defined.

A very useful equivalent picture of this construction can be phrased in the language of the theory of embeddings. (See, *e.g.*, [Giulini 2014] for more on this.) In particular, the global hyperbolicity condition (2.3.1) implies the existence of a one-parameter family of embedding maps

$$i_t : \Sigma \to \mathscr{M} \tag{2.3.2}$$

of a ("time-evolving") Cauchy surface Σ into the spacetime ("at different times"), such that $\Sigma_t = i_t(\Sigma) \subset \mathscr{M}$ constitute the (spacelike) Riemannian slices of $\mathscr{M}$. In particular, for any spatial point $p \in \Sigma$, the two spacetime points $i_{t_1}(p) \in \Sigma_{t_1}$ and $i_{t_2}(p) \in \Sigma_{t_2}$ are

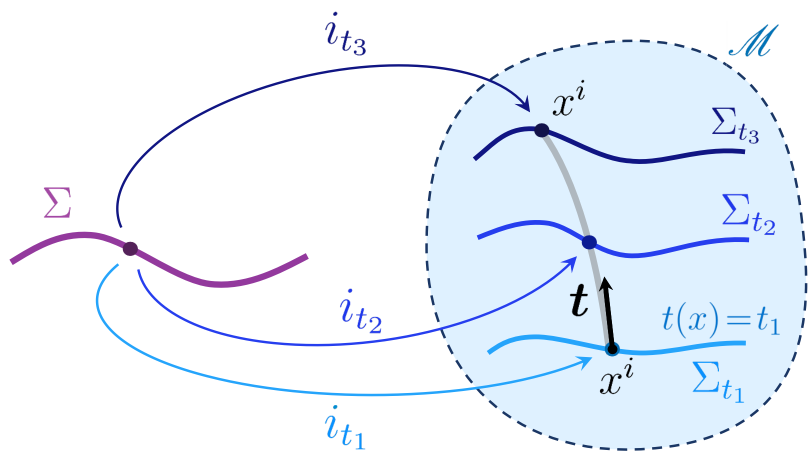

FIGURE 2.3. A depiction (in $(1+1)$ dimensions) of the spacetime $\mathscr{M}$ as constituted by a family of embedded submanifolds Σ_t obtained from an embedding map $i_t : \Sigma \to \mathscr{M}$. Three such submanifolds are shown at three different times, with the time flow vector field identifying the spatial coordinates between them.

identified as the "same spatial point at different times". Specifying such a one-parameter family (2.3.2) of embeddings is mathematically equivalent to specifying a time function and a time flow vector field in $\mathscr{M}$. See Fig. 2.3 for a visual representation.

Before defining additional structures, we can already use the above to specify what we mean precisely by the "time derivative" of a given quantity. Let $\boldsymbol{A} \in \mathscr{T}^k{}_l(\mathscr{M})$ be any tensor in our spacetime. Then we define its time rate of change, or "time derivative," $\dot{\boldsymbol{A}}$ simply as its Lie derivative $\mathcal{L}$ along the time flow vector field $\boldsymbol{t}$,

$$\dot{\boldsymbol{A}} = \mathcal{L}_{\boldsymbol{t}} \boldsymbol{A} \,. \tag{2.3.3}$$

Let n^a be the unit (future-oriented) normal to Σ. Note that this is uniquely determined once $t(x^a)$ is specified. (In particular, $\boldsymbol{n} = -(\boldsymbol{\nabla} t)/\sqrt{(\boldsymbol{\nabla} t)\cdot(\boldsymbol{\nabla} t)}$, with the minus sign ensuring the future orientation.) We thus have a natural spatial volume (three-) form $\epsilon^{\Sigma}_{abc} = \epsilon^{\mathscr{M}}_{abcd} n^d$ induced by the spacetime volume form $\boldsymbol{\epsilon}_{\mathscr{M}}$ (through projection with $\boldsymbol{n}$). However, as in the Lagranigian formulation, it will be convenient to work instead with flat volume forms: in particular, $\boldsymbol{e}$ on $\mathscr{M}$ (as before), and $\mathbf{e}$ on Σ. The latter is notationally distinguished from the former by writing it in upright rather than italic font (though the context usually leaves little danger for confusion), and its components are obtained by projecting $\boldsymbol{e}$ with $\boldsymbol{t}$, *i.e.* $\mathrm{e}_{abc} = e_{abcd} t^d$.

Although we will endeavor to refrain from entering here into excessive geometrical technicalities, a few further definitions and citations of some mathematical results are useful before returning—as we promise we shall, and thereby in a very illustrative way—to physics.

First, let $\boldsymbol{h} = \boldsymbol{g}|_\Sigma$ be the metric induced by $\boldsymbol{g}$ on Σ. In abstract index notation, we could equivalently translate this into the object h_{ij} *i.e.* as a tensor in Σ (an embedded submanifold of $\mathscr{M}$), or as h_{ab}, *i.e.* a tensor on the full spacetime $\mathscr{M}$. In principle and unless otherwise made explicit, we prefer to retain the meaning of all abstract geometric expressions as referent to quantities living in $\mathscr{M}$ (so, *e.g.*, $\boldsymbol{h}$ a priori equates to the spacetime field h_{ab}). One can always project these into any submanifold $\mathscr{U} \subset \mathscr{M}$, in particular by contracting the expressions (over all indices) with the induced metric $\boldsymbol{g}_{\mathscr{U}}$ corresponding thereto, whenever desired. Henceforth we use the notation $(\cdot)|_{\mathscr{U}}$ to refer precisely to such a projection of any quantity $(\cdot)$ onto a submanifold $\mathscr{U}$.

Now, the spatial metric $\boldsymbol{h}$ naturally determines a compatible derivative operator $\boldsymbol{\mathcal{D}}$ on Σ. In turn, $\boldsymbol{\mathcal{D}}$ defines in the usual way the (spatial) Riemann tensor $\mathcal{R}_{abcd}$ of $\boldsymbol{h}$ on Σ, written in calligraphic font to distinguish it from the (Roman font) spacetime Riemann tensor R_{abcd} of $\boldsymbol{g}$. The *extrinsic curvature* (or *second fundamental form*) of Σ is also defined in the usual way, as the derivative of the normal vector; equivalently, it can also be shown to equal half the normal Lie derivative of the spatial metric:

$$K_{ab} = h_{ac} \nabla_b n^c = \frac{1}{2} \mathcal{L}_{\boldsymbol{n}} h_{ab} \, . \tag{2.3.4}$$

Using these definitions, one can prove by direct computation the following relations for the projections onto Σ of the spacetime Riemann tensor and one normal projection of the spacetime Riemann tensor (see, *e.g.*, Chapter 3 of [Bojowald 2011] for the step-by-step computations):

$$R_{abcd}|_\Sigma = \mathcal{R}_{abcd} - K_{ad} K_{bc} + K_{ac} K_{bd} \, , \tag{2.3.5}$$

$$R_{\boldsymbol{n}abc}|_\Sigma = \mathcal{D}_c K_{ab} - \mathcal{D}_b K_{ac} \, . \tag{2.3.6}$$

The first equation (2.3.5) is usually called the *Gauss equation*, and the second equation (2.3.5) is called the *Peterson-Mainardi-Codazzi equation* or (especially common in physics, although historically unfair, as we will shortly clarify) simply the *Codazzi equation*. These are classic results in the theory of embeddings, first discovered in the pioneering days of differential geometry in the early-to-mid 19th century. See [Abbena et al. 2006] for more historical and mathematical details.

The Gauss equation (2.3.5) was first obtained by its eponym [Gauss 1827] in two dimensions. It became known as the *Theorema Egregium* ("remarkable theorem"), and has since then remained one of the most famous results in geometry. It tells us how the curvature—that is, the Riemann tensor—of the embedded surface (Σ) relates to that of the entire manifold ($\mathscr{M}$) through the extrinsic curvature ($\boldsymbol{K}$).

The Peterson-Mainardi-Codazzi equation (2.3.6) was first obtained by [Peterson 1853], and later independently by [Mainardi 1856] and [Codazzi 1868]. It expresses the projection onto the hypersurface (Σ) of one normal projection of the Riemann tensor of the entire manifold ($\mathscr{M}$) in terms of derivatives of the extrinsic curvature ($\boldsymbol{K}$).

These are completely general conditions that are satisfied by any embedding—in this case, of a three-dimensional spacelike hypersurface into a spacetime. It will be especially interesting for our purposes to consider the contracted form of these equations: in particular, by contracting an appropriate pair of spacetime indices in (2.3.5)-(2.3.6) and re-expressing the results in terms of the Einstein tensor $\boldsymbol{G} = \boldsymbol{R} - \frac{1}{2}R\boldsymbol{g}$, one finds by direct computation that

$$2\, G_{\boldsymbol{nn}}|_{\Sigma} = \mathcal{R} - \boldsymbol{K} : \boldsymbol{K} + K^2 \,, \tag{2.3.7}$$

$$G_{\boldsymbol{n}a}|_{\Sigma} = \mathcal{D}^b K_{ab} - \mathcal{D}_a K \,. \tag{2.3.8}$$

We stress once again that these are purely geometrical requirements that the embedding must satisfy. Yet, though we have apparently said nothing so far about physics, the cognizant reader will appreciate that setting the above equations to zero gives precisely the "time-time" and "time-space" Einstein equations in typical canonical form, and we already know from our earlier discussion at the end of the previous section that these contain no second time derivatives. Hence it seems that we have already obtained the canonical constraints of GR without yet essentially doing anything (except anticipating the vacuum Einstein equation) from the point of view of physics! All that we have done is to set up the hypersurface embedding, as is needed for our subsequent canonical formulation (of any field theory).

Finally, it is worth pointing out here one final remark which follows directly from these geometrical identities, and which also sheds some insight into physics. In particular, one can contract the Gauss equation (2.3.5) to express the Ricci scalar R of $\boldsymbol{g}$ in terms of the Ricci scalar $\mathcal{R}$ of $\boldsymbol{h}$ and the extrinsuc curvature $\boldsymbol{K}$ of Σ. The result is [Giulini 2014]:

$$R = \mathcal{R} + \left(\boldsymbol{K} : \boldsymbol{K} - K^2\right) + 2\boldsymbol{\nabla} \cdot \left(K\boldsymbol{n} - \boldsymbol{\nabla}_{\boldsymbol{n}}\boldsymbol{n}\right) \,. \tag{2.3.9}$$

Up to a factor, this is of course simply the Lagrangian of GR in the bulk (*i.e.* the Einstein-Hilbert Lagrangian). As the final term is a divergence and hence will only contribute a boundary term, we see from this that the gravitational action in the bulk is simply:

$$\mathcal{S}_{\mathrm{G}}|_{\mathrm{int}(\mathscr{V})} = \frac{1}{2\kappa}\int_{\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{M}} \left[\mathcal{R} + \left(\boldsymbol{K} : \boldsymbol{K} - K^2\right)\right] \,. \tag{2.3.10}$$

We have, in this way, a heuristic conceptual link to the meaning of the Lagrangian in classical particle mechanics as the "kinetic minus potential energy": the spatial curvature scalar

$\mathcal{R}$ can be regarded as minus the "gravitational potential energy" (so that, the greater the curvature, the greater the magnitude of the "potential energy", negatively-signed overall) and the extrinsic curvature terms ($\boldsymbol{K} : \boldsymbol{K} - K^2$) as the "gravitational kinetic energy" (an analogy that becomes clearer later when we see how the extrinsic curvature is essentially equivalent to the canonical gravitational momentum, such that these are "momentum squared" terms). Of course this analogy is rather vague and not meant to be taken too formally; we will carefully treat the basic questions surrounding notions of "gravitational energy" at the end of this chapter once we have established the full canonical formulation.

2.3.2. Phase space. Now that we have a splitting of our spacetime,

$$\mathscr{M} = \bigcup_t \Sigma_t \simeq \mathbb{R} \times \Sigma \,, \tag{2.3.11}$$

into Cauchy surfaces Σ with all major geometrical constructions that will be needed in place, we may confidently return to physics.

The next step in the canonical formulation of a theory is to introduce the fields by prescribing what is referred to as a *configuration* $\varphi = \{\varphi^A(x^i)\}_A$ on Σ. Physically, this is understood to describe the "instantaneous" configuration of the spacetime fields $\psi(x^a)$, at a particular "time" t (correspondent to a particular Σ_t in the spacetime foliation). Thus it is usually (though by no means necessarily, as one has freedom in how exactly to proceed) defined simply by a direct projection onto Σ of ψ, *i.e.* $\varphi = \psi|_\Sigma$. (In fact, while this is a natural starting point, often this definition does not by itself strictly suffice; in particular, there may be important degrees of freedom lost in the projection, and one must devise a procedure for taking these into account too. As we shall see and elaborate upon, this happens in GR.)

Once the configuration variables φ of the theory have all been defined, one defines

$$\mathscr{Q} = \left\{\varphi\left(x^i\right)\right\} \tag{2.3.12}$$

to be the set of *all possible* (*i.e.*, physically/mathematically permissible) configurations of the collection of fields φ^A. This is a functional space referred to as the *configuration space* of the theory.

Now one must also prescribe what are referred to as the *canonical* (or *conjugate*) *momenta* $\pi = \{\pi_A(x^i)\}_A$ of the fields φ^A, such that π_A is dual to φ^A in all indices. We will see momentarily how the Lagrangian $\mathcal{L}[\psi]$ can be used to devise such a prescription (given the definition of φ). Once this is in hand, the set

$$\mathscr{P} = \{(\varphi, \pi)\} = T^*\mathscr{Q} \tag{2.3.13}$$

of all possible configurations and momenta taken together will simply constitute the cotangent bundle of the configuration space, $T^*\mathscr{Q}$. This is called the *phase space*[2] of the theory.

Phase space (for any field theory) is a particular example of what is referred to as a *symplectic manifold*. Such objects have been extensively studied by geometers, and today symplectic geometry is a broad and fruitful area of mathematics in its own right. (See, *e.g.*, the reviews/books [Silva 2008, 2006; Hofer and Zehnder 2011].) Historically, this field originated precisely from the advent of classical Hamiltonian particle mechanics (much as variational calculus and functional analysis were heavily precipitated by classical Lagrangian particle mechanics), and so it is worth our time to briefly offer a description of phase space in symplectic language before moving on to formulate the Hamiltonian equations of motion.

In general, a symplectic manifold $(\mathscr{W};\boldsymbol{\omega})$ is any $2n$-dimensional manifold $\mathscr{W}$ equipped with a two-form $\boldsymbol{\omega}$, called a *symplectic form*, provided that the latter satisfies two conditions: *(i)* $\boldsymbol{\omega}$ is closed (*i.e.* $\mathrm{d}_{\mathscr{W}}\boldsymbol{\omega} = 0$, where $\mathrm{d}_{\mathscr{W}}$ is the exterior derivative on $\mathscr{W}$); *(ii)* $\boldsymbol{\omega}$ is non-degenerate (*i.e.* at any point $p \in \mathscr{W}$ and for any vectors $\boldsymbol{X}, \boldsymbol{Y} \in T_p\mathscr{W}$, if $\imath_{\boldsymbol{X}}\imath_{\boldsymbol{Y}}\boldsymbol{\omega}_p = 0$ $\forall \boldsymbol{Y} \in T_p\mathscr{W}$ where $\imath$ is the interior product on $\mathscr{W}$, then $\boldsymbol{X} = 0$).

Having claimed that the phase space $\mathscr{P} = \{(\varphi^A(x^i), \pi_A(x^i))\}$ is an example of a symplectic manifold, we ought to show it by producing the symplectic form. In order to do this, we first need therefore to further clarify what we mean by an exterior derivative $\mathrm{d}_{\mathscr{P}}$ on $\mathscr{P}$. Because we are dealing here with functional spaces, we must use the *functional exterior derivative* on $\mathscr{P}$ which we denote by $\delta = \mathrm{d}_{\mathscr{P}}$ (not to be confounded in meaning with the $\delta\psi^A$ on $\mathscr{M}$ from the Lagrangian analysis); see [Crnković 1987; Crnković and Witten 1989] for more technical details on this. For example, $\delta\varphi^A(x^i)$ and $\delta\pi_A(x^i)$ are one-forms on $\mathscr{P}$, and so for any functional (zero-form) $F[\varphi^A, \pi_A]$ on $\mathscr{P}$, for example, we have that the action of δ yields the one-form

$$\delta F\,[\varphi, \pi] = \int_\Sigma \mathbf{e} \left(\frac{\delta F}{\delta\varphi^A(x^i)}\delta\varphi^A(x^i) + \frac{\delta F}{\delta\pi_A(x^i)}\delta\pi_A(x^i) \right) , \tag{2.3.14}$$

where $\delta F/\delta f(x^i)$ indicates the functional derivative of F, as defined in the previous section and restricted to functionals on Σ. A wedge product can be naturally defined to obtain p-forms, and the action of δ also easily generalizes thereto.

We are now ready to write down the symplectic form $\boldsymbol{\omega}$ for our phase space $(\mathscr{P};\boldsymbol{\omega})$. It is possible to show (a result known generally as the Darboux theorem) that, at least

[2] The origin of the nomenclature is from statistical mechanics, where many of these methods were first developed. See [Davies 1977; Sklar 1995; H. R. Brown et al. 2009] for good historical accounts.

locally, $\boldsymbol{\omega}$ is always given by:

$$\boldsymbol{\omega} = \int_{\Sigma} \mathbf{e}\, \delta\pi_A \wedge \delta\varphi^A \,, \tag{2.3.15}$$

which can be checked to satisfy the symplectic form conditions[3]. In the case that there is only one (tensorial) field variable in φ, it can furthermore be proved that the symplectic form $\boldsymbol{\omega}$ is also the volume form $\boldsymbol{\epsilon}_{\mathscr{P}}$, conventionally denoted as $\boldsymbol{\epsilon}_{\mathscr{P}} = \boldsymbol{\Omega}$, on $\mathscr{P}$, that is to say, we have $\boldsymbol{\Omega} = \boldsymbol{\omega}$. If there are N fields in φ, then a volume form $\boldsymbol{\Omega}$ on $\mathscr{P}$ is given simply by the N-th exterior power of the symplectic form, in particular $\boldsymbol{\Omega} = [(-1)^{N(N-1)/2}/N!]\boldsymbol{\omega}^{\wedge N}$.

We will need just a few more definitions and results before proceeding to the equations of motion. Let $\mathscr{F}(\mathscr{P}) = \{F : \mathscr{P} \to \mathbb{R}\}$ be the set of all functionals (zero-forms) on the phase space $(\mathscr{P}; \boldsymbol{\omega})$. For any $F[\varphi^A, \pi_A] \in \mathscr{F}(\mathscr{P})$, the symplectic form $\boldsymbol{\omega}$ defines a vector field $\boldsymbol{X}_F \in T\mathscr{P}$ associated to F, referred to as the *Hamiltonian vector field* (HVF) of F, via the relation

$$\imath_{\boldsymbol{X}_F} \boldsymbol{\omega} = -\delta F \,, \tag{2.3.16}$$

where $\imath$ is the interior product on $\mathscr{P}$. We furthermore define an operation $\{\cdot, \cdot\} : \mathscr{F} \times \mathscr{F} \to \mathbb{R}$ known as the *Poisson bracket*, which for any two functionals $F, G \in \mathscr{F}$ is given by:

$$\{F, G\} = \int_{\Sigma} \mathbf{e} \left(\frac{\delta F}{\delta\varphi^A(x^i)} \frac{\delta G}{\delta\pi_A(x^i)} - \frac{\delta F}{\delta\pi_A(x^i)} \frac{\delta G}{\delta\varphi^A(x^i)} \right) . \tag{2.3.17}$$

Finally, let $\boldsymbol{Y} \in T\mathscr{P}$ be any vector field in $\mathscr{P}$. We can use it to define the action of a map $\Phi_t^{(\boldsymbol{Y})} : \mathscr{P} \times \mathbb{R} \to \mathscr{P}$ by requiring that $\Phi_t^{(\boldsymbol{Y})}$ moves points around in $\mathscr{P}$ along the integral curves of $\boldsymbol{Y}$ (as a function of the parameter $t \in \mathbb{R}$). Mathematically, this means that such a map is defined by the ordinary differential equation:

$$\frac{\mathrm{d}\Phi_t^{(\boldsymbol{Y})}}{\mathrm{d}t} = \boldsymbol{Y} \circ \Phi_t^{(\boldsymbol{Y})} \,. \tag{2.3.18}$$

Any map satisfying (2.3.18) is called a *flow*, and $\boldsymbol{Y}$ is called its (infinitesimal) *generator*. If, additionally, $\Phi_t^{(\boldsymbol{Y})}$ preserves the symplectic form under the pullback, *i.e.* if $(\Phi_t^{(\boldsymbol{Y})})^* \boldsymbol{\omega} = \boldsymbol{\omega}$, then $\Phi_t^{(\boldsymbol{Y})}$ is called a *canonical transformation* (or, in geometry, a *symplectomorphism*). It

[3] (φ, π) are then referred to as Darboux coordinates, symplectic coordinates, or canonical coordinates.

is useful and easy to prove that $\Phi_t^{(\boldsymbol{Y})}$ is a canonical transformation if and only if $\boldsymbol{Y}$ is an HVF[4].

2.3.3. The Hamiltonian and the equations of motion. We define the *Hamiltonian functional* $H \in \mathscr{F}(\mathscr{P})$, henceforth simply the *Hamiltonian*, to be a phase space functional of the form

$$H\left[\varphi^A, \pi_A\right] = \int_\Sigma \mathbf{e}\,\mathcal{H}\left(\varphi^A, \pi_A\right) , \tag{2.3.19}$$

where $\mathcal{H}$ is a local function of (φ^A, π_A), referred to as the *Hamiltonian density*—which, provided no confusion is created, we will also simply call the Hamiltonian—such that the field equations of the original spacetime field theory (for ψ in $\mathscr{M}$) are equivalent to the *canonical* (or *Hamiltonian*) *equations of motion* for the phase space variables,

$$\dot{\varphi}^A = \left\{\varphi^A, H\right\} = \frac{\delta H}{\delta \pi_A} , \tag{2.3.20}$$

$$\dot{\pi}^A = \left\{\pi^A, H\right\} = -\frac{\delta H}{\delta \varphi^A} , \tag{2.3.21}$$

with the last equality in each line following from the general definition of the Poisson bracket (2.3.17).

More generally, the time derivative (Lie derivative along the time flow vector field) of any functional $F \in \mathscr{F}(\mathscr{P})$ is given by the Poisson bracket with the Hamiltonian H, which we can write as

$$\dot{F} = \{F, H\} = \boldsymbol{X}_H(F) , \tag{2.3.22}$$

where $\boldsymbol{X}_H \in T\mathscr{P}$ is the HVF of the Hamiltonian,

$$\boldsymbol{X}_H = \int_\Sigma \mathbf{e} \left(\frac{\delta H}{\delta \pi_A} \frac{\delta}{\delta \varphi^A} - \frac{\delta H}{\delta \varphi^A} \frac{\delta}{\delta \pi_A} \right) . \tag{2.3.23}$$

Put differently, the time evolution of any quantity through $\mathscr{P}$ is represented by the integral curves of the HVF of the Hamiltonian, $\boldsymbol{X}_H$. We refer to the flow $\Phi_t^{(\boldsymbol{X}_H)} : \mathscr{P} \times \mathbb{R} \to \mathscr{P}$ generated thereby as the *Hamiltonian flow*, and for simplicity we will henceforth denote it simply as $\Phi_t = \Phi_t^{(\boldsymbol{X}_H)}$. Following our discussion at the end of the last subsection, the fact that $\boldsymbol{X}_H$ is an HVF guarantees that Φ_t is a canonical transformation. In particular, it

[4] Let $\Phi_t^{(\boldsymbol{Y})}$ be such that $(\Phi_t^{(\boldsymbol{Y})})^* \boldsymbol{\omega} = \boldsymbol{\omega}$. Equivalently, $\mathcal{L}_{\boldsymbol{Y}} \boldsymbol{\omega} = 0$. Recall Cartan's "magic formula," which tells us that the action of the Lie derivative when acting on forms can be expressed as $\mathcal{L}_{\boldsymbol{Y}} = \imath_{\boldsymbol{Y}} \circ \mathrm{d} + \mathrm{d} \circ \imath_{\boldsymbol{Y}}$. In our case, as we have seen, d is the functional exterior derivative δ, so we have $0 = \mathcal{L}_{\boldsymbol{Y}} \boldsymbol{\omega} \Leftrightarrow 0 = \imath_{\boldsymbol{Y}}(\delta\boldsymbol{\omega}) + \delta(\imath_{\boldsymbol{Y}} \boldsymbol{\omega})$. But $\boldsymbol{\omega}$ is closed (*i.e.* $\delta\boldsymbol{\omega} = 0$, the first symplectic form property), so this is equivalent to $0 = \delta(\imath_{\boldsymbol{Y}} \boldsymbol{\omega}) \Leftrightarrow \imath_{\boldsymbol{Y}} \boldsymbol{\omega} = \delta F$ for some $F \in \mathscr{F}(\mathscr{P})$. This is the same as the definition (2.3.16), *i.e.* it is equivalent to saying that $\boldsymbol{Y}$ is an HVF (specifically, $\boldsymbol{Y} = \boldsymbol{X}_{-F}$).

preserves the symplectic (as well as volume) form:

$$\mathcal{L}_{\boldsymbol{X}_H}\boldsymbol{\omega} = 0 = \mathcal{L}_{\boldsymbol{X}_H}\boldsymbol{\Omega}\,. \tag{2.3.24}$$

This result is commonly known as is *Louville's theorem.*

We have expended much effort so far on developing the technical machinery for a canonical analysis without yet prescribing the recipe for explicitly computing the most important pieces: the canonical momenta and the Hamiltonian itself! Assuming a Lagrangian formulation of the theory exists, the definition typically ascribed to the former is:

$$\pi_A = \frac{\partial \mathcal{L}}{\partial \dot{\varphi}^A}\,, \tag{2.3.25}$$

and to the latter:

$$\mathcal{H} = \dot{\varphi}^A \pi_A - \mathcal{L}\,, \tag{2.3.26}$$

called the *Legendre transform* of the Lagrangian.

Let us consider in turn the subtleties presented by these definitions. First, the definition of the canonical momenta (2.3.25) may be seen as expressing π_A as a function of $(\varphi^A, \dot{\varphi}^A)$, given explicitly by the $\dot{\varphi}^A$ partial of $\mathcal{L}(\varphi^A, \dot{\varphi}^A)$ on the RHS. A one-to-one correspondence between the configuration time derivatives $\dot{\varphi}^A$ and the momenta π_A exists if and only if it is possible to invert this mapping, *i.e.* to write all $\dot{\varphi}^A$ as functions of π_A (and possibly φ^A). If this is *not* possible (as we will see in GR), then some of the $\dot{\varphi}^A$ will not represent true "dynamical" degrees of freedom, but instead will define the constraints.

Let us see how this works in a bit more detail. Let $f : T\mathscr{Q} \to T^*\mathscr{Q}$ denote the mapping taking points from the set $\{(\varphi, \dot{\varphi})\}$, which is simply the tangent space $T\mathscr{Q}$ of the configuration space, into the phase space variables $\{(\varphi, \pi)\}$ according to the rule (2.3.25), *i.e.*

$$\begin{aligned} f : T\mathscr{Q} \to & T^*\mathscr{Q} \\ \left(\varphi^A, \dot{\varphi}^A\right) \mapsto & \left(\varphi^A, \pi_A\left(\varphi^A, \dot{\varphi}^A\right)\right) = \left(\varphi^A, \partial_{\dot{\varphi}^A}\mathcal{L}\left(\varphi^A, \dot{\varphi}^A\right)\right)\,. \end{aligned} \tag{2.3.27}$$

The inverse function theorem tells us that for $f^{-1} : T^*\mathscr{Q} \to T\mathscr{Q}$ to exist, we must have $\det(\mathrm{Jac}(f)) \neq 0$. In particular, this requires $\det(W_{AB}) \neq 0$, *i.e.* the non-degeneracy of the matrix W_{AB} given by

$$W_{AB} = \frac{\partial \pi_A}{\partial \dot{\varphi}^B} = \frac{\partial^2 \mathcal{L}}{\partial \dot{\varphi}^B \partial \dot{\varphi}^A}\,, \tag{2.3.28}$$

where in the last equality we have used the momentum definition (2.3.25). It should not surprise the reader that this is essentially the same as the principal symbol of the general Euler-Lagrange equations (2.2.10) that we encountered earlier!

Thus, if the function (2.3.27) does not have an inverse (that is, it has an inverse only on a restriction of its domain), or equivalently the matrix (2.3.28) is degenerate, then not all n

of the $\dot{\varphi}^A$ can be solved for in terms of the π_A, and those which cannot will consequently avoid picking up an additional time derivative in the equations of motion. These therefore define constraints on the second-order equations for the true "dynamical" degrees of freedom in (φ^A, π_A).

Suppose f has $\tilde{m}$ degeneracy directions, *i.e.* $\tilde{m}$ of the $\dot{\varphi}^A$ map trivially onto the π_A. This is equivalent to the existence of $\tilde{m}$ phase space functionals $\tilde{\zeta}_{\tilde{\mathrm{j}}} \in \mathscr{F}(\mathscr{P})$, $\forall 1 \leq \tilde{\mathrm{j}} \leq \tilde{m}$, which identically vanish for solutions satisfying the equations of motion of the theory, *i.e.*

$$0 = \tilde{\zeta}_{\tilde{\mathrm{j}}} \,. \tag{2.3.29}$$

Such constraints are called *primary* constraints.

As we will see (and as the notation using tildes anticipates), these may actually not be the only phase space constraints for our theory; in particular, consistency conditions involving the primary constraints $\tilde{\zeta}_{\tilde{\mathrm{j}}}$ may imply additional, *independent* constraints—and we shall concretely see how so in the next subsection. For the moment, let us simply assume henceforth that *all* the constraints of the theory—however they are obtained—are collected into the set $\zeta = \{\zeta_{\mathrm{j}}\}_{\mathrm{j}}$, with the index j here running over a possibly larger range than just from 1 to $\tilde{m}$, and with the first $\tilde{m}$ of them being the primary constraints just discussed (and notationally identified with tildes), *i.e.* $\zeta_{\mathrm{j}} = \tilde{\zeta}_{\tilde{\mathrm{j}}}$ for $1 \leq \mathrm{j} = \tilde{\mathrm{j}} \leq \tilde{m}$.

The $\tilde{\zeta}_{\tilde{\mathrm{j}}}$ may be regarded as coordinates locally orthogonal to the image of the function (2.3.27), such that locally $\mathscr{P}$ has a complete set of coordinates given by $(\varphi^A, \pi_A, \tilde{\zeta}_{\tilde{\mathrm{j}}})$. This is illustrated visually in Fig 2.4. For convenience, we henceforth redefine the set π in $\mathscr{P}$ to include not only those momenta π_A which can be solved for, but also the primary constraints $\tilde{\zeta}_{\tilde{\mathrm{j}}}$, *i.e.* $\pi = \{\pi_A, \tilde{\zeta}_{\tilde{\mathrm{j}}}\}$, such that an arbitrary point in $\mathscr{P}$ is still labeled as (φ, π). Henceforth, though we will continue to simply call it the "phase space" if the context is clear enough, we will formally refer to $\mathscr{P} = \{(\varphi, \pi)\}$ as the *unconstrained*, or *full phase space.*

A few more definitions follow naturally from these considerations and will be useful for us to establish here before moving on. We offer in Fig. 2.5 a visual depiction to make the story a little bit easier to follow.

We define the *primary constraint surface* $\tilde{\mathscr{C}} \subseteq \mathscr{P}$ as the submanifold of the phase space where only the primary constraints $\tilde{\zeta}_{\tilde{\mathrm{j}}}$ vanish (with no conditions assumed on non-primary constraints ζ_{j} for $\mathrm{j} \geq \tilde{m} + 1$, if any exist):

$$\tilde{\mathscr{C}} = \left\{ (\varphi, \pi) \in \mathscr{P} \middle| \, 0 = \tilde{\zeta} \right\} . \tag{2.3.30}$$

Meanwhile, the *constraint surface* or, for emphasis, *full constraint surface* $\mathscr{C} \subseteq \tilde{\mathscr{C}}$ is the submanifold of $\mathscr{P}$ where *all* the constraints (the primary $\tilde{\zeta}$ plus any other constraints, ζ in total) are satisfied:

$$\mathscr{C} = \{ (\varphi, \pi) \in \mathscr{P} | \, 0 = \zeta \} . \tag{2.3.31}$$

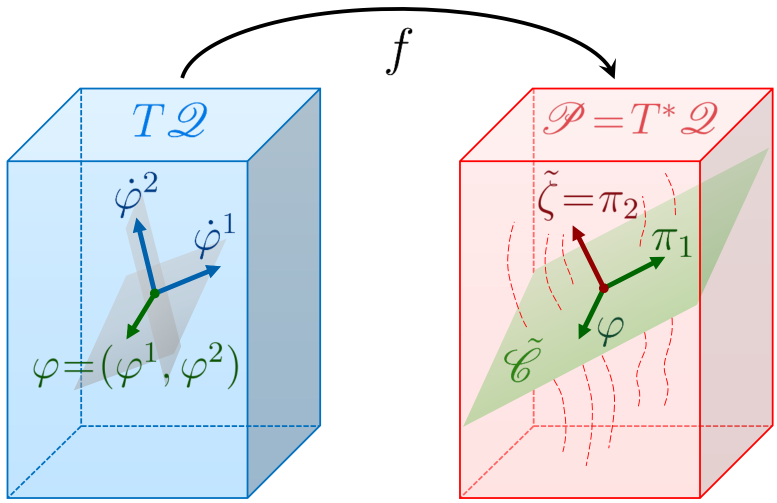

FIGURE 2.4. An illustration of the transformation $f : T\mathscr{Q} \to T^*\mathscr{Q}$ from the configuration space tangent bundle into the phase space $\mathscr{P} = T^*\mathscr{Q}$. For example, suppose we have a two-dimensional configuration $\varphi = \{\varphi^1, \varphi^2\}$ (visually represented as one dimension), and correspondingly $\dot{\varphi} = \{\dot{\varphi}^1, \dot{\varphi}^2\}$. Suppose that here, $\dot{\varphi}^2$ is in the kernel of this map, *i.e.* it maps trivially to π_2 such that the only primary constraint is $0 = \pi_2 = \tilde{\zeta}$. The primary constraint surface $\tilde{\mathscr{C}}$ thus has coordinates $\{\varphi, \pi_1\}$.

We will accordingly also find it useful to define operations of equality under primary and full constraint imposition. In particular, we use the symbols "$\tilde{=}$" and "$\mathring{=}$" respectively to indicate these, such that $a \mathrel{\tilde{=}} b$ means that $a|_{\tilde{\mathscr{C}}} = b|_{\tilde{\mathscr{C}}}$, and $a \mathrel{\mathring{=}} b$ that $a|_{\mathscr{C}} = b|_{\mathscr{C}}$.

Now, let us turn to a discussion of the definition of the Hamiltonian (2.3.26) as a Legendre transform of $\mathcal{L} : T\mathscr{Q} \to \mathbb{R}$. In particular, we have—just as in the case of the momentum definition (2.3.25)—a functional prescription in terms of $(\varphi, \dot{\varphi})$ (*i.e.*, as a functional on the tangent bundle $T\mathscr{Q}$). In other words, the definition gives us $\mathcal{H}[\varphi, \dot{\varphi}]$. How can we know, in general, that this $\mathcal{H}$ is also—as it ought to be for this procedure to make sense—a well-defined functional $\mathcal{H}[\varphi, \pi]$ on the phase space (cotangent bundle) $\mathscr{P} = T^*\mathscr{Q}$, *irrespective* of the existence of constraints? (I.e, how can we know that under the mapping $f : T\mathscr{Q} \to T^*\mathscr{Q}$ given by Eqn. (2.3.27), the transformation of $\mathcal{H}$ always "avoids" any degeneracy directions that may arise from its non-invertibility?) An easy way to see this is by computing the exterior derivative of the Legendre transform $\mathcal{H} = \dot{\varphi}^A \pi_A - \mathcal{L}$ (2.3.26),

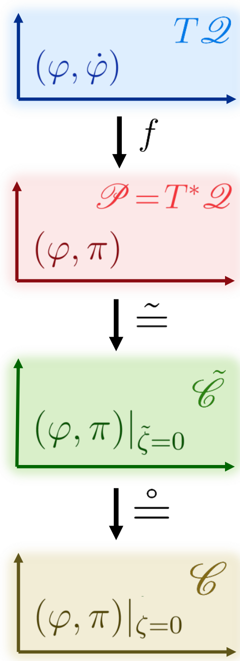

FIGURE 2.5. An illustration of the configuration space tangent bundle $T\mathscr{Q}$, the phase space $\mathscr{P} = T^*\mathscr{Q}$, the primary constraint surface $\tilde{\mathscr{C}}$, and the full constraint surface $\mathscr{C}$ with their respective coordinates, and the maps/operations that respectively transform from one of these spaces to the next.

immediately yielding

$$\delta\mathcal{H} = \int_\Sigma \mathbf{e}\left[\dot{\varphi}^A\delta\pi_A + \pi_A\delta\dot{\varphi}^A - \frac{\partial\mathcal{L}}{\partial\varphi^A}\delta\varphi^A - \frac{\partial\mathcal{L}}{\partial\dot{\varphi}^A}\delta\dot{\varphi}^A\right] \tag{2.3.32}$$

$$= \int_\Sigma \mathbf{e}\left[-\frac{\partial\mathcal{L}}{\partial\varphi^A}\delta\varphi^A + \dot{\varphi}^A\delta\pi_A\right], \tag{2.3.33}$$

where to obtain the last equality the momentum definition $\pi_A = \partial\mathcal{L}/\partial\dot{\varphi}^A$ (2.3.25) was once again used, such that no term linear in $\delta\dot{\varphi}$ remains. Hence the exterior derivative (2.3.33) of the Hamiltonian functional $\mathcal{H}$ is always a linear combination *only* of $\delta\varphi^A$ and $\delta\pi_A$, making it a well-defined one-form on $\mathscr{P}$, implying that $\mathcal{H}$ itself is a well-defined zero-form (functional) on $\mathscr{P}$.

Now we must verify that this $\mathcal{H} = \dot{\varphi}^A\pi_A - \mathcal{L}$ indeed gives the correct (canonical) equations of motion (2.3.20)-(2.3.21) for the theory. For their formulation, it will in fact be more convenient to first define what is referred to as the *total* Hamiltonian, which we denote as $\tilde{\mathcal{H}} \in \mathscr{F}(\mathscr{P})$, given by:

$$\tilde{\mathcal{H}} = \mathcal{H} - \tilde{\lambda}^{\tilde{\mathsf{j}}}\tilde{\zeta}_{\tilde{\mathsf{j}}}\,, \quad \tilde{H} = \int_\Sigma \mathbf{e}\,\tilde{\mathcal{H}}\,. \tag{2.3.34}$$

This is simply the Hamiltonian $\mathcal{H}$ we have been working with so far (obtained via the Legendre transform), plus an arbitrary linear combination of the primary constraints $\tilde{\zeta}_{\tilde{\mathsf{j}}}$, with $\tilde{m}$ coefficient functions $\tilde{\lambda}^{\tilde{\mathsf{j}}} \in \mathscr{F}(\mathscr{P})$ playing the role of Lagrange multipliers. We shall see the usefulness of this momentarily. We remark for now that by construction, $\tilde{\mathcal{H}}$ will coincide with $\mathcal{H}$ on the (primary and thus also full) constraint surface, $\tilde{\mathcal{H}} \mathrel{\tilde{\approx}} \mathcal{H} \mathrel{\mathring{=}} \tilde{\mathcal{H}}$.

Combining $\mathcal{H} = \dot{\varphi}^A\pi_A - \mathcal{L}$ (2.3.26) and $\tilde{\mathcal{H}} = \mathcal{H} - \tilde{\lambda}\cdot\tilde{\zeta}$ (2.3.34), one can isolate for the Lagrangian as: $\mathcal{L} = \dot{\varphi}^A\pi_A - \tilde{\mathcal{H}} - \tilde{\lambda}\cdot\tilde{\zeta}$. Integrating both sides over (a finite range of) t, and then taking their variation (*i.e.* viewing the arguments of both sides as one-parameter families in λ, in the sense of the previous section, and applying $\partial_\lambda|_{\lambda=0}$), a straightforward computation shows that the stationary action principle $0 = \delta\mathcal{S}[\psi] = \int \mathrm{d}t\mathbf{e}\,\mathcal{L}$ is equivalent to the following system of equations in $\mathscr{P}$:

$$\dot{\varphi}^A = \left\{\varphi^A, \tilde{H}\right\} + \tilde{\zeta}_{\tilde{\mathsf{j}}}\frac{\delta\tilde{\lambda}^{\tilde{\mathsf{j}}}}{\delta\pi_A}\,, = \frac{\delta\tilde{H}}{\delta\pi_A} + \tilde{\zeta}_{\tilde{\mathsf{j}}}\frac{\delta\tilde{\lambda}^{\tilde{\mathsf{j}}}}{\delta\pi_A}\,, \tag{2.3.35}$$

$$\dot{\pi}^A = \left\{\pi^A, \tilde{H}\right\} - \tilde{\zeta}_{\tilde{\mathsf{j}}}\frac{\delta\tilde{\lambda}^{\tilde{\mathsf{j}}}}{\delta\varphi^A}\,, = -\frac{\delta\tilde{H}}{\delta\varphi^A} - \tilde{\zeta}_{\tilde{\mathsf{j}}}\frac{\delta\tilde{\lambda}^{\tilde{\mathsf{j}}}}{\delta\varphi^A}\,. \tag{2.3.36}$$

So on the constraint surface $\mathscr{C}$, we indeed recover the standard form of the canonical equations of motion in terms of the total Hamiltonian:

$$\dot{\varphi}^A \mathrel{\mathring{=}} \left\{\varphi^A, \tilde{H}\right\}\,, \tag{2.3.37}$$

$$\dot{\pi}^A \mathrel{\mathring{=}} \left\{\pi^A, \tilde{H}\right\}\,. \tag{2.3.38}$$

2.3.4. Constraints and the reduced phase space. Let us now address the subtleties that the presence of constraints poses to our Hamiltonian analysis.

First, suppose that all primary constraints are indeed satisfied, *i.e.* $\tilde{\zeta}_{\tilde{\mathsf{j}}} = 0$, so that we are on $\tilde{\mathscr{C}}$ in $\mathscr{P}$. For ease in following the discussion, consulting again Figs. 2.4 and 2.5 is useful. Observe that the vanishing of $\tilde{\zeta}$ on $\tilde{\mathscr{C}}$ necessarily implies that their time derivatives $\dot{\tilde{\zeta}}$ vanish thereon too. These are called *consistency conditions*:

$$0 \mathrel{\tilde{\approx}} \dot{\tilde{\zeta}}_{\tilde{\mathsf{j}}} \mathrel{\mathring{=}} \left\{\tilde{\zeta}_{\tilde{\mathsf{j}}}, \tilde{H}\right\} \mathrel{\mathring{=}} \left\{\tilde{\zeta}_{\tilde{\mathsf{j}}}, H\right\} - \tilde{\lambda}^{\tilde{\mathsf{k}}}\left\{\tilde{\zeta}_{\tilde{\mathsf{j}}}, \tilde{\zeta}_{\tilde{\mathsf{k}}}\right\}\,, \tag{2.3.39}$$

where in the last equality we have simply used the definition of the total Hamiltonian $\tilde{\mathcal{H}} = \mathcal{H} - \tilde{\lambda}\cdot\tilde{\zeta}$ (2.3.34).

From this we see that the multiplicative functions (Lagrange multipliers) $\tilde{\lambda}^{\tilde{\mathsf{j}}}$ can be fully determined (*i.e.* $\tilde{m}$ independent equations for them can be obtained by inverting the above equation) if and only if $\det(\{\tilde{\zeta}_{\tilde{\mathsf{j}}}, \tilde{\zeta}_{\tilde{\mathsf{k}}}\}) \neq 0$. Otherwise, the system formed by the vanishing of the primary constraints coupled with the consistency conditions (2.3.39) is under-determined, and therefore has to be augmented by further equations, which we can now interpret as constituting the additional constraints in ζ_{j} (those for $\mathsf{j} \geq \tilde{m}+1$).

To obtain these equations, let $\{\hat{v}^{\tilde{\mathsf{j}}}_{\hat{\mathsf{j}}}\}$ be the set of $\hat{m} \leq \tilde{m}$ null-eigenvectors, indexed in the set by $1 \leq \hat{\mathsf{j}} \leq \hat{m}$ (and with components indexed by $\tilde{\mathsf{j}}$), of the matrix $\{\tilde{\zeta}_{\tilde{\mathsf{j}}}, \tilde{\zeta}_{\tilde{\mathsf{k}}}\}$, *i.e.* we have $0 = \hat{v}^{\tilde{\mathsf{j}}}_{\hat{\mathsf{j}}}\{\tilde{\zeta}_{\tilde{\mathsf{j}}}, \tilde{\zeta}_{\tilde{\mathsf{k}}}\}$. Multiplying the consistency condition (2.3.39) on the left by $\hat{v}^{\tilde{\mathsf{j}}}_{\hat{\mathsf{j}}}$ yields the *independent* equations

$$0 \mathrel{\hat{=}} \hat{v}^{\tilde{\mathsf{j}}}_{\hat{\mathsf{j}}} \left\{\tilde{\zeta}_{\tilde{\mathsf{j}}}, H\right\} = \hat{\zeta}_{\hat{\mathsf{j}}}\,, \tag{2.3.40}$$

which we define as $\hat{\zeta}$. These $\hat{m}$ equations are referred to as *secondary* constraints.

Assuming that with these, one now obtains a complete system, then one takes the complete set of $(\tilde{m}+\hat{m})$ constraints ζ_{j} to be simply the set of the $(\tilde{m})$ primary and $(\hat{m})$ secondary constraints $\zeta = \{\tilde{\zeta}, \hat{\zeta}\}$. (In the index notation, we set $\zeta_{\tilde{m}+\mathsf{k}} = \hat{\zeta}_{\hat{\mathsf{k}}}$, $\forall 1 \leq \mathsf{k} = \hat{\mathsf{k}} \leq \hat{m}$.) This happens to be the case in GR (which, as we shall see, has a total of eight constraints, with $\tilde{m}=4$ primary constraints and $\hat{m}=4$ secondary constraints)[5].

Now we turn to a different and also very useful way of looking at how to classify the (total set of) constraints ζ. It may be motivated by reflecting upon the following question: assuming we are indeed only interested in those field configutations which satisfy the constraints, *i.e.* only in points (φ, π) living on $\mathscr{C}$, could we simply regard $\mathscr{C}$ in some operational sense as an "effective" phase space? The answer in general is *no*: in particular, $\mathscr{C}$ will *not* in general be a symplectic manifold.

To be more precise, let $i : \mathscr{C} \to \mathscr{P}$ be the embedding of $\mathscr{C}$ in $\mathscr{P}$. We denote by $\boldsymbol{\omega}|_{\mathscr{C}} = i^*\boldsymbol{\omega}$ the pullback of the symplectic form $\boldsymbol{\omega}$ of the phase space $\mathscr{P}$ to the constraint surface $\mathscr{C}$. In general, as we will show presently, $\boldsymbol{\omega}|_{\mathscr{C}}$ is unfortunately *not* a symplectic form on $\mathscr{C}$. Thus $\boldsymbol{\omega}|_{\mathscr{C}}$ is sometimes instead referred to as the *presymplectic* form.

In order to make progress, the following classification of constraint functions $\zeta \in \mathscr{F}(\mathscr{P})$ is useful.

- A constraint $\zeta \in \mathscr{F}(\mathscr{P})$ is called a *first-class constraint* if its HVF $\boldsymbol{X}_\zeta$ is *everywhere* tangent to $\mathscr{C}$ (*i.e.* $\boldsymbol{X}_\zeta \in T\mathscr{C}$).
- A constraint $\zeta \in \mathscr{F}(\mathscr{P})$ is called a *second-class constraint* if its HVF $\boldsymbol{X}_\zeta$ is *nowhere* tangent to $\mathscr{C}$ (*i.e.* $\nexists p \in \mathscr{P}$ so that $(\boldsymbol{X}_\zeta)_p \in T_p\mathscr{C}$).

Now suppose that among the constraint functions ζ in our field theory, say $\zeta_1 = f$, is first-class. One finds that the interior product between its HVF and the presymplectic form vanishes: that is, $\imath_{\boldsymbol{X}_f}\boldsymbol{\omega}|_{\mathscr{C}} = i^*(\imath_{\boldsymbol{X}_f}\boldsymbol{\omega}) = i^*(\boldsymbol{\nabla} f) = 0$, using the definition of the

[5] If, on the other hand, one still does not have a complete set of equations by arriving at the secondary constraints, the same process must be repeated until one does: the secondary constraints imply their own consistency conditions, those in turn will yield the tertiary constraints, and so on if necessary.

HVF. This means that $\boldsymbol{\omega}|_{\mathscr{C}}$ is degenerate (with the degeneracy directions spanned by the HVFs of the first-class constraints), and hence cannot be a symplectic form. Conversely, $\boldsymbol{\omega}|_{\mathscr{C}}$ will be a symplectic form (on all of $\mathscr{C}$) if and only if all constraints ζ are second-class constraints.

Once again it may appear that we are getting too lost in mathematical abstractions, but this is where we gain a meaningful insight into physics. It happens that in many classical field theories of interest, including GR and EM, all constraints ζ_{j} turn out to be first-class. The degeneracy directions of the presymplectic form $\boldsymbol{\omega}|_{\mathscr{C}}$ in these theories correspond to what at the Lagrangian level are seen as *gauge transformations*: that is, maps of of the fields ψ which do not change the Lagrangian $\mathcal{L}(\psi, \boldsymbol{\nabla}\psi)$. For example, this corresponds to the $\mathbb{U}(1)$ symmetry of EM and the diffeomorphism invariance of GR. We will see explicitly how the latter works when we finally arrive at the canonical formulation of GR in the next section.

Just before doing so, it is salient to address a slightly more technical question, one upon which we will not dwell further in this chapter than the next few paragraphs, but which will reappear in our work on entropy in Chapter 4. Namely, it is the issue of how exactly we do, in fact, recover a symplectic structure for some "effective" subset of the phase space that interests us for the meaningful study of dynamics—which, so far, has intuitively meant the constraint surface $\mathscr{C} \subset \mathscr{P}$. In principle, a symplectic form on $\mathscr{C}$ *could* be obtained if one directly factors out the HVFs of the constraints (which span its kernel) living on $T\mathscr{C}$, by simply identifying all points on the orbits of their flow in $\mathscr{C}$. These are concordantly called *gauge orbits.* Thus, one could work with a factor space $\tilde{\mathscr{P}} \subset \mathscr{C}$ defined simply as the space of gauge orbits in $\mathscr{C}$, and which therefore is, by construction, symplectic (with the factored presymplectic form).

However, depending on the desired aim of implementing the canonical construction of a field theory, taking such an approach can turn out to be problematic. This happens in particular if the theory happens to be diffeomorphism invariant (such as GR), in which case "time evolution" in the sense defined at the beginning of this subsection (of the "instantaneous configuration" in the Hamiltonian theory) can equivalently be regarded as effected by spacetime diffeomorphisms (of the full metric $\boldsymbol{g}$ in the Lagrangian theory). Hence moving to the space of gauge orbits in $\mathscr{C}$ essentially renders the dynamics nonexistent: they become entirely trivial, because they are essentially factored out of $\tilde{\mathscr{P}}$, leaving one with no more sense of "motion through phase space". This issue is developed in clear and lengthy detail in [Schiffrin and Wald 2012].

There exist two possible solutions for ameliorating this difficulty—that is, for obtaining a symplectic structure out of $\boldsymbol{\omega}|_{\mathscr{C}}$ which *does* still preserve a nontrivial notion of "time evolution":

(1) Instead of passing to the space of gauge orbits, one may instead choose a representative of each gauge orbit [Schiffrin and Wald 2012]. The idea is that one can find a surface $\mathscr{S} \subset \mathscr{C}$ such that each gauge orbit in $\mathscr{C}$ intersects $\mathscr{S}$ once and only once. (In fact, sometimes a family of such surfaces that work in localised regions of $\mathscr{C}$ is needed, but we keep our discussion here simplified.) The choice of $\mathscr{S}$ is not unique, and so taking a different surface $\mathscr{S}'$ effectively amounts to a change of description—the freedom of which, in the context of our spacetime splitting, corresponds to "time evolution" (*i.e.* change of representative Cauchy surface in spacetime) on one hand and the associated spatial diffeomorphisms on the other. As we shall discuss further, this is what is effectively encoded in the constraints of GR. It is beyond our scope here to enter further into the concrete technicalities of this procedure; for the interested reader, they are elaborated in [Schiffrin and Wald 2012]. The key point is essentially that the subspace $\mathscr{S}$ of the constraint surface $\mathscr{C}$ resulting from such a construction can be shown to be symplectic,. Therefore, one can work with the symplectic form $\boldsymbol{\omega}|_{\mathscr{S}}$ obtained by pulling back $\boldsymbol{\omega}|_{\mathscr{C}}$ to $\mathscr{S}$.

(2) A specific choice of gauge may be explicitly fixed, such that the combination of the constraints ζ_j coupled with the gauge-fixing conditions becomes second-class. This idea is developed further in Chapter 3 of [Bojowald 2011]. One can by such a procedure obtain a symplectic structure on a subspace of the constraint surface $\mathscr{S} \subset \mathscr{C}$ where the (explicitly chosen) gauge-fixing conditions are satisfied, and where one will thus have a symplectic form $\boldsymbol{\omega}|_{\mathscr{S}}$ (for the fixed gauge).

In other to keep our discussion general, unless otherwise stated, we refer to the symplectic manifold $(\mathscr{S}; \boldsymbol{\omega}|_{\mathscr{S}})$ as the *reduced phase space* irrespective of whether procedure (a) or (b) is used to define it.

2.4. Canonical formulation of general relativity

Now that we have established the procedure for producing a canonical formulation of any field theory given that a Lagrangian formulation exists, let us apply it to GR.

2.4.1. Canonical variables. As the only physical field is the metric $\boldsymbol{g}$, the first immediately suggestible choice for a configuration variable φ is the metric $\boldsymbol{h}$ induced by $\boldsymbol{g}$ on each Cauchy surface Σ, *i.e.* $\boldsymbol{h} = \boldsymbol{g}|_\Sigma$. In index notation, this is given by

$$h_{ab} = g_{ab} + n_a n_b \,. \tag{2.4.1}$$

Now, observe that taking $\boldsymbol{h}$ to be the only gravitational configuration variable would not suffice: there are ten independent field variables in $\boldsymbol{g}$, and $\boldsymbol{h}$ accounts for only six of these! We would not obtain a complete set of equations of motion from this procedure unless all (ten) field variables present in the spacetime metric are mapped to the same number of field variables in φ.

Intuitively, the four degrees of freedom "missing" from $\boldsymbol{h}$ are the "time-time" and "time-space" components of the spacetime metric $\boldsymbol{g}$, which may be regarded as the projections $g_{\boldsymbol{tt}} = g_{ab}t^a t^b$ and $(\boldsymbol{g} \cdot \boldsymbol{t})|_\Sigma = g_{ab}t^a h^{bc}$. There is a one-to-one correspondence between these spacetime metric projections and the choice of the time flow vector field $\boldsymbol{t}$ itself. (In other words, these four degrees of freedom simply encode the freedom in identifying spatial points between Cauchy slices.) However, note that we cannot include $\boldsymbol{t}$ as such in φ to account for these degrees of freedom, as $\boldsymbol{t}$ of course does not live in Σ. In particular, in general it has nonvanishing projections both normally and orthogonally to a Cauchy slice. As the former is a scalar, $\boldsymbol{n} \cdot \boldsymbol{t}$, which may thus be regarded as a function on Σ, and the latter is a vector, $\boldsymbol{t}|_\Sigma = \boldsymbol{t} \cdot \boldsymbol{h}$, in the tangent space of Σ, we may correspondingly take these quantities to account for the full set of spacetime metric degrees of freedom in φ. They are referred to as the *lapse function* and *shift vector* respectively:

$$N = -\,\boldsymbol{n} \cdot \boldsymbol{t} \in \mathscr{F}(\Sigma)\,, \tag{2.4.2}$$

$$\boldsymbol{N} = \boldsymbol{t}|_\Sigma = \boldsymbol{h} \cdot \boldsymbol{t} \in T\Sigma\,. \tag{2.4.3}$$

Note that this implies $\boldsymbol{t} = N\boldsymbol{n} + \boldsymbol{N}$.

Now we have all the pieces for writing down the configuration space of GR:

$$\varphi_{\mathrm{G}} = (N, \boldsymbol{N}, \boldsymbol{h})\,, \tag{2.4.4}$$

where all configuration variables are properly defined on Σ and ultimately encode the full set of (ten) field variables present in the spacetime metric $\boldsymbol{g}$.

The next step is to write the gravitational action (2.2.14) in terms of $\varphi_{\mathrm{G}} = (N, \boldsymbol{N}, \boldsymbol{h})$ and $\dot{\varphi}_{\mathrm{G}} = (\dot{N}, \dot{\boldsymbol{N}}, \dot{\boldsymbol{h}})$ instead of the spacetime metric $\boldsymbol{g}$. The calculation is long but generally straightforward (using all the definitions we have established), and we omit writing it explicitly here. The result can be found, *e.g.*, in Chapter 3 of [Bojowald 2011].

With $\mathcal{L}_{\mathrm{G}}(\varphi_{\mathrm{G}}, \dot{\varphi}_{\mathrm{G}})$ in hand, one can then use it to determine the set of canonical momenta corresponding to the configuration φ_{G}. Let $\pi^{(N)}$, $\pi_a^{(\boldsymbol{N})}$ and $\pi_{(\boldsymbol{h})}^{ab}$ denote, respectively, the canonical momenta of N, N^a and h_{ab}, such that the total set of canonical momenta is $\pi_{\mathrm{G}} = (\pi^{(N)}, \boldsymbol{\pi}^{(\boldsymbol{N})}, \boldsymbol{\pi}_{(\boldsymbol{h})})$. We follow standard convention and henceforth drop the "$(\boldsymbol{h})$" from the canonical momentum of the metric, so we simply write $\pi_{(\boldsymbol{h})}^{ab} = \pi^{ab}$ and

$$\pi_{\mathrm{G}} = (\pi^{(N)}, \boldsymbol{\pi}^{(\boldsymbol{N})}, \boldsymbol{\pi})\,. \tag{2.4.5}$$

Each of these can be computed as the appropriate partials of $\mathcal{L}_{\mathrm{G}}(N, \boldsymbol{N}, \boldsymbol{h}, \dot{N}, \dot{\boldsymbol{N}}, \dot{\boldsymbol{h}})$ using the general canonical momentum definition (2.3.25):

$$\pi^{(N)} = \frac{\partial \mathcal{L}_{\mathrm{G}}}{\partial \dot{N}} = 0\,, \tag{2.4.6}$$

$$\boldsymbol{\pi}^{(\boldsymbol{N})} = \frac{\partial \mathcal{L}_{\mathrm{G}}}{\partial \dot{\boldsymbol{N}}} = 0\,, \tag{2.4.7}$$

$$\boldsymbol{\pi} = \frac{\partial \mathcal{L}_{\mathrm{G}}}{\partial \dot{\boldsymbol{h}}} = \sqrt{h}\left(\boldsymbol{K} - K\boldsymbol{h}\right)\,, \tag{2.4.8}$$

where $K = \mathrm{tr}(\boldsymbol{K})$.

2.4.2. Constraints. Notice that the canonical momenta $\pi^{(N)}$ and $\boldsymbol{\pi}^{(\boldsymbol{N})}$ corresponding respectively to the lapse and the shift both vanish identically. These equations, therefore, identify precisely the degeneracy directions of the map $f : T\mathscr{Q} \to T^*\mathscr{Q}$ (2.3.27) discussed earlier, which in this case maps $(N, \boldsymbol{N}, \boldsymbol{h}, \dot{N}, \dot{\boldsymbol{N}}, \dot{\boldsymbol{h}}) \mapsto (N, \boldsymbol{N}, \boldsymbol{h}, \pi_{(N)}, \boldsymbol{\pi}_{(\boldsymbol{N})}, \boldsymbol{\pi})$. Consequently, $\pi^{(N)} = 0 = \boldsymbol{\pi}^{(\boldsymbol{N})}$ can be taken directly to be the ($\tilde{m} = 4$) primary constraints of GR, $\tilde{\zeta} = \{\frac{1}{\sqrt{h}}\pi^{(N)}, \frac{1}{\sqrt{h}}\boldsymbol{\pi}^{(\boldsymbol{N})}\}$ where we have introduced the (nonvanishing) factors of $\frac{1}{\sqrt{h}}$ for convenience. We write these constraints as

$$\tilde{\zeta}_{\mathrm{G}} = \frac{1}{\sqrt{h}}\pi^{(N)}\,, \tag{2.4.9}$$

$$\tilde{\boldsymbol{\zeta}}_{\mathrm{G}} = \frac{1}{\sqrt{h}}\boldsymbol{\pi}^{(\boldsymbol{N})}\,, \tag{2.4.10}$$

(with the realization that the indices $\tilde{\mathrm{j}}$ on the primary constraints that we were using earlier are in this case spacetime indices).

Physically, this means that the lapse and shift are not dynamical variables. Thus, there exists freedom in choosing them. Equivalently, there is freedom in choosing the time flow vector field in the spacetime foliation, which in turn translates into the freedom of how to identify spatial points at different instants of time. In this way, we can see that these primary constraints (2.4.9)-(2.4.10) are a manifestation of the coordinate freedom of GR.

Next we must ask: are these all the constraints? To investigate, let us compute their time derivatives and equate them to zero on the primary constraint surface (amounting to the consistency conditions (2.3.39)):

$$0 \mathrel{\tilde{=}} \dot{\tilde{\zeta}}_{\mathrm{G}} = \dot{\pi}^{(N)} \mathrel{\mathring{=}} \left\{\dot{\pi}^{(N)}, \tilde{H}_{\mathrm{G}}\right\} \mathrel{\mathring{=}} \left\{\dot{\pi}^{(N)}, H_{\mathrm{G}}\right\} = -\sqrt{h}C\,, \tag{2.4.11}$$

$$0 \mathrel{\tilde{=}} \dot{\tilde{\boldsymbol{\zeta}}}_{\mathrm{G}} = \dot{\boldsymbol{\pi}}^{(\boldsymbol{N})} \mathrel{\mathring{=}} \left\{\dot{\boldsymbol{\pi}}^{(\boldsymbol{N})}, \tilde{H}_{\mathrm{G}}\right\} \mathrel{\mathring{=}} \left\{\dot{\boldsymbol{\pi}}^{(N)}, H_{\mathrm{G}}\right\} = -\sqrt{h}\boldsymbol{C}\,. \tag{2.4.12}$$

The RHS's—whatever they are—have to vanish, and so (up to a factor of $\sqrt{h}$, extracted for convenience) are identified as the secondary constraints, typically denoted *vis-à-vis* our earlier notation as $C = \hat{\zeta}_{\mathrm{G}}$ and $\boldsymbol{C} = \hat{\boldsymbol{\zeta}}_{\mathrm{G}}$ (thus the $\hat{\mathrm{j}}$ indices in $\hat{\zeta}_{\hat{\mathrm{j}}}$ are also spacetime indices here). This is because the equations $\{\dot{\pi}^{(N)}, H_{\mathrm{G}}\} \mathrel{\mathring{=}} 0 \mathrel{\mathring{=}} \{\dot{\boldsymbol{\pi}}^{(N)}, H_{\mathrm{G}}\}$ specify precisely the degeneracy directions of the matrix of Poisson brackets of the primary constraints, $\{\tilde{\zeta}^{\mathrm{G}}_{\mathrm{j}}, \tilde{\zeta}^{\mathrm{G}}_{\mathrm{k}}\}$. Direct computation of the brackets $\{\dot{\pi}^{(N)}, H_{\mathrm{G}}\}$ and $\{\dot{\boldsymbol{\pi}}^{(N)}, H_{\mathrm{G}}\}$ yields the expressions:

$$C = -\mathcal{R} + \frac{1}{h}\left(\boldsymbol{\pi} : \boldsymbol{\pi} - \frac{1}{2}\pi^2\right) , \tag{2.4.13}$$

$$\boldsymbol{C} = \boldsymbol{\mathcal{D}} \cdot \left(\frac{1}{\sqrt{h}}\boldsymbol{\pi}\right) , \tag{2.4.14}$$

where $\mathcal{R}$ is the Ricci scalar of $\boldsymbol{h}$ and $\pi = \mathrm{tr}(\boldsymbol{\pi})$.

Observe that these are precisely the contracted Gauss-Peterson-Mainardi-Codazzi equations, (2.3.7) and (2.3.8) respectively, encountered earlier as inevitable geometrical conditions on the hypersurface embedding! In the present context, they are known as the *Hamiltonian constraint* (2.4.13) and *momentum constraint* (2.4.14), and they complete the set of (eight) constraints of GR. The existence of these secondary constraints is physically related, just as the primary ones although via a slightly different mechanism, to coordinate or "gauge" freedom in GR.

In particular, the Hamiltonian constraint implies a time function redefinition freedom, $t(x) \mapsto t'(x)$ (invariance under a "change of time coordinate"). This essentially means that one can perform the slicing of spacetime into Cauchy surfaces completely as one wishes. The detailed proof of the equivalence between this freedom and the Hamiltonian constraint is a bit subtle and involves the introduction of a few additional technical constructs that we would like to avoid here. Thus we simply omit it, and we refer the interested reader to Chapter 3 of [Bojowald 2011].

The momentum constraint also implies some form of gauge freedom—in this case, spatial diffeomorphism invariance, *i.e.*, the freedom to transform the three-metric $\boldsymbol{h}$ by the action of a diffeomorphism $\phi : \Sigma \to \Sigma$ within the Cauchy surface, *i.e.* $\boldsymbol{h} \mapsto \phi_* \boldsymbol{h}$, without changing the equations of motion. This may be regarded as the freedom to choose a different spatial coordinate system (within Σ). We leave the proof of this claim to the end of this section, where we will be able to show it quite succinctly using the symplectic structure.

We summarize the constraints of GR in Table 1. A visual depiction of the gauge freedoms related to each of the constraints is shown in Fig. 2.6.

We end our discussion on constraints here by remarking on the subtle difference in the meaning of "coordinate freedom" in GR implied by the primary versus the secondary constraints. The secondary constraints—the momentum and Hamiltonian constraints—are respectively equivalent to re-labeling spatial coordinates within a Cauchy surface, and re-labeling the time coordinate (the Cauchy surface foliation within the spacetime). Instead,

	The constraints of GR			
Type	**Primary**		**Secondary**	
Name	*Lapse momentum constraint*	*Shift momentum constraint*	*Hamiltonian constraint*	*Momentum constraint*
Definition	$0 = \pi^{(N)}$	$0 = \boldsymbol{\pi}^{(\boldsymbol{N})}$	$0 = C$	$0 = \boldsymbol{C}$
Eqn.	(2.4.9)	(2.4.10)	(2.4.11)	(2.4.12)
DoF	Time flow vector field invariance		Time function invariance	Diffeomorphism invariance on Σ
Map	$\boldsymbol{t} \mapsto \tilde{\boldsymbol{t}}$		$t(x^a) \mapsto t'(x^a)$	$\boldsymbol{h} \mapsto \phi_* \boldsymbol{h}$

TABLE 1. The constraints of GR. These are classified into primary and secondary constraints, with the name, equation, and DoF (degree of freedom) associated to each as well as the map permitted by the latter.

the primary constraints—the vanishing of the lapse and shift momentum—are about the freedom in how one identifies spatial coordinates on a particular Cauchy surface to coordinates on other embedded Cauchy surfaces in the future (themselves possessing their own spatial diffeomorphism invariance), via the time flow vector field. Nevertheless, these two sorts of freedoms are not completely independent of each other, a fact which is encoded in the consistency requirement $\nabla_{\boldsymbol{t}} t = 1$.

2.4.3. The Hamiltonian and the equations of motion. The computation of the gravitational Hamiltonian $\mathcal{H}_{\mathrm{G}}$, and hence from this, the total gravitational Hamiltonian $\tilde{\mathcal{H}}_{\mathrm{G}} = \mathcal{H}_{\mathrm{G}} - \tilde{\lambda}\tilde{\zeta}_{\mathrm{G}} - \tilde{\boldsymbol{\lambda}} \cdot \tilde{\boldsymbol{\zeta}}_{\mathrm{G}}$ needed to obtain the equations of motion, now follows by directly applying the recipe outlined in the previous section for general field theories. We have laid out all the basic ingredients and from here, conceptually, this is relatively straightforward, however the computation itself turns out to be quite lengthy. Detailed step-by-step presentations can be found, *e.g.*, in Chapter 3 of [Bojowald 2011] or Chapter 4 of [Poisson 2007]. The result is

$$\tilde{H}_{\mathrm{G}}[\varphi_{\mathrm{G}}, \pi_{\mathrm{G}}] = \frac{1}{2\kappa} \int_{\Sigma} \mathbf{e} \sqrt{h} \left[\lambda \cdot \zeta + 2\boldsymbol{\mathcal{D}} \cdot \left(\frac{1}{\sqrt{h}} \boldsymbol{N} \cdot \boldsymbol{\pi} \right) \right] . \tag{2.4.15}$$

Notice that, in the bulk ($\mathrm{int}(\Sigma)$), we only have a linear combination of (all) constraints, $\lambda \cdot \zeta = \lambda^{\mathrm{j}} \zeta_{\mathrm{j}}$. In particular, these are formed by linear contributions from the primary constraints $\tilde{\zeta}_{\mathrm{G}}$ and $\tilde{\boldsymbol{\zeta}}_{\mathrm{G}}$ (added, respectively, with multipliers $\tilde{\lambda}$ and $\tilde{\boldsymbol{\lambda}}$ to obtain $\tilde{\mathcal{H}}_{\mathrm{G}}$ from $\mathcal{H}_{\mathrm{G}}$) and secondary constraints C (2.4.13) and $\boldsymbol{C}$ (2.4.14) (added with multipliers obtained

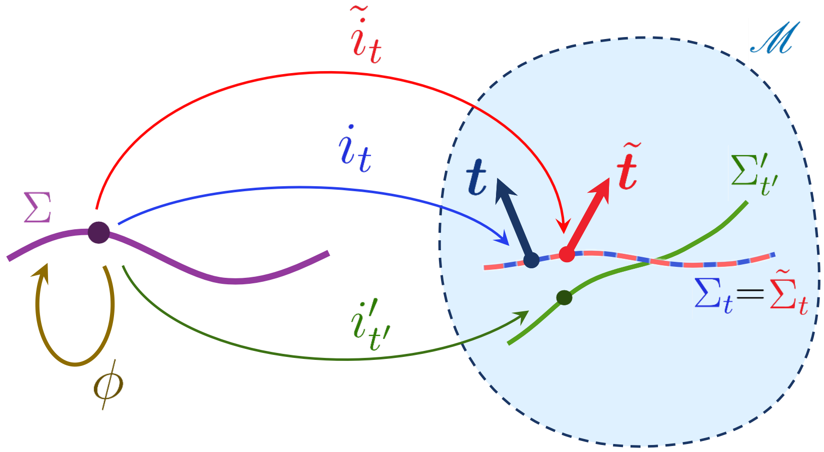

FIGURE 2.6. A visual representation of the "gauge freedoms" of GR. The embedding $i_t : \Sigma \to \mathscr{M}$ is shown in blue, along with the transformations on this embedding permitted by the constraints, shown in different colours. In particular, the primary constraints imply that we can change i_t to a new embedding $\tilde{i}_t$, shown in red, resulting from a change in the time flow vector field $\boldsymbol{t} \mapsto \tilde{\boldsymbol{t}}$ (or equivalently, $(N, \boldsymbol{N}) \mapsto (\tilde{N}, \tilde{\boldsymbol{N}})$). The embedded surface itself does not change, *i.e.* $\tilde{\Sigma}_t = \tilde{i}_t(\Sigma) = i_t(\Sigma) = \Sigma_t$, but the identification of spatial coordinates on sequential Cauchy surfaces in the family of embeddings does. On the other hand, the Hamiltonian constraint implies the freedom to change from i_t to $i'_{t'}$, shown in green, which is a change of foliation, or time function redefinition $t(x^a) \mapsto t'(x^a)$, such that $\Sigma'_{t'} = i'_{t'}(\Sigma)$ does not coincide with $\Sigma_t = i_t(\Sigma)$. Finally, the momentum constraint implies the freedom to map the spatial metric $\boldsymbol{h}$ in Σ by a diffeomorphism ϕ, $\boldsymbol{h} \mapsto \phi_* \boldsymbol{h}$.

from the computation of the Legendre transform):

$$\lambda \cdot \zeta = -\frac{\tilde{\lambda}}{\sqrt{h}} \pi^{(N)} - \frac{\tilde{\boldsymbol{\lambda}}}{\sqrt{h}} \cdot \boldsymbol{\pi}^{(\boldsymbol{N})} + NC - 2\boldsymbol{N} \cdot \boldsymbol{C} \,. \tag{2.4.16}$$

The canonical equations of motion, again following a lengthy but straightforward computation, can now be obtained from what we have established. The result is:

$$\dot{h}_{ab} \mathrel{\mathring{=}} \left\{h_{ab}, \tilde{H}_{\rm G}\right\} \mathrel{\mathring{=}} 2\left[\frac{N}{\sqrt{h}}\left(\pi_{ab} - \frac{1}{2}\pi h_{ab}\right) + D_{(a}N_{b)}\right], \tag{2.4.17}$$

$$\begin{aligned}\dot{\pi}^{ab} \mathrel{\mathring{=}} \left\{\pi^{ab}, \tilde{H}_{\rm G}\right\} \mathrel{\mathring{=}} \frac{N}{\sqrt{h}}\left[-\mathcal{R}^{ab} + \frac{1}{2}h^{ab}\left(\mathcal{R} + \boldsymbol{\pi} : \boldsymbol{\pi} - \frac{1}{2}\pi^2\right) - 2\pi^{ac}\pi_c{}^b + \pi\pi^{ab}\right] \\ + \sqrt{h}\left[D^aD^bN - h^{ab}D^2N + \boldsymbol{D}\cdot\left(\frac{\boldsymbol{N}\pi^{ab}}{\sqrt{h}}\right)\right] - 2\pi^{c(a}D_cN^{b)},\end{aligned} \tag{2.4.18}$$

where $D^2 = \boldsymbol{D}\cdot\boldsymbol{D}$.

2.4.4. Symplectic structure and gauge freedom. Typically what is referred to as *the symplectic form of GR* is the symplectic form on the primary constraint surface $\tilde{\mathscr{C}} = \{(N, \boldsymbol{N}, \boldsymbol{h}, \boldsymbol{\pi})\}$, denoted (with the correspondent slight abuse of notation) by $\boldsymbol{\omega}$. (A symplectic form on the full GR phase space $\mathscr{P} = \left\{\left(N, \boldsymbol{N}, \boldsymbol{h}, \pi^{(N)}, \boldsymbol{\pi}^{(\boldsymbol{N})}, \boldsymbol{\pi}\right)\right\}$ may of course easily be defined by the simple addition of lapse and shift momentum terms, but for analyzing the symplectic structure one can assume these automatically to be zero without loss of generality.). In this case, $\boldsymbol{\omega}$ is also the volume form $\boldsymbol{\Omega} = \boldsymbol{\epsilon}_{\tilde{\mathscr{C}}}$, and is given by

$$\boldsymbol{\omega} = \boldsymbol{\Omega} = \int_\Sigma \mathbf{e}\,\delta\pi^{ab}\wedge\delta h_{ab}\,. \tag{2.4.19}$$

We will see more on this in our work on entropy in Chapter 4.

For now, let us use everything we have established to prove that the momentum constraint $0 = \boldsymbol{C} = \boldsymbol{\mathcal{D}}\cdot(\boldsymbol{\pi}/\sqrt{h})$ (2.4.12) is related to spatial diffeomorphism degrees of freedom in the theory. The idea of this proof appears in Appendix E of [Wald 1984], and we formalize it here by relating it to the symplectic structure (2.4.19).

Let $\mathcal{G}_{\boldsymbol{\xi}} : \tilde{\mathscr{C}} \to \tilde{\mathscr{C}}$ be an infinitesimal spatial gauge transformation determined by any vector field $\boldsymbol{\xi} \in T\Sigma$. In particular, we know that under this map, the metric $\boldsymbol{h}$ must transform as $\boldsymbol{h} \mapsto \boldsymbol{h} + \mathcal{L}_{\boldsymbol{\xi}}\boldsymbol{h}$ where, using the definition of the Lie derivative, $\mathcal{L}_{\boldsymbol{\xi}}h_{ab} = \mathcal{D}_{(a}\xi_{b)}$. Under such a map, the symplectic form transforms as:

$$\begin{aligned}\boldsymbol{\omega} = \int_\Sigma \mathbf{e}\,\delta\pi^{ab}\wedge\delta h_{ab} \mapsto &\int_\Sigma \mathbf{e}\,\delta\pi^{ab}\wedge\delta\left(h_{ab} + \mathcal{D}_{(a}\xi_{b)}\right) \\ = &\int_\Sigma \mathbf{e}\,\delta\pi^{ab}\wedge\delta h_{ab} + \int_\Sigma \mathbf{e}\,\delta\pi^{ab}\wedge\delta\mathcal{D}_{(a}\xi_{b)} \\ = &\,\boldsymbol{\omega} + \int_\Sigma \mathbf{e}\,\delta\pi^{ab}\wedge\mathcal{D}_{(a}\delta\xi_{b)}\,.\end{aligned} \tag{2.4.20}$$

As $\mathcal{G}_{\boldsymbol{\xi}}$ is a gauge transformation, it must preserve the symplectic structure, *i.e.* it must act as the identity on the symplectic form: $(\mathcal{G}_{\boldsymbol{\xi}})^*\boldsymbol{\omega} = \boldsymbol{\omega}$. Thus, $\boldsymbol{\omega} \mapsto \boldsymbol{\omega}$ in (2.4.20) if and only if the integral term in the last equality vanishes. A simple integration by parts with the appropriate handling of the volume form and the assumption of vanishing $\boldsymbol{\xi}$ on $\partial\Sigma$ turns this into the requirement that

$$0 = \int_{\Sigma} \boldsymbol{\epsilon}_{\Sigma} \left[\mathcal{D}_{(a} \left(\frac{1}{\sqrt{h}} \delta\pi^{ab} \right) \right] \wedge \delta\xi_{b)} \, . \tag{2.4.21}$$

Since this is true for any vector field $\boldsymbol{\xi}$, we find that this is equivalent to demanding that, everywhere on Σ,

$$0 = \boldsymbol{\mathcal{D}} \cdot \left(\frac{1}{\sqrt{h}} \boldsymbol{\pi} \right) , \tag{2.4.22}$$

precisely the momentum constraint! This proves that those configurations of $\boldsymbol{h}$ and $\boldsymbol{\pi}$ in the primary constraint surface $\tilde{\mathscr{C}} \subset \mathscr{P}$ which enjoy spatial gauge freedom, *i.e.* which are diffeomorphism invariant, must be those satisfying the momentum constraint.

2.5. Applications

In this section, we elaborate a bit further on and offer some specific examples of applications of the canonical formulation of GR. We structure the discussion into four broad topics: mathematical relativity, numerical relativity, quantum gravity, and gravitational energy-momentum definitions.

2.5.1. Mathematical relativity. From a mathematical point of view, the vacuum Einstein equation $\boldsymbol{G} = 0$ on a spacetime $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ constitutes a system of ten coupled second-order quasi-linear PDEs for the ten components $g_{\alpha\beta}$ of $\boldsymbol{g}$ in some local coordinates $\{x^{\alpha}\}$. Formulated as a canonical problem, the four secondary constraints (the Hamiltonian and momentum constraints) form an elliptic system of PDEs on Σ, which determines the permissible initial data for $\boldsymbol{h}$ and $\boldsymbol{\pi}$. The canonical equations (*i.e.* the time evolution problem) for $\boldsymbol{h}$ and $\boldsymbol{\pi}$, together with a gauge fixing condition (equivalently, a choice of $\boldsymbol{t}$, or N and $\boldsymbol{N}$), then form a hyperbolic system of PDEs.

Many important mathematical issues, of intimate concern also for the physical meaning of GR, can be studied from this perspective. In particular, while we have devoted much discussion to the formulation of the equations of motion themselves, we have said little about the general character of the class of solutions that these equations may admit. We should of course expect this class to be not only wide enough to include configurations of the gravitational field actually observed in nature, but also to avoid "pathological" developments of any given (permissible) initial configurations. Broadly speaking, the aim of *mathematical relativity* is to rigorously address these sorts of problems via mathematical analysis and PDE theory.

A classic problem of in this area, which we outline now briefly, is that of the well-posedness of GR—or generally, that of a given field theory. In any theory, if the time evolution of a (complete) set of canonical fields φ^{A} and their conjugate momenta π_{A} is uniquely determined, *i.e.* if solutions to all canonical equations $\dot{\varphi} = \delta H/\delta\pi$ and $\dot{\pi} = -\delta H/\delta\varphi$ (supplemented with constraints if applicable) exist and are unique, then the theory is said to possess an *initial value formulation.*

In addition to this, theories in physics are usually also expected to satisfy the following two properties:

(1) In a suitable sense, "small" changes in initial data $(\varphi|_{\Sigma_0}, \pi|_{\Sigma_0})$ on some Σ_0 should only produce correspondingly "small" changes in the solution $(\varphi|_{\Sigma_t}, \pi|_{\Sigma_t})$, $t > 0$, over any fixed compact region of $\mathscr{M}$.

(2) Changes in the initial data $(\varphi|_{\Sigma_0}, \pi|_{\Sigma_0})$ in a given subset of Σ_0 should not produce changes in the solution outside the causal future of that subset.

Physically, the first condition is understood to mean that the theory has basic predictive power, since initial conditions can always be measured only to finite accuracy. The second condition is essentially an expression of the relativity principle that information cannot propagate faster than light.

If a theory possesses an initial value formulation satisfying (1) and (2), then it is said that the theory is *well-posed*[6].

Local well-posedness of GR was first proved by Choquet-Bruhat in [Fourès-Bruhat 1952]. The idea of the proof is to work in the *harmonic* (H) gauge, *i.e.* a gauge where locally coordinates $\{x^\alpha_{\mathsf{H}}\}$, called harmonic coordinates, satisfy the harmonic equation $\nabla^2 x^\alpha_{\mathsf{H}} = 0$, where $\nabla^2 = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}$. In such a gauge, as shown, *e.g.*, in Chapter 10 of [Wald 1984], the vacuum Einstein equation is

$$0 = R^{\mathsf{H}}_{\alpha\beta} = -\frac{1}{2} g^{\gamma\delta}_{\mathsf{H}} \partial_\gamma \partial_\delta g^{\mathsf{H}}_{\alpha\beta} + l_{\alpha\beta}(\boldsymbol{g}^{\mathsf{H}}, \boldsymbol{\partial}\boldsymbol{g}^{\mathsf{H}})\,, \tag{2.5.1}$$

with the final term representing all lower (first and zeroth) order terms in the PDE, in this case in all (time and space) coordinates. There is then a theorem [Leray 1952], similar in style to the famous Cauchy-Kovalevskaya theorem for linear hyperbolic PDEs (see, *e.g.*, Chapter 4 of [Evans 1998]) but applicable to a certain pertinent class of quasilinear hyperbolic PDEs, including the harmonic gauge Einstein equation (2.5.1), from which well-posedness of the latter can be shown.

Since the 1950s, mathematical relativity has developed into a field in its own right. See, *e.g.*, the extensive textbook [Choquet-Bruhat 2009]. Aside from issues of well-posedness, which are of current concern to modified theories of gravity (where this issue generally complexifies), another direction of investigation is that of global properties of solutions—*e.g.*, global uniqueness of solutions to the Einstein vacuum equation, first proved in [Choquet-Bruhat and Geroch 1969]. Additionally, there is also the problem of the perturbative stability of known exact solutions to the Einstein equation. For example, the (global

[6] Our definition follows that of [Wald 1984]. For a more technical overview, see [Hilditch 2013].

nonlinear) stability of Minkowski spacetime was famously proved in [Christodoulou and Klainerman 1993], while the stability of the Kerr spacetime remains an open problem. See [Coley 2018, 2017] for recent comprehensive statements of current open problems in this area.

2.5.2. Numerical relativity. The relevance of obtaining numerical solutions of GR hardly requires amplification: from investigating the general behaviour and mathematical character of the Einstein equation and/or (usually quantum/holographic gravity motivated) modifications thereof (see, *e.g.*, the review [Cardoso et al. 2012]), to direct application in gravitational wave astronomy (see, *e.g.*, the review [Duez and Zlochower 2018]); numerical solutions can give clues not only to the validity of mathematical conjectures, but also to the sorts of astrophysical observations that may be achievable and interesting to study. For general recent reviews of the field, see *e.g.* [Lehner and Pretorius 2014; Sarbach and Tiglio 2012].

Canonical methods are generally the most preferred approach for numerically obtaining (especially in the very strong field regime) dynamical solutions of GR. Another formalism commonly used in numerical relativity aside from the canonical decomposition is the *null* (or *characteristic*) *decomposition*, involving the choice of one or two of the coordinates to be null.

Carrying out a numerical evolution in GR in the canonical picture then essentially begins with a specification of initial data $(\Sigma_0, \boldsymbol{h}_0, \boldsymbol{\pi}_0)$ for the dynamical fields, along with a lapse N_0 and shift $\boldsymbol{N}_0$, chosen in such a way that the four secondary constraints $C[\varphi_0^{\mathrm{G}}, \pi_0^{\mathrm{G}}] = 0 = \boldsymbol{C}[\varphi_0^{\mathrm{G}}, \pi_0^{\mathrm{G}}]$, which are four elliptic PDEs on the initial value surface, hold. Then, one evolves this initial data via the twelve hyperbolic canonical equations of motion, which can be shown to guarantee the preservation in time of the secondary constraints. As for the freedom implied by the primary constraints, this must somehow be taken into account in the numerical evolution also (recall that this means the freedom to "re-identify" spatial points via a transformation of the time flow vector field from one Cauchy slice to another). In practice this is typically achieved by some explicit specification either of $(N, \boldsymbol{N})$, or directly of coordinate ("gauge") conditions. In this way, ensuring the satisfaction of the primary constraints for numerical solutions is typically not so problematic.

Instead, what is problematic and what has stifled the progress of numerical relativity for a long time is dealing with the numerical propagation of the secondary constraints. While the PDEs ensure these should remain identically zero, numerically of course one will always have the accumulation of some non-zero error in time (starting from initial data exactly satisfying the secondary constraints). A direct numerical implementation of the standard canonical GR equations in the exact form developed here, for example, leads to exponentially growing modes in this error. (These equations are said to be *weakly* *hyperbolic*.) To overcome this issue, (so-called *strongly hyperbolic*) reformulations of the initial value problem, friendlier to computational stability, are required.

Today, two main such methods are generally in use, which have been found to keep the secondary constraints under adequate numerical control: the so-called *generalized harmonic formulation* and the *BSSN* [Baumgarte and Shapiro 1998; Shibata and Nakamura 1995] (or sometimes *BSSNOK*, including the additional authors of [Nakamura et al. 1987]) *formulation*.

In generalized harmonic formulations, the idea is to work in gauges similar to the harmonic gauge mentioned in the previous subsection, generalizing the harmonic equation on the coordinates to include a source. The first successful simulation of a binary black hole merger was achieved via a harmonic gauge approach in [Pretorius 2005].

The BSSN formulation begins with a conformal rescaling of the spatial metric $\boldsymbol{h}$: in particular, one takes $\tilde{\boldsymbol{h}} = \psi^{-4}\boldsymbol{h}$ to be the dynamical configuration variable, where ψ is a conformal factor chosen such that $\det(\tilde{\boldsymbol{h}}) = 1$. One proceeds from this to define additional phase space variables via similar rescalings, and then to obtain the canonical equations of motion for these variables via the procedures we have outlined in the general canonical (ADM) case. The equations obtained at this point are still generally numerically unstable, but what turns out to solve the issue is a clever addition of the secondary constraints to the dynamical evolution equations. As the former are identically zero analytically, their addition to the latter does not affect the solution in theory. Yet it does seem to affect, in a desirable way, numerical stability in practice. This has been shown "empirically" by its widespread success in strongly dynamical simulations, and rigorously for perturbations of Minkowski space [Alcubierre et al. 2000].

For more on this topic, see the textbooks [Baumgarte and Shapiro 2010; Alcubierre 2012; Shibata 2015].

2.5.3. Quantum gravity. The mathematical formalism of quantum physics has, from its roots up to modern particle theories, largely taken shape in the basic language of canonical formulations. Thus modest approaches towards investigating the question of quantizing gravity have often begun with canonical formulations of GR. (Indeed, recall that the very first canonical formulation of GR was developed for this purpose [Pirani and Schild 1950].)

It turns out that the typical canonical formulation of GR we have developed in this chapter is not itself easily amenable to standard (canonical) quantization procedures. That is, an immediately suggestible idea for quantizing GR would be to turn the set of dynamical classical phase space variables $(\boldsymbol{h}, \boldsymbol{\pi})|_{\mathscr{C}}$—*i.e.*, those in the constraint surface (where all the constraints $\zeta = 0$ have been solved, and which therefore may be regarded as containing all observable information without gauge arbitrariness)—into quantum Hilbert space operators, with their Poisson bracket relations becoming commutators (Dirac brackets, with the appropriate factor of i$\hbar$). This turns out to be very difficult to carry out in practice, as a complete set of observables to characterize the constraint surface $\mathscr{C}$ of GR seems too difficult to construct explicitly. Furthermore, even if this is possible, quantizing only variables in $\mathscr{C}$ by construction leaves out “off-shell” information about the solutions (in $\mathscr{P}\backslash\mathscr{C}$), which may in fact be necessary from a quantum point of view (as, *e.g.*, possible additional contributions in a path integral formulation).

This being the case, one may next contemplate the possibility of working instead with quantization on the full classical phase space $\mathscr{P}$ (or perhaps just the primary constraint surface $\tilde{\mathscr{C}}$). Thus, in addition to the dynamical variables, one also promotes the constraint functions ζ_{j} to Hilbert space operators satisfying the condition that they annihilate any quantum state $|\psi\rangle$ that is a solution of the theory, $\zeta_{\mathrm{j}}|\psi\rangle = 0$. This type of procedure, known as Dirac quantization, also suffers from a number of technical problems (*e.g.*, ambiguities in the choice of the order of the factors in the constraint operator, anomalies etc.).

Much greater progress towards the quantization of GR has instead been made by pursuing *first-order formulations* of the theory, *i.e.* formulations that produce first-order equations of motion directly at the Lagrangian level. While of course the Einstein-Hilbert action $\mathcal{S}_{\mathrm{EH}}[\boldsymbol{g}] = \frac{1}{2\kappa}\int \boldsymbol{e}\,\sqrt{-g}R[\boldsymbol{g}]$ yields as we have seen second-order equations of motion for the field $\boldsymbol{g}$, a simple example of a first-order Lagrangian formulation of GR is the *Palatini action*: $\mathcal{S}_{\mathrm{P}}[\boldsymbol{g},\boldsymbol{\Gamma}] = \frac{1}{2\kappa}\int \boldsymbol{e}\,\sqrt{-g}R[\boldsymbol{\Gamma}]$, where both the metric $\boldsymbol{g}$ and the connection $\boldsymbol{\Gamma}$ are regarded as physical fields and which, by the vanishing of the variation with respect to $\boldsymbol{g}$, yields first-order equations of motion for $\boldsymbol{\Gamma}$ (which turn out to be Christoffel symbols by the vanishing of the variation with respect to $\boldsymbol{\Gamma}$).

First-order formulations of GR that have proved most useful to quantization programs have been in the framework of *tetrads*. These can be very roughly defined as a set of vector fields $\mathfrak{e}_I^a \in T\mathscr{M}$, labeled by an “internal” index $I = 0,1,2,3$, which provides an orthonormal basis of the tangent space at each point, *i.e.* $\{\mathfrak{e}_I^a\}_{I=0}^3$ are such that

$$g_{ab}\mathfrak{e}_I^a\mathfrak{e}_J^b = \eta_{IJ} = \mathrm{diag}(-1,1,1,1) \tag{2.5.2}$$

is the Minkowski metric in the internal coordinates. The dual to a tetrad, to which we associate the index-free notation $\boldsymbol{\mathfrak{e}}^I$,

$$\boldsymbol{\mathfrak{e}}^I = \mathfrak{e}_a^I = \eta^{IJ}\mathfrak{e}_J^b g_{ab}\,, \tag{2.5.3}$$

is a one-form on spacetime called a *co-tetrad*; it may be regarded as encoding the same geometrical/physical information as the metric—and thus, one may consider a second order action $\mathcal{S}[\boldsymbol{\mathfrak{e}}]$ for GR.

Let us see briefly how one can devise a tetrad formulation of GR which is first order. See Chapter 3 of [Wald 1984] and Chapter 6 of [Bojowald 2011] for more details. First, one can show that the exterior derivative of the co-tetrad (2.5.3) takes the form

$$\mathrm{d}\mathfrak{e}_I = \mathfrak{e}^J \wedge \boldsymbol{\omega}_{IJ} \,, \tag{2.5.4}$$

where the $\boldsymbol{\omega}_{IJ}$ are all one-forms (on $\mathscr{M}$) known as *connection one-forms*[7]. Thanks to antisymmetry, all $\boldsymbol{\omega}_{IJ}$ can be determined completely from (2.5.4) in terms of derivatives of the co-tetrads $\mathfrak{e}^I$.

Now consider the spacetime Riemann tensor twice contracted into internal indices, $R_{abIJ} = R_{abcd}\mathfrak{e}^a_I\mathfrak{e}^b_J$. It is possible to show that this is in fact a collection of spacetime two-forms $\boldsymbol{R}_{IJ}$, labelled by two internal indices. These can be expressed as a functions of only the connection one-forms $\boldsymbol{\omega}$:

$$\boldsymbol{R}_{IJ}\left[\boldsymbol{\omega}\right] = \mathrm{d}\boldsymbol{\omega}_{IJ} + \boldsymbol{\omega}_{IK} \wedge \boldsymbol{\omega}^K{}_J \,. \tag{2.5.5}$$

The vacuum Einstein equation can in this case be obtained from the following action, taken to be a functional of the co-tetrads $\mathfrak{e}$ and the connection one-forms $\boldsymbol{\omega}$ (and written with respect to the flat space volume element $\boldsymbol{e}$):

$$\mathcal{S}\left[\mathfrak{e}, \boldsymbol{\omega}\right] = \frac{1}{2\kappa} \int e_{IJKL}\mathfrak{e}^I \wedge \mathfrak{e}^J \wedge \boldsymbol{R}^{KL}\left[\boldsymbol{\omega}\right] \,. \tag{2.5.6}$$

In loop quantum gravity, for example, one carries out the quantization in a canonical formulation derived from an action very similar to (2.5.6) above, called the *Holst action* [S. Holst 1996]. Essentially, it simply adds an additional topological term which does not affect the classical equations of motion, but the inclusion of which turns out to provide crucial space for development of the theory at the quantum level. The coupling constant of this term, generally denoted by γ, is known as the Barbero-Immirzi parameter and, while freely specifiable at the mathematical level, is thought to play the role of a physical constant in loop quantum gravity [Barbero G. 1995; Immirzi 1997].

With the addition of this term, a canonical analysis can be carried out following similar methods as those we have seen in this chapter. The most useful such formulation for loop quantum gravity was developed using *Ashtekar variables*, originally introduced in [Ashtekar 1987]. These are closely related to $(\boldsymbol{\omega}, \mathfrak{e})$, and have proven to be very useful to work with for the purposes of attempting to quantize the theory.

While much progress has been made in the last few decades following these lines, a full theory of quantum gravity remains an open problem in physics today. For more, see the textbook [Rovelli 2007] and the recent review [Ashtekar et al. 2015].

2.5.4. Gravitational energy-momentum. The issue of defining gravitational energy-momentum, and conservation principles in GR more generally, is a notoriously

[7] We follow standard notational convention for the connection one-forms in this subsection; these are of course not to be confused with the symplectic form generally denoted by $\boldsymbol{\omega}$.

subtle one for a multitude of reasons. While this will be treated in far greater detail in Chapter 5, the key point of this problem has a simple physical explanation in the equivalence principle [Misner et al. 1973]: in brief, it is impossible to define a sensible notion of *local* gravitational energy-momentum (in a similar style as one typically does for matter), *i.e.* as a volume density, simply because it is always possible to "transform away" any local gravitational field (at any given spacetime point). Thus a total gravitational energy-momentum as a volume integral of any local density cannot be meaningfully defined in GR.

The solution generally accepted to circumvent this problem is instead to define and work with what are called *quasilocal* definitions of gravitational energy-momentum: namely, *surface densities* (rather than volume densities) which, when integrated over the *boundary* (rather than the interior) of some spatial volume, yield meaningful definitions of the total energy-momentum of that volume.

Today, there exist a number of proposals for energy-momentum formulas in this style, often intended to be valid for arbitrary (closed) spatial regions within a Cauchy surface Σ and in agreement with each other in various limits. See the reviews [Jaramillo and Gourgoulhon 2011; Szabados 2004] for comprehensive summaries. Nevertheless, it is generally expected that such definitions should agree when applied to the entire Cauchy surface Σ, and in particular, that they should recover the definitions motivated by canonical formulations.

Indeed, as we have seen, canonical methods of the sort developed in this chapter treat the entire Cauchy surface as the dynamical "system" of interest, and are therefore restricted to (possibly) providing elucidation on the meaning of energy-momentum only for this entire system, *i.e.* the entire space. For asking questions about "sub-systems" of Σ (*i.e.* finite spatial regions), further geometrical constructions are necessary, and often take the form of worldtube boundary splittings or similar strategies. More on this in Chapter 5.

For now, let us consider the most classic result for an energy definition within canonical GR, the *ADM energy*, applicable to a vacuum spacetime which is *asymptotically flat*. This means that the spacetime is a development of an initial data set $(\Sigma, \boldsymbol{h}, \boldsymbol{\pi})$ such that, outside some compact subset of Σ, there exist coordinates $\{x^i\}$ in which the components of $\boldsymbol{h}$ and $\boldsymbol{\pi}$ satisfy the following "fall-off" conditions in terms of $r = (x^i x_i)^{1/2}$:

$$h_{ij} - \delta_{ij} = \mathcal{O}\left(r^{-1}\right) , \quad \pi_{ij} = \mathcal{O}\left(r^{-2}\right) , \tag{2.5.7}$$

$$\partial^n h_{ij} = \mathcal{O}\left(r^{-(n+1)}\right) , \quad \partial^n \pi_{ij} = \mathcal{O}\left(r^{-(n+2)}\right) , \quad \forall n \geq 1 . \tag{2.5.8}$$

This formalizes the idea that, for "sufficiently large" r, the spacetime is "sufficiently close" to Minkowski. Therefore such spacetimes can be physically interpreted to describe "isolated systems".

Now consider again the gravitational Hamiltonian $H_{\mathrm{G}} = \frac{1}{2\kappa}\int_\Sigma \mathbf{e}\sqrt{h}[NC - 2\boldsymbol{N}\cdot\boldsymbol{C} + 2\boldsymbol{\mathcal{D}}\cdot(\frac{1}{\sqrt{h}}\boldsymbol{N}\cdot\boldsymbol{\pi})]$. On the constraint surface $\mathscr{C}$, only the divergence in the integrand is in general non-zero. Using Stokes' theorem, one may write it as a (closed) boundary integral:

$$H_{\mathrm{G}} \mathrel{\hat{=}} -\frac{1}{\kappa}\oint_{\mathscr{S}} \boldsymbol{\epsilon}_{\mathscr{S}}\left(Nk - \frac{N_a r_b \pi^{ab}}{N\sqrt{\sigma}}\right), \tag{2.5.9}$$

where $\mathscr{S} = \partial\Sigma \simeq \mathbb{S}^2$ is the Cauchy surface boundary (topologically a two-sphere), with unit normal r^a and induced metric $\boldsymbol{\sigma} = \boldsymbol{h}|_{\partial\Sigma}$, and where $k = \mathrm{tr}(\boldsymbol{k})$ is the trace of the extrinsic curvature $\boldsymbol{k}$ of $\mathscr{S}$ in Σ. This $H_{\mathrm{G}}|_{\mathscr{C}}$ (2.5.9) is sometimes called the "solution-valued" Hamiltonian.

For an asymptotically flat $(\Sigma, \boldsymbol{h}, \boldsymbol{\pi})$, *i.e.* satisfying (2.5.7)-(2.5.8), choose a time flow vector field $\boldsymbol{t}$ such that as $r \to \infty$ we have $\boldsymbol{t} \to \boldsymbol{n}$, or equivalently $N \to 1$ and $\boldsymbol{N} \to 0$. This means that asymptotically, spatial points on one time slice of the spacetime are identified directly along the normal with those on a future time slice. Such a $\boldsymbol{t}$ is said to generate an *asymptotic time translation*, and so the evaluation of the gravitational Hamiltonian $H_{\mathrm{G}}|_{\mathscr{C}}$ for this type of $\boldsymbol{t}$ can be interpreted physically as the total gravitational energy of Σ. It is referred to as the ADM energy [Arnowitt et al. 1962], and we see by inspection from Eq. (2.5.9) that it is given by:

$$E_{\mathrm{ADM}} = -\frac{1}{\kappa}\lim_{r\to\infty}\oint_{\mathscr{S}} \boldsymbol{\epsilon}_{\mathscr{S}}\, k\,, \tag{2.5.10}$$

the integral of a surface energy density given by k. This can be shown to recover, for example, the mass parameter in exact black hole spacetimes.

A notion of gravitational momentum can also be defined in this setting. Yet, the asymptotic flatness conditions (2.5.7)-(2.5.8) alone do not suffice, as was first analyzed in detail in [Regge and Teitelboim 1974]. In particular, a momentum definition also requires the *Regge-Teitelboim parity conditions*:

$$\partial^n h_{ij}^{\mathrm{odd}} = \mathcal{O}\left(r^{-(n+2)}\right), \quad \partial^n \pi_{ij}^{\mathrm{even}} = \mathcal{O}\left(r^{-(n+3)}\right), \tag{2.5.11}$$

where $f^{\mathrm{odd}}(x) = f(x) - f(-x)$ is the odd part of a function and $f^{\mathrm{even}}(x) = f(x) + f(-x)$ the even part.

If these parity conditions are satisfied, then it is possible to define an ADM linear momentum, for example, by an application of Noether's theorem to asymptotic space translations. (See Chapter 3 of [Bojowald 2011].) The result is:

$$P^a_{\mathrm{ADM}} = -\frac{1}{\kappa}\lim_{r\to\infty}\oint_{\mathscr{S}} \mathrm{d}^2x\, r_b \pi^{ab}\,. \tag{2.5.12}$$

A similar formula for an ADM angular momentum may be defined from asymptotic rotations[8].

It has been proven that data $(\Sigma, \boldsymbol{h}, \boldsymbol{\pi})$ which satisfy the Regge-Teitleboim parity conditions (2.5.11) are dense among asymptotically flat data (satisfying (2.5.7)-(2.5.8)) in a suitable weighted Sobolev space [Corvino and Schoen 2006].

[8] However, there is arbitrariness in such a formula since it refers, in asymptotic coordinates, to an origin of rotations which may lie outside the asymptotic region [Bojowald 2011].

CHAPTER 3

General Relativistic Perturbation Theory

Chapter summary. This chapter offers a rigorous presentation of perturbation methods in general relativity. Many problems of interest, in gravitational physics generally, often involve phenomena that are "very close", in some suitable sense, to a known exact solution of the theory. This permits the expression of quantities of interest in the form of infinite Taylor series about the known, "background" value, and simplifies the problem to that of computing the terms in these series up to the desired order of accuracy. Such tactics form the basis of computing corrections to the motion of a moving object in general relativity, specifically as caused by self-force effects—a topic that we treat in extensive detail in Chapter 5.

We begin in Section 3.1 with a brief introduction, outlining the basic idea behind the general philosophy of perturbation theory in general relativity. Essentially, the view is that one is trying to solve analytically intractable equations defined on an abstract ("perturbed") spacetime that one cannot construct explicitly, but one that is nonetheless "close enough" to a known exact solution of the theory (the "background"). What must be done, in this case, is to transport these equations to the background manifold under a map—in particular, a diffeomorphism—identifying the different spacetimes, thus turning them into solvable Taylor series on an explicitly known mathematical space. This basic picture, from both a physical and mathematical point of view, lends sensible meaning to the heuristic idea of "adding a perturbation on top of a background".

Section 3.2 is dedicated to formalizing these ideas mathematically. Special attention is paid to the issue of perturbative gauge freedom. In particular, the choice of the map relating the "perturbed" and "background" spacetimes is not unique, and a change to a different map is shown to correspond to a perturbative gauge transformation.

Then in Sections 3.3 and 3.4, we summarize the main ideas and results that have been obtained from the application of perturbation methods to black hole spacetimes. Respectively, these sections consider perturbations to the Schwarzschild-Droste and Kerr spacetimes. The perturbative equations in the former case (the Regge-Wheeler and Zerilli equations) are presented in the context of a canonical analysis, and that in the latter case (the Teukolsky equation) from the point of view of the Newman-Penrose formalism.

Teoria general relativista de les pertorbacions (chapter summary translation in Catalan). Aquest capítol ofereix una presentació rigorosa dels mètodes de pertorbació en la relativitat general. Molts problemes d'interès, en la física gravitacional generalment, solen implicar fenòmens "molt propers", en algun sentit adequat, a una solució exacta coneguda de la teoria. Això permet l'expressió de quantitats d'interès en forma de sèries infinites de Taylor al voltant del valor conegut, de "fons", i simplifica el problema per computar els termes d'aquestes sèries fins a l'ordre de precisió desitjat. Aquestes tàctiques constitueixen la base del càlcul de correccions al moviment d'un objecte en la relativitat general, concretament causada per efectes d'auto-força, un tema que tractem detalladament al capítol 5.

Comencem a la secció 3.1 amb una breu introducció, que descriu la idea bàsica de la filosofia general de la teoria de les pertorbacions en la relativitat general. Essencialment, es tracta de resoldre equacions analíticament intractables definides en un espai abstracte ("pertorbat") que no es pot construir explícitament, però que és "prou proper" a una solució exacta coneguda de la teoria (el "fons"). El que s'ha de fer, en aquest cas, és transportar aquestes equacions al fons sota un mapa - en concret, un difomorfisme - identificant els diferents espais-temps, convertint-les així en sèries de Taylor solucionables en un espai matemàtic explícitament conegut. Aquesta imatge bàsica, tant des del punt de vista físic com matemàtic, dóna un sentit raonable a la idea heurística "d'afegir una pertorbació en un fons".

La secció 3.2 es dedica a formalitzar matemàticament aquestes idees. Es presta una atenció especial al problema de la llibertat de mesura pertorbativa. En particular, l'elecció del mapa relacionant els espais-temps "pertorbats" i de "fons" no és única, i es mostra que un canvi a un mapa diferent correspon a una transformació de mesura pertorbativa.

A continuació, a les Seccions 3.3 i 3.4, resumim les idees i resultats principals que s'han obtingut a partir de l'aplicació de mètodes de pertorbació als espais-temps de forats negres. Respectivament, aquestes seccions consideren pertorbacions als espais-temps de Schwarzschild-Droste i Kerr. Les equacions pertorbatives en el primer cas (les equacions Regge-Wheeler i Zerilli) es presenten en el context d'una anàlisi canònica, i en el segon cas (l'equació de Teukolsky) des del punt de vista del formalisme de Newman-Penrose.

Théorie générale relativiste des perturbations (chapter summary translation in French). Ce chapitre propose une présentation rigoureuse des méthodes de perturbation dans la relativité générale. De nombreux problèmes d'intérêt, dans la physique gravitationnelle en général, impliquent souvent des phénomènes « très proches », dans un sens approprié, d'une solution exacte connue de la théorie. Cela permet d'exprimer des quantités d'intérêt sous la forme d'une série infinie de Taylor autour de la valeur « de fond » connue et simplifie le problème en se limitant au calcul des termes de ces séries jusqu'à l'ordre de précision souhaité. De telles tactiques constituent la base du calcul des corrections apportées au mouvement d'un objet en mouvement dans la relativité générale, en particulier à cause des effets de la force propre - un sujet que nous traitons en détail au chapitre 5.

Nous commençons à la section 3.1 par une brève introduction, décrivant l'idée de base de la philosophie générale de la théorie des perturbations dans la relativité générale. L'essentiel, c'est que l'on essaie de résoudre des équations analytiquement insolubles définies sur un espace-temps abstrait (« perturbé ») qu'on ne peut pas construire explicitement, mais qu'est néanmoins « suffisamment proche » d'une solution exacte connue de la théorie (« le fond »). Ce qui doit être fait, dans ce cas, est de transporter ces équations vers le fond usant une application - en particulier, un difféomorphisme - identifiant les différents espaces-temps, les transformant ainsi en séries de Taylor résolubles sur un espace mathématique explicitement connu. Cette image de base, d'un point de vue physique et mathématique, donne un sens raisonnable à l'idée heuristique « d'ajouter une perturbation au-dessus d'un fond ».

La section 3.2 est consacrée à la formalisation mathématique de ces idées. Une attention particulière est accordée à la question de la liberté de jauge perturbative. En particulier, le choix de l'application reliant les espaces-temps « perturbé » et « de fond » n'est pas unique et une modification apportée à une carte différente correspond à une transformation perturbative de jauge perturbative.

Ensuite, dans les sections 3.3 et 3.4, nous résumons les idées principales et les résultats qui ont été obtenus à partir de l'application de méthodes de perturbation aux espaces-temps de trous noirs. Respectivement, ces sections traitent des perturbations des espaces-temps de Schwarzschild-Droste et de Kerr. Les équations perturbatives dans le premier cas (les équations de Regge-Wheeler et Zerilli) sont présentées dans le contexte d'une analyse canonique et cela dans le second cas (l'équation de Teukolsky) du point de vue du formalisme de Newman-Penrose.

3.1. Introduction

Many problems in GR, when exact or fully numerical solutions cannot be obtained or are impracticable, may be amenable instead to treatment via perturbation theory. That is, one often encounters situations where the desired solution to the Einstein equation, though infeasible to obtain explicitly, is nonetheless "sufficiently close", in some suitable sense, to a known exact solution of the theory. This allows one then to obtain approximate solutions in the form of Taylor series about the known exact solution. Undoubtedly the most famous example of such solutions is that of plane gravitational waves (in the simplest case, on a flat spacetime background).

The heuristic notion that one often starts with in thinking about perturbations is the following. Suppose a background quantity $\mathring{Q}$ (such as the metric), in a background manifold $\mathring{\mathscr{M}}$, is known explicitly as the solution to an equation of interest $E[\mathring{Q}] = 0$ (in $\mathring{\mathscr{M}}$). One then imagines adding a "small" perturbation δQ to this known background quantity, and then assuming that the approximate solution which one seeks is $Q \approx \mathring{Q} + \delta Q$, or more generally $Q = \mathring{Q} + \lambda \delta Q + \mathcal{O}(\lambda^2)$ where λ is a "small" expansion parameter. Then, one inserts this form of Q into the equation of interest $E[Q] = 0$ which is thereby expanded and solved, order by order (up to the desired order), in λ.

This point of view of perturbations is in many cases sufficient for simple calculations, e.g. plane gravitational waves can usefully and simply be thought of as wave-like perturbations "on top of" Minkowski for practical purposes. However, often this perspective is too limiting, and in particular a careful treatment of the self-force problem—the main topic of Chapter 5—requires us to be a bit more precise about exactly what we mean by "a perturbation on top of a background."

Let us begin with a simple question: formally speaking, where (i.e. in what space) does Q live as a mathematical quantity? Clearly it must live on $\mathring{\mathscr{M}}$, as this is the (exact) manifold that we know, and in which we know how to carry out calculations. Nonetheless, Q is the solution to a ("perturbed") equation $E[Q] = 0$ which is, obviously, *not* the exact equation $E[\mathring{Q}] = 0$ for the background solution $\mathring{Q}$ on $\mathring{\mathscr{M}}$. So where does the equation $E[Q] = 0$ really come from?

The answer, of course, is that it is an equation we do not know how to solve exactly, in a manifold which we also do not know exactly (and precisely due to which one designs a perturbation procedure to deal with the problem in the first place). Let $\mathscr{M}_{(\lambda)}$ denote this "true", (analytically) unsolvable manifold. The "true" quantity $Q_{(\lambda)}$ lives here, and satisfies the equation $E_{(\lambda)}[Q_{(\lambda)}] = 0$ on $\mathscr{M}_{(\lambda)}$ exactly. In fact, what we shall ultimately need to work with is a *one-parameter family* of such objects (manifolds and related equations) in λ; we develop this in detail in the next section, but for the moment, to continue setting out the general idea, it is enough to think of $\mathring{\mathscr{M}}$ and $\mathscr{M}_{(\lambda)}$ as just two manifolds.

Now, the "true" manifold $\mathscr{M}_{(\lambda)}$ is assumed to be diffeomorphic to the background $\mathring{\mathscr{M}}$, so that there exists a diffeomorphism $\varphi : \mathring{\mathscr{M}} \to \mathscr{M}_{(\lambda)}$ which identifies spacetime points in the background with points in the "perturbed" spacetime. The "perturbed" equation $E[Q] = 0$ to be solved on $\mathring{\mathscr{M}}$ is then nothing more than the transport (under φ) of the "true" equation $E_{(\lambda)}[Q_{(\lambda)}] = 0$ from $\mathscr{M}_{(\lambda)}$ to $\mathring{\mathscr{M}}$, and so Q is understood as the transport of $Q_{(\lambda)}$ to the background, i.e. $Q = \varphi^* Q_{(\lambda)}$. In other words, what one is really solving is $\varphi^*(E_{(\lambda)}[Q_{(\lambda)}]) = E[Q] = 0$ on $\mathring{\mathscr{M}}$. If λ is "small" in some suitable sense (to be defined more precisely in the next section), then this produces Taylor series on $\mathring{\mathscr{M}}$ in λ, which one then solves order by order.

While this may sound a bit abstract, there is very a sensible physical meaning to this perspective. The "background", strictly speaking, does not exist as an object of study in the "real" world, one "on top of" which one "adds" perturbations. Rather, the background is a mathematical idealization—a crucial one, as it provides the stage upon which we know how to do calculations—which is "close enough" to the "true" world as to permit the representation of quantities of interest in the form of infinite Taylor series thereabout. The latter are just transports to the background of equations that we do not know how to deal with directly in the "real" spacetime, and permit one to arrive at approximations by truncating the Taylor series at the desired order in the perturbation parameter.

This geometrical view of perturbation theory not only renders the technical construction conceptually well-motivated, as we shall see, it thereby also avoids running into any dangerous ambiguities in the interpretation of the perturbation quantities, especially vis-á-vis the delicate issue of perturbative gauge transformations. Indeed, it is worth remembering that in the history of GR much confusion has been created by insufficiently careful treatments of general relativistic perturbations and their related gauge issues, which have often taken a long time to clarify[1].

Given the complexity of the self-force problem, and the fact that gauge issues have proven notoriously difficult therein also, we choose in this chapter to develop perturbation theory from the geometrical perspective just outlined. It will prove indispensable for our work on the self-force in Chapter 5.

3.2. General formulation of perturbation theory

3.2.1. Setup. Our exposition of perturbation theory in this subsection follows closely the treatment of [Bruni et al. 1997]. See also Chapter 7 of [Wald 1984] for a simpler treatment of this topic but following the same philosophy.

Let $\lambda \geq 0$ represent our perturbation parameter. It is a purely formal parameter, in the sense that it should be set equal to unity at the end of any computation and serves only to indicate the order of the perturbation. To formalize the ideas outlined in this chapter introduction, we begin by defining a *one-parameter family* of spacetimes $\{(\mathscr{M}_{(\lambda)}, \boldsymbol{g}_{(\lambda)}, \boldsymbol{\nabla}_{(\lambda)})\}_{\lambda \geq 0}$, where $\boldsymbol{\nabla}_{(\lambda)}$ is the connection compatible with the metric $\boldsymbol{g}_{(\lambda)}$ in $\mathscr{M}_{(\lambda)}, \forall \lambda \geq 0$, such that $(\mathscr{M}_{(0)}, \boldsymbol{g}_{(0)}, \boldsymbol{\nabla}_{(0)}) = (\mathring{\mathscr{M}}, \mathring{\boldsymbol{g}}, \mathring{\boldsymbol{\nabla}})$ is a known, exact spacetime—the "background." See Fig. 3.1 for a visual depiction. For notational convenience, any object with a sub-scripted "(0)" (from a one-parameter perturbative family) is equivalently

[1] This is particularly true in the history of cosmological perturbation theory before the work of [Bardeen 1980], who was the first to formulate it in terms of gauge-invariant quantities. We will not comment more on this particular topic in this thesis; see e.g. [Mukhanov et al. 1992; Brandenberger 2004].

written with an overset "$\circ$" instead. For the GSF problem, $\mathring{\boldsymbol{g}}$ is usually the Schwarzschild-Droste[2] or Kerr metric. Then, one should establish a way of smoothly relating the elements of this one-parameter family (between each other) such that calculations on any $\mathscr{M}_{(\lambda)}$ for $\lambda > 0$—which may be, in principle, intractable analytically—can be mapped to calculations on $\mathring{\mathscr{M}}$ in the form of infinite (Taylor) series in λ—which, provided $\mathring{\mathscr{M}}$ is chosen to be a known, exact spacetime, become tractable, order-by-order, in λ.

Thus, it is convenient to define a (five-dimensional, Lorentzian) product manifold $\mathscr{N} = \mathscr{M}_{(\lambda)} \times \mathbb{R}^{\geq}$, the natural differentiable structure of which is given simply by the direct product of those on $\mathscr{M}_{(\lambda)}$ and the non-negative real numbers (labeling the perturbation parameter), $\mathbb{R}^{\geq} = \{\lambda \in \mathbb{R} | \lambda \geq 0\}$. For any one-parameter family of (k, l)-tensors $\{\boldsymbol{A}_{(\lambda)}\}_{\lambda \geq 0}$ such that $\boldsymbol{A}_{(\lambda)} \in \mathscr{T}^{k}{}_{l}(\mathscr{M}_{(\lambda)})$, $\forall \lambda \geq 0$, we define $\mathsf{A} \in \mathscr{T}^{k}{}_{l}(\mathscr{N})$ by the relation

$$\mathsf{A}^{\alpha_1 \cdots \alpha_k}{}_{\beta_1 \cdots \beta_l}(p, \lambda) = A_{(\lambda)}^{\alpha_1 \cdots \alpha_k}{}_{\beta_1 \cdots \beta_l}(p) \,, \quad \forall p \in \mathscr{M}_{(\lambda)} \text{ and } \forall \lambda \geq 0 \,. \tag{3.2.1}$$

Henceforth any such tensor living on the product manifold will be denoted in serif font—instead of Roman font, which remains reserved for tensors living on $(3 + 1)$-dimensional spacetimes. Furthermore, any spacetime tensor (except for volume forms) or operator written without a sub- or super-scripted (λ) lives on $\mathring{\mathscr{M}}$. Conversely, any tensor (except for volume forms) or operator living on $\mathscr{M}_{(\lambda)}$, $\forall \lambda > 0$, is indicated via a sub- or (equivalently, if notationally more convenient) super-scripted (λ), e.g. $\boldsymbol{A}_{(\lambda)} = \boldsymbol{A}^{(\lambda)} \in \mathscr{T}^{k}{}_{l}(\mathscr{M}_{(\lambda)})$ is always tensor in $\mathscr{M}_{(\lambda)}$. The volume form of any (sub-)manifold $\mathscr{U}$ is always simply denoted by the standard notation $\boldsymbol{\epsilon}_{\mathscr{U}}$ (and is always understood to live on $\mathscr{U}$).

Let $\Phi^{\mathsf{X}}_{(\lambda)} : \mathscr{N} \to \mathscr{N}$ be a one-parameter group of diffeomorphisms generated by a vector field $\mathsf{X} \in T\mathscr{N}$. (That is to say, the integral curves of X define a flow on $\mathscr{N}$ which connects any two leaves of the product manifold.) For notational convenience, we denote its restriction to maps from the background to a particular perturbed spacetime (identified by a particular value of $\lambda > 0$) as

$$\varphi^{\mathsf{X}}_{(\lambda)} = \Phi^{\mathsf{X}}_{(\lambda)}|_{\mathring{\mathscr{M}}} : \mathring{\mathscr{M}} \to \mathscr{M}_{(\lambda)} \tag{3.2.2}$$

$$p \mapsto \varphi^{\mathsf{X}}_{(\lambda)}(p) \,. \tag{3.2.3}$$

The choice of X—equivalently, the choice of $\varphi^{\mathsf{X}}_{(\lambda)}$—is not unique; there exists freedom in choosing it, and for this reason, X—equivalently, $\varphi^{\mathsf{X}}_{(\lambda)}$—is referred to as the perturbative

[2] Commonly, this is referred to simply as the "Schwarzschild metric". Yet, it has long gone unrecognized that Johannes Droste, then a doctoral student of Lorentz, discovered this metric independently and announced it only four months after Schwarzschild [Droste 1916b; Droste 1916a; Schwarzschild 1916; Rothman 2002], so for the sake of historical fairness, throughout this work, we use the nomenclature "Schwarzschild-Droste metric" instead.

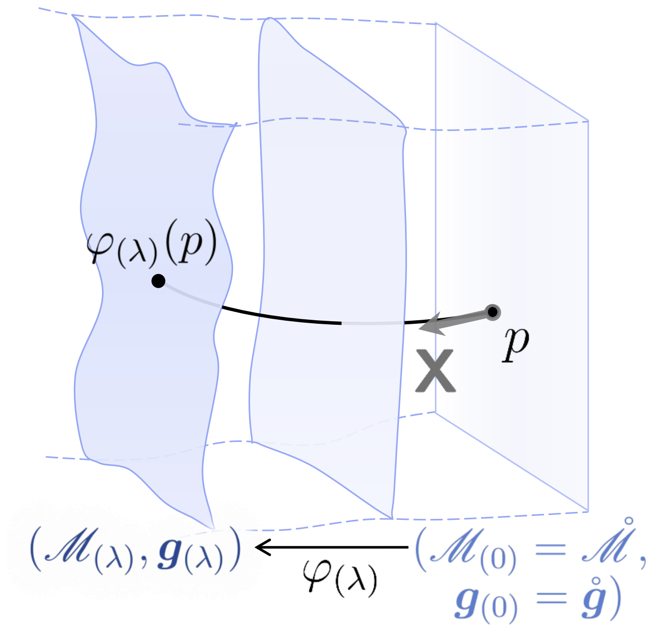

FIGURE 3.1. Representation of a one-parameter family of spacetimes $\{\mathscr{M}_{(\lambda)}\}_{\lambda\geq 0}$ used for perturbation theory. Each of the $\mathscr{M}_{(\lambda)}$ are depicted visually in $(1+1)$ dimensions, as leaves of a (five-dimensional) product manifold $\mathscr{N} = \mathscr{M}_{(\lambda)} \times \mathbb{R}$, with the coordinate $\lambda \geq 0$ representing the perturbative expansion parameter. A choice of a map (or gauge) $\varphi_{(\lambda)} : \mathring{\mathscr{M}} \to \mathscr{M}_{(\lambda)}$, the flow of which is defined by the integral curves of a vector field $\mathbf{X} \in T\mathscr{N}$, gives us a way of identifying any point $p \in \mathring{\mathscr{M}} = \mathscr{M}_{(0)}$ on the "background" to one on some "perturbed" ($\lambda > 0$) spacetime, i.e. $p \mapsto \varphi_{(\lambda)}(p)$.

"gauge". We may work with any different gauge choice $\mathbf{Y}$ generating a different map $\varphi^{\mathbf{Y}}_{(\lambda)} : \mathring{\mathscr{M}} \to \mathscr{M}_{(\lambda)}$. If we do not need to render the issue of gauge specification explicit, we may drop the superscript and, instead of $\varphi^{\mathbf{X}}_{(\lambda)}$, we simply write $\varphi_{(\lambda)}$.

Consider now the transport under $\varphi^{\mathbf{X}}_{(\lambda)}$ of any tensor $\boldsymbol{A}_{(\lambda)} \in \mathscr{T}^k{}_l(\mathscr{M}_{(\lambda)})$ from a perturbed spacetime to the background manifold. We always denote the transport of any such tensor by simply dropping the (λ) sub- or super-script and optionally including a superscript to indicate the gauge—that is, $\forall \boldsymbol{A}_{(\lambda)} \in \mathscr{T}^k{}_l(\mathscr{M}_{(\lambda)})$,

$$(\varphi^{\mathbf{X}}_{(\lambda)})^* \boldsymbol{A}_{(\lambda)} = \boldsymbol{A}^{\mathbf{X}} = \boldsymbol{A} \in \mathscr{T}^k{}_l(\mathring{\mathscr{M}})\,, \tag{3.2.4}$$

and similarly the transport of $\nabla_{(\lambda)}$ to $\mathring{\mathscr{M}}$ is ∇. We know, moreover, that we can express any such $\boldsymbol{A}$ as a Taylor series around its background value, $\boldsymbol{A}_{(0)} = \mathring{\boldsymbol{A}}$ in $\mathring{\mathscr{M}}$. This follows from the Taylor expansion of $\Phi^*_{(\lambda)}\mathbf{A}$ in $\mathscr{N}$ along with the definition of the Lie derivative $\mathcal{L}$ and the group properties of $\Phi_{(\lambda)}$ [Bruni et al. 1997]:

$$\boldsymbol{A} = \mathring{\boldsymbol{A}} + \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} \mathcal{L}^n_{\mathbf{X}} \mathbf{A}|_{\mathring{\mathscr{M}}} \tag{3.2.5}$$

$$= \mathring{\boldsymbol{A}} + \sum_{n=1}^{\infty} \lambda^n \delta^n \boldsymbol{A}\,, \tag{3.2.6}$$

where, in the last equality, we have defined $\delta^n \boldsymbol{A} = (1/n!)(\partial^n_\lambda \boldsymbol{A})|_{\lambda=0}$ and so the (gauge-dependent) first-order perturbation is $\delta^1 \boldsymbol{A} = \delta \boldsymbol{A} = \delta \boldsymbol{A}^{\mathbf{X}}$. Note that the symbol δ^n, $\forall n$, can be thought of as an operator $\delta^n = (1/n!)\partial^n_\lambda|_{\lambda=0}$ that acts upon and extracts the $\mathcal{O}(\lambda^n)$ part of any tensor in $\mathring{\mathscr{M}}$. We refer to

$$\Delta \boldsymbol{A} = \boldsymbol{A} - \mathring{\boldsymbol{A}} = \sum_{n=1}^{\infty} \lambda^n \delta^n \boldsymbol{A} \tag{3.2.7}$$

as the (full) perturbation (in the background) of $\boldsymbol{A}$.

In particular, we have that the background value of the perturbed metric $\boldsymbol{g} = (\varphi^{\mathbf{X}}_{(\lambda)})^* \boldsymbol{g}_{(\lambda)}$ is $\mathring{\boldsymbol{g}}$ and we denote its first-order perturbation for convenience and according to convention as $\boldsymbol{h} = \delta \boldsymbol{g}$. (It is unfortunate that the convention for denoting the spatial three-metric on a Cauchy slice, as in the previous chapter, is usually the same; we henceforth clarify which of these two we are talking about if the context does not make it sufficiently apparent.) Thus we have

$$\boldsymbol{g} = \mathring{\boldsymbol{g}} + \lambda \boldsymbol{h} + \mathcal{O}(\lambda^2)\,, \tag{3.2.8}$$

where we have omitted explicitly specifying the gauge ($\mathbf{X}$) dependence for now.

Let us define one further piece of notation that we shall later need to use: let $\mathring{\boldsymbol{\Gamma}}$ and $\boldsymbol{\Gamma} = (\varphi^{\mathbf{X}}_{(\lambda)})^* \boldsymbol{\Gamma}_{(\lambda)}$ denote the Christoffel symbols (living on $\mathring{\mathscr{M}}$) associated respectively with $\mathring{\boldsymbol{g}}$ and $\boldsymbol{g}$, defined in the usual way (as the connection coefficients between their respective compatible covariant derivatives and the partial derivative). Then their difference,

$$\boldsymbol{C} = (\varphi^{\mathbf{X}}_{(\lambda)})^* \boldsymbol{\Gamma}_{(\lambda)} - \mathring{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma} - \mathring{\boldsymbol{\Gamma}}\,, \tag{3.2.9}$$

is the connection coefficient relating ∇ and $\mathring{\nabla}$ on $\mathring{\mathscr{M}}$, which is in fact a tensor. Note that $\mathring{\boldsymbol{C}} = 0$, i.e. $\boldsymbol{C} = \lambda \delta \boldsymbol{C} + \mathcal{O}(\lambda^2)$. In particular, it is given by

$$C^a{}_{bc} = \frac{\lambda}{2} \mathring{g}^{ad} \left(\mathring{\nabla}_b h_{cd} + \mathring{\nabla}_c h_{bd} - \mathring{\nabla}_d h_{bc} \right) + \mathcal{O}\left(\lambda^2\right)\,. \tag{3.2.10}$$

3.2.2. Perturbed Einstein equations. In this setting, then, what one is often interested in is computing the metric perturbation $\Delta \boldsymbol{g}$ for a known background metric $\mathring{\boldsymbol{g}}$. This means transporting the vacuum Einstein equation $\boldsymbol{R}_{(\lambda)}[\boldsymbol{g}_{(\lambda)}] = 0$ on $\mathscr{M}_{(\lambda)}$ to $\mathscr{M}_{(0)}$. In this way, an approximate solution can be expediently obtained for the metric of the spacetime of interest up to the desired order in λ. In principle, a number of technical subtleties must also be kept in mind whenever a procedure of this sort is implemented [Wald 1984]: *(i)* For an n-th order approximation (in λ), it is in general difficult to estimate the $(n+1)$-th order *error*. Thus, it may be problematic to determine just how "small" λ needs to be in order for this perturbative scheme to be valid to sufficient accuracy. *(ii)* The existence of a one-parameter family $\{\boldsymbol{g}_{(\lambda)}\}$ implies the existence of a solution $\boldsymbol{h}$ to the linearized field equation. However, the converse is not true. Thus, merely solving for $\boldsymbol{h}$ does not guarantee that one will have *linearization stability*, i.e. a corresponding exact solution $\boldsymbol{g}_{(\lambda)}$ in $\mathscr{M}_{(\lambda)}$.

Now, to transform the vacuum Einstein equation $\boldsymbol{R}_{(\lambda)}[\boldsymbol{g}_{(\lambda)}] = 0$ on $\mathscr{M}_{(\lambda)}$ into an equation on $\mathscr{M}_{(0)}$, let us begin by considering the definition of the Riemann tensor: for any $(0,1)$-tensor $\boldsymbol{\omega}_{(\lambda)}$ on $\mathscr{M}_{(\lambda)}$, we have $\nabla^{(\lambda)}_a \nabla^{(\lambda)}_b \omega^{(\lambda)}_c - \nabla^{(\lambda)}_b \nabla^{(\lambda)}_a \omega_c = R^{(\lambda)}_{abc}{}^{d} \omega^{(\lambda)}_d$. The transport of this equation to $\mathscr{M}_{(0)}$, using the fact that the tensor transport commutes with contractions and denoting $\boldsymbol{\omega} = \varphi^*_{(\lambda)} \boldsymbol{\omega}_{(\lambda)}$, is simply:

$$\nabla_a \nabla_b \omega_c - \nabla_b \nabla_a \omega_c = R_{abc}{}^{d} \omega_d, \tag{3.2.11}$$

where $\boldsymbol{\nabla}$ is the transport to to $\mathscr{M}_{(0)}$ of the derivative operator $\boldsymbol{\nabla}_{(\lambda)}$ on $\mathscr{M}_{(\lambda)}$ (compatible with $\boldsymbol{g}_{(\lambda)}$). Inserting $\nabla_a \omega_b = \mathring{\nabla}_a \omega_b - C^c{}_{ab} \omega_c$ where $\boldsymbol{C}$ is the connection coefficient [Eq. (3.2.10)] into the LHS of the transported Riemann formula [Eq. (3.2.11)], a straightforward computation turns this into a relation (in $\mathring{\mathscr{M}}$) between the perturbed and background values of the Riemann tensor:

$$R_{abc}{}^{d} = \mathring{R}_{abc}{}^{d} - 2\mathring{\nabla}_{[a} C^d{}_{b]c} + 2C^e{}_{c[a} C^d{}_{b]e}. \tag{3.2.12}$$

Using the fact that the background metric $\mathring{\boldsymbol{g}}$ satisfies the vacuum Einstein equation, $\mathring{R}_{ac} = 0$, we contract the above to get:

$$R_{ac} = -2\mathring{\nabla}_{[a} C^b{}_{b]c} + 2C^e{}_{c[a} C^b{}_{b]e}. \tag{3.2.13}$$

Inserting the metric expansion [Eq. (3.2.8)] into the definition of $\boldsymbol{C}$ [Eq. (3.2.10)], and then this into Eq. (3.2.13), the perturbed vacuum equation $\boldsymbol{R} = 0$ on $\mathscr{M}_{(0)}$ becomes an infinite set of equations at each order in λ. Carrying out the calculation to second order yields:

$$\mathcal{O}(1): \quad 0 = \mathring{R}_{ac}[\mathring{\boldsymbol{g}}] \tag{3.2.14}$$

$$\mathcal{O}(\lambda): \quad 0 = \delta R_{ac}[\boldsymbol{h}] = -\tfrac{1}{2}\mathring{\nabla}_a \mathring{\nabla}_c h - \tfrac{1}{2}\mathring{\Box} h_{ac} + \nabla^b \nabla_{(c} h_{a)b} \tag{3.2.15}$$

$$\mathcal{O}(\lambda^2): \quad 0 = \delta R_{ac}[\delta^2 \boldsymbol{g}] + \delta^2 R_{ac}[\boldsymbol{h}], \tag{3.2.16}$$

where $\delta^n R_{ac}$ is the n-th order part of the expansion of R_{ac} (in powers of λ), and $\mathring{\Box} = \mathring{\nabla}^b \mathring{\nabla}_b$ is the background wave operator.

3.2.3. Gauge transformations. We now turn to the subtle problem of gauge transformations in perturbation theory. Thus far, we have been working with a one-parameter group of diffeomorphisms $\Phi^{\mathbf{X}}_{(\lambda)} : \mathscr{N} \to \mathscr{N}$ generated by the vector field $\mathbf{X} \in T\mathscr{N}$. What this does, in essence, is to prescribe an identification between points on the different leaves $\mathscr{M}_{(\lambda)}$ of $\mathscr{N}$ along the integral curves of $\mathbf{X}$ (and in particular, between points on the background and any given perturbed spacetime via $\varphi^{\mathbf{X}}_{(\lambda)} = \Phi^{\mathbf{X}}_{(\lambda)}|_{\mathring{\mathscr{M}}}$). However, this identification is not unique; there is freedom in choosing the vector field $\mathbf{X}$ (equivalently, the map $\Phi^{\mathbf{X}}_{(\lambda)}$), referred to as the *gauge choice*, and a change of this vector field (equivalently, the associated map) is called a *gauge transformation*.

To understand the effect of performing a gauge transformation, let $\Phi^{\mathbf{X}}_{(\lambda)} : \mathscr{N} \to \mathscr{N}$ and $\Phi^{\mathbf{Y}}_{(\lambda)} : \mathscr{N} \to \mathscr{N}$ be two different (one-parameter groups of) diffeomorphisms, defined by the integral curves of two different vector fields, $\mathbf{X}$ and $\mathbf{Y}$ respectively, on $\mathscr{N}$ (such that $X^4 = \lambda = Y^4$). See Fig. 3.2.

According to the discussion above, we will obtain two different values of the perturbation in any (k,l)-tensor, $\Delta \boldsymbol{A}^{\mathbf{X}} = \boldsymbol{A}^{\mathbf{X}} - \mathring{\boldsymbol{A}}$ and $\Delta \boldsymbol{A}^{\mathbf{Y}} = \boldsymbol{A}^{\mathbf{Y}} - \mathring{\boldsymbol{A}}$ respectively, depending on which map (or vector field) we use. This is sometimes referred to as "gauge ambiguity" in the calculation of the perturbation, and it is said that $\mathbf{A}$ is (totally) *gauge invariant* if $\Delta \boldsymbol{A}^{\mathbf{X}} = \Delta \boldsymbol{A}^{\mathbf{Y}}$ for any $\mathbf{X} \neq \mathbf{Y}$. The Stewart-Walker lemma [J. M. Stewart and Walker 1974] (see also Chapter 1 of [J. Stewart 1993]) tells us that this happens if and only if $\mathring{\boldsymbol{A}}$ vanishes, is a constant scalar field, or is a linear combination of products of Kronecker deltas with constant coefficients. In general, however, this is not necessarily the case, and so it is important to understand how perturbations change under a gauge transformation.

Let us now define a one-parameter family of diffeomorphisms $\Psi_{(\lambda)} : \mathring{\mathscr{M}} \to \mathring{\mathscr{M}}$ on the background by:

$$\Psi_{(\lambda)} = \varphi^{\mathbf{X}}_{(-\lambda)} \circ \varphi^{\mathbf{Y}}_{(\lambda)} \,. \tag{3.2.17}$$

What this does is to move points in the background along the integral curves of $\mathbf{Y}$ into the perturbed spacetimes, and then along the integral curves of $\mathbf{X}$ "in reverse," back onto the background. (Note that this does *not*, in general, form a group.) Then observe that $\boldsymbol{A}^{\mathbf{X}}$ and $\boldsymbol{A}^{\mathbf{Y}}$ are related by

$$\boldsymbol{A}^{\mathbf{Y}} = [(\Phi^{\mathbf{Y}}_{(\lambda)})^* \mathbf{A}]_{\mathring{\mathscr{M}}} \tag{3.2.18}$$

$$= [(\Phi^{\mathbf{Y}}_{(\lambda)})^* \circ [(\Phi^{\mathbf{X}}_{(\lambda)})^*]^{-1} \circ (\Phi^{\mathbf{X}}_{(\lambda)})^* \mathbf{A}]_{\mathring{\mathscr{M}}} \tag{3.2.19}$$

$$= [(\Phi^{\mathbf{Y}}_{(\lambda)})^* \circ (\Phi^{\mathbf{X}}_{(-\lambda)})^* \circ (\Phi^{\mathbf{X}}_{(\lambda)})^* \mathbf{A}]_{\mathring{\mathscr{M}}} \tag{3.2.20}$$

$$= [\Psi^*_{(\lambda)} \circ (\Phi^{\mathbf{X}}_{(\lambda)})^* \mathbf{A}]_{\mathring{\mathscr{M}}} \tag{3.2.21}$$

$$= \Psi^*_{(\lambda)} \boldsymbol{A}^{\mathbf{X}} \,, \tag{3.2.22}$$

where in the second line we have introduced the identity, and in the following lines we have applied the definitions established so far. Now, according to theorems the proofs of

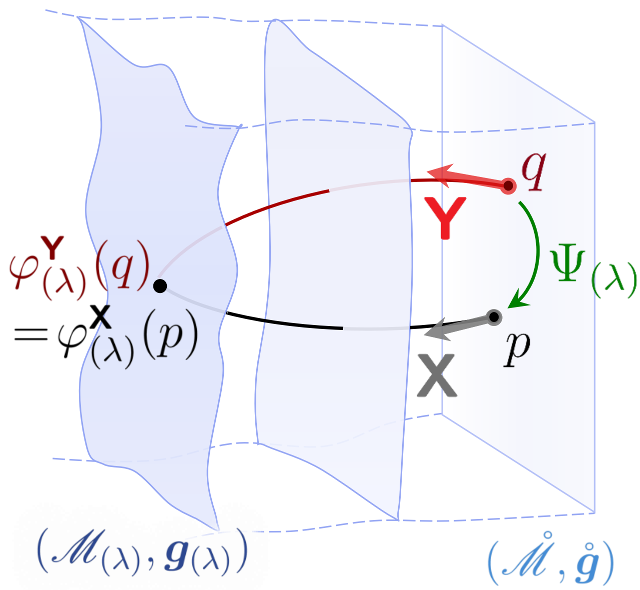

FIGURE 3.2. A gauge transformation consists in choosing a different vector field in $T\mathscr{N}$, or equivalently a different associated diffeomorphism, for identifying points between the background and the perturbed spacetimes. In this illustration, the point $p \in \mathring{\mathscr{M}}$ is mapped under the flow of $\mathbf{X}$ to the same point in $\mathscr{M}_{(\lambda)}$ as is $q \in \mathring{\mathscr{M}}$ under the flow of $\mathbf{Y}$ (for $p \neq q$ and $\mathbf{X} \neq \mathbf{Y}$). One thus has a gauge transformation on the background $q \mapsto \Psi_{(\lambda)}(q) = p$.

which can be found in [Bruni et al. 1997], for any one-parameter family of diffeomorphisms $\Psi_{(\lambda)} : \mathring{\mathscr{M}} \to \mathring{\mathscr{M}}$ [Eq. (3.2.17)], there exists an infinite sequence of one-parameter groups of diffeomorphisms $\{\psi^{(n)}_{(\lambda)} : \mathring{\mathscr{M}} \to \mathring{\mathscr{M}}\}_{n=1}^{\infty}$ such that $\Psi_{(\lambda)} = \cdots \circ \psi^{(n)}_{(\lambda^n/n!)} \circ \cdots \circ \psi^{(2)}_{(\lambda^2/2!)} \circ \psi^{(1)}_{(\lambda)}$. Moreover, the transport under $\Psi_{(\lambda)}$ of any tensor field $\boldsymbol{A}$ on $\mathring{\mathscr{M}}$ has the following series expansion in λ:

$$\Psi^*_{(\lambda)}\boldsymbol{A} = \sum_{l_1,l_2,l_3,\ldots=0}^{\infty} \frac{\lambda^{(\sum_{j=1}^{\infty} j l_j)}}{\prod_{k=1}^{\infty}(k!)^{l_k} l_k!} \mathcal{L}^{l_1}_{\boldsymbol{\xi}_{(1)}} \mathcal{L}^{l_2}_{\boldsymbol{\xi}_{(2)}} \mathcal{L}^{l_3}_{\boldsymbol{\xi}_{(3)}} \cdots \boldsymbol{A}\,, \tag{3.2.23}$$

where $\boldsymbol{\xi}_{(n)} \in T\mathring{\mathscr{M}}$ is the vector field in the background generating the flow of each $\psi^{(n)}_{(\lambda)}$.

Applying the above theorem [Eq. (3.2.23)] to the relation between $\boldsymbol{A}^{\mathsf{X}}$ and $\boldsymbol{A}^{\mathsf{Y}}$ [Eq. (3.2.22)], one obtains:

$$\boldsymbol{A}^{\mathsf{Y}} = \boldsymbol{A}^{\mathsf{X}} + \lambda \mathcal{L}_{\boldsymbol{\xi}_{(1)}} \boldsymbol{A}^{\mathsf{X}} + \frac{\lambda^2}{2}\left(\mathcal{L}^2_{\boldsymbol{\xi}_{(1)}} + \mathcal{L}_{\boldsymbol{\xi}_{(2)}}\right)\boldsymbol{A}^{\mathsf{X}} + \mathcal{O}(\lambda^3)\,. \tag{3.2.24}$$

Substituting series expansions [Eq. (3.2.7)] for $\boldsymbol{A}^{\mathsf{X}}$ and $\boldsymbol{A}^{\mathsf{Y}}$ into the above and demanding that the resulting expression holds order by order yields:

$$\begin{aligned}\Delta \boldsymbol{A}^{\mathsf{Y}} =& \Delta \boldsymbol{A}^{\mathsf{X}} + \lambda\left\{\mathcal{L}_{\boldsymbol{\xi}_{(1)}} \mathring{\boldsymbol{A}}\right\} \\ &+ \lambda^2\left\{\frac{1}{2}\left(\mathcal{L}^2_{\boldsymbol{\xi}_{(1)}} + \mathcal{L}_{\boldsymbol{\xi}_{(2)}}\right)\mathring{\boldsymbol{A}} + \mathcal{L}_{\boldsymbol{\xi}_{(1)}}\left[(\mathcal{L}_{\mathsf{X}}\mathsf{A})_{\mathring{\mathscr{M}}}\right]\right\} + \mathcal{O}(\lambda^3),\end{aligned} \tag{3.2.25}$$

where $\boldsymbol{\xi}_{(1)} = \boldsymbol{Y} - \boldsymbol{X}$, and $\boldsymbol{\xi}_{(2)} = [\boldsymbol{X}, \boldsymbol{Y}]$.

To see the relation between a perturbative gauge transformation and a "change of coordinates", let us apply the above to the situation where A is simply a coordinate function x^α. Then, one can easily check that

$$\Psi^*_{(\lambda)} x^\alpha = x^\alpha + \lambda \xi^\alpha_{(1)} + \frac{\lambda^2}{2}\left(\xi^\beta_{(1)}\partial_\beta \xi^\alpha_{(1)} + \xi^\alpha_{(2)}\right) + \mathcal{O}(\lambda^3)\,. \tag{3.2.26}$$

3.2.4. The Lorenz and transverse-traceless gauges. It is convenient to introduce the *trace-reversed* metric perturbation,

$$\tilde{\boldsymbol{h}} = \boldsymbol{h} - \frac{1}{2} h \mathring{\boldsymbol{g}}\,. \tag{3.2.27}$$

Then, under a gauge transformation generated by some vector field $\boldsymbol{\xi}_{(1)} = \boldsymbol{\xi}$ (to first order), we have $\boldsymbol{h}^{\mathsf{X}} \mapsto \boldsymbol{h}^{\mathsf{Y}} = \boldsymbol{h}^{\mathsf{X}} + \mathcal{L}_{\boldsymbol{\xi}} \mathring{\boldsymbol{g}}$ and so,

$$\tilde{h}^{\mathsf{Y}}_{ab} = \tilde{h}^{\mathsf{X}}_{ab} + \nabla_a \xi_b + \nabla_b \xi_a - (\mathring{\boldsymbol{\nabla}} \cdot \boldsymbol{\xi})\mathring{g}_{ab}\,, \tag{3.2.28}$$

$$\Rightarrow \mathring{\nabla}^b \tilde{h}^{\mathsf{Y}}_{ab} = \mathring{\nabla}^b \tilde{h}^{\mathsf{X}}_{ab} + \mathring{\Box}\xi_a + \mathring{R}_{ab}\xi^b = \mathring{\nabla}^b \tilde{h}^{\mathsf{X}}_{ab} + \mathring{\Box}\xi_a\,, \tag{3.2.29}$$

where in the last line we have used the contraction of the Riemann formula on $\mathring{\mathscr{M}}$ and finally the fact that $\mathring{\boldsymbol{R}} = 0$. If we choose $\boldsymbol{\xi}$ to be a solution of the equation $\mathring{\Box}\xi_a = -\mathring{\nabla}^b \tilde{h}^{\mathsf{X}}_{ab}$, then the gauge defined by Y is known as the *Lorenz gauge*. In this case we denote $\mathsf{Y} = \mathsf{L}$, whereupon we have

$$\mathring{\nabla}^b \tilde{h}^{\mathsf{L}}_{ab} = 0\,. \tag{3.2.30}$$

This does not, however, completely fix all of the available gauge freedom, for we could still add to $\boldsymbol{\xi}$ any other vector $\boldsymbol{\zeta}$ satisfying $\mathring{\Box}\boldsymbol{\zeta} = 0$, and the Lorenz gauge condition [Eq. (3.2.30)] would still hold. Performing a further gauge transformation generated by such a $\boldsymbol{\zeta}$, the trace of $\boldsymbol{h}^{\mathsf{L}}$ transforms as $h^{\mathsf{L}} \mapsto h^{\mathsf{L}} + \mathring{\boldsymbol{\nabla}} \cdot \boldsymbol{\zeta}$. We can now choose $\boldsymbol{\zeta}$ to be a solution of $\boldsymbol{\nabla} \cdot \boldsymbol{\zeta} = -h^{\mathsf{L}}$. (Note that this is consistent, since the trace of the linearized Einstein equation [Eq. (3.2.15)] assuming the Lorenz gauge condition [Eq. (3.2.30)] is $\mathring{\Box} h^{\mathsf{L}} = 0$, and so taking the Laplacian of the equation $\mathring{\boldsymbol{\nabla}} \cdot \boldsymbol{\zeta} = -h^{\mathsf{L}}$ leads to both sides vanishing if $\mathring{\Box}\zeta = 0$.) In other words, the first-order metric perturbation can also be made traceless; combined with the Lorenz gauge condition [Eq. (3.2.30)] (which now is satisfied also by the metric perturbation itself, since it equals its trace-reversed part), this leads to the *transverse-traceless* gauge, defined by the gauge vector T,

$$\mathring{\nabla}^b h^{\mathsf{T}}_{ab} = 0, \quad h^{\mathsf{T}} = 0\,. \tag{3.2.31}$$

In this gauge, the Einstein equation [Eq. (3.2.15)] reduces to a very simple wave-type equation:

$$\mathring{\Box} h^{\mathsf{T}}_{ac} + 2\mathring{R}_a{}^b{}_c{}^d h^{\mathsf{T}}_{bd} = 0\,. \tag{3.2.32}$$

If the background is Minkowski space, the above simplifies to the elementary wave equation $\mathring{\Box}\boldsymbol{h}^{\mathsf{T}} = 0$. This readily admits plane wave solutions; for example, choosing a coordinate system such that the direction of propagation of the plane wave is aligned with the Cartesian z-direction, one finds the solution:

$$h^{\mathsf{T}}_{\alpha\beta} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \exp\left(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}\right)\,. \tag{3.2.33}$$

See Chapter 7 of [Carroll 2003] for more discussion on this solution and related elementary aspects of gravitational waves. See also the textbook [Maggiore 2007] for a more involved development of this topic.

3.3. Perturbations of the Schwarzschild-Droste spacetime

3.3.1. The Schwarzschild-Droste spacetime. In this section we will consider the problem of perturbations to the Schwarzschild-Droste spacetime $(\mathring{\mathscr{M}}, \mathring{\boldsymbol{g}}, \mathring{\boldsymbol{\nabla}})$ the metric of which is given, in Schwarzschild coordinatres, by:

$$\mathring{\boldsymbol{g}} = -f\left(r\right)\mathrm{d}t^2 + \frac{1}{f\left(r\right)}\mathrm{d}r^2 + r^2\boldsymbol{g}_{\mathbb{S}^2}\,, \tag{3.3.1}$$

where $f(r) = 1 - 2M/r$ is the Schwarzschild function and $\boldsymbol{g}_{\mathbb{S}^2}$ is the metric on the unit two-sphere. We furthermore denote the derivative operator compatible with $\boldsymbol{g}_{\mathbb{S}^2}$, for ease of readability, simply as $\boldsymbol{\nabla}_{\mathbb{S}^2}$ for the remainder of this section.

It will be instructive to cast this into canonical language, using the notation and tools established in the previous chapter. Thus, we reserve here the notation $\boldsymbol{h}$ to refer to the spatial three-metric, and all metric perturbations will be denoted strictly using the δ notation in order to avoid confusion.

Cauchy surfaces $\mathring{\Sigma}$ for this spacetime can most easily be defined by the constancy of the Schwarzschild time coordinate, i.e. $t = \mathrm{const.}$ in (3.3.1). In this case the (background) canonical variables can be read off the metric directly by inspection:

$$\mathring{N} = \sqrt{f(r)}\,, \tag{3.3.2}$$

$$\mathring{\boldsymbol{N}} = 0\,, \tag{3.3.3}$$

$$\mathring{\boldsymbol{h}} = \frac{1}{f(r)}\mathrm{d}r^2 + r^2 \boldsymbol{g}_{\mathbb{S}^2}\,, \tag{3.3.4}$$

$$\mathring{\boldsymbol{\pi}} = 0\,. \tag{3.3.5}$$

Of course, the (background) canonical equations of motion are in this case trivial.

Now, observe that it is possible to perform a further foliation of the Schwarzschild Cauchy surfaces $\mathring{\Sigma}$ themselves each into a set of two-spheres, $\mathring{\Sigma} \simeq \mathbb{R}\times\mathbb{S}^2$. This is in fact a general consequence of the spherical symmetry of the Schwarzschild-Droste spacetime, and is true irrespective of the (background) coordinate choice. In particular, let us define

$$\mathscr{S}_r = \left\{ p \in \mathring{\Sigma} :\ \left(x^i x_i\right)^{1/2}\Big|_p = r \right\} \simeq \mathbb{S}^2_r \tag{3.3.6}$$

to be r-radius two-spheres embedded in $\mathring{\Sigma}$, topologically equivalent to r-radius two spheres $\mathbb{S}^2_r$ embedded in $\mathbb{R}^3$. Then we may write

$$\mathring{\Sigma} = \bigcup_{r>2M} \mathscr{S}_r \simeq \bigcup_{r>2M} \mathbb{S}^2_r\,. \tag{3.3.7}$$

Furthermore let r^a denote the outward-pointing unit vector normal to $\mathscr{S}_r$.

3.3.2. Historical development of perturbation approaches. In principle, the perturbation problem for $\delta\boldsymbol{g}$ in a given background spacetime involves solving for ten degrees of freedom (determined by the ten linearized Einstein equations [Eq. (3.2.15)]). Yet, just as the symmetries of the exact Schwarzschild-Droste spacetime—i.e., the fact that it is static and spherically-symmetric—tremendously reduce the number of degrees of freedom and simplify the obtention of the explicit background solution, they do so too for its perturbations.

Historically, the starting point for this problem has been to profit from these symmetries from the outset by expanding all expressions of interest in spherical harmonics, thus reducing the problem to studying the evolution of spherical harmonic modes. We proceed to define these following [Martel and Poisson 2005].

Concretely, any function on $\mathring{\mathscr{M}}$ can be expanded in the form of a series in the standard scalar spherical harmonic functions $Y^{lm} : \mathbb{S}^2 \to \mathbb{R}$, which are defined by the eigenvalue PDE

$$0 = \left[\triangle_{\mathbb{S}^2} + l\,(l+1)\right] Y^{lm}\,. \tag{3.3.8}$$

This can be used to account for all coordinate components of the metric perturbation $\delta\boldsymbol{g}$ in the time-time, radial-radial, and time-radial directions, which thanks to our $(2+1+1)$ splitting may be viewed as well-defined functions on $\mathring{\mathscr{M}}$. Hence the set of functions $\{Y^{lm}\}$

forms a complete basis for these. However, we will also have vector and tensor degrees of freedom induced by $\delta\boldsymbol{g}$ on each $\mathscr{S}_r$, for which tensorial generalizations of the scalar spherical harmonics Y^{lm} are needed.

One defines *even-parity* vector harmonics as the derivatives of the scalars, $\boldsymbol{\nabla}_{\mathbb{S}^2} Y^{lm}$, and *odd-parity* vector harmonics by contracting the Levi-Civita symbol $\epsilon_{\mathfrak{i}\mathfrak{j}}$ with the derivatives, $\epsilon_{\mathfrak{j}}{}^{\mathfrak{i}}\nabla^{\mathfrak{j}}_{\mathbb{S}^2} Y^{lm}$. (We remind the reader that Fraktur indices $\mathfrak{i}$, $\mathfrak{j}$, $\mathfrak{k}$,... are used throughout for indicating components of tensors living on topological two-spheres.) The set

$$\mathscr{B}_{\mathrm{VH}} = \left\{ \nabla^{\mathfrak{i}}_{\mathbb{S}^2} Y^{lm}, \epsilon_{\mathfrak{j}}{}^{\mathfrak{i}} \nabla^{\mathfrak{j}}_{\mathbb{S}^2} Y^{lm} \right\} \tag{3.3.9}$$

then forms a complete basis for expanding vectors on two-spheres. Similarly, one defines *even-parity* ($(0, 2)$-) tensor harmonics as the quantities $g^{\mathbb{S}^2}_{\mathfrak{i}\mathfrak{j}} Y^{lm}$ and $\nabla^{\mathbb{S}^2}_{\mathfrak{i}} \nabla^{\mathbb{S}^2}_{\mathfrak{j}} Y^{lm}$ (or any two independent linear combinations of these), and *odd-parity* ($(0, 2)$-) tensor harmonics as $\epsilon^{\mathfrak{k}}{}_{(\mathfrak{i}} \nabla^{\mathbb{S}^2}_{\mathfrak{j})} \nabla^{\mathbb{S}^2}_{\mathfrak{k}} Y^{lm}$. The set

$$\mathscr{B}_{\mathrm{TH}} = \left\{ g^{\mathbb{S}^2}_{\mathfrak{i}\mathfrak{j}} Y^{lm}, \nabla^{\mathbb{S}^2}_{\mathfrak{i}} \nabla^{\mathbb{S}^2}_{\mathfrak{j}} Y^{lm}, \epsilon^{\mathfrak{k}}{}_{(\mathfrak{i}} \nabla^{\mathbb{S}^2}_{\mathfrak{j})} \nabla^{\mathbb{S}^2}_{\mathfrak{k}} Y^{lm} \right\} \tag{3.3.10}$$

then forms a complete basis for expanding ($(0, 2)$-) tensors on two-spheres. The basis $\{Y^{lm}\}$ of scalar harmonics is itself also referred to as *even-parity*. The procedure for generating higher tensorial harmonics can be continued onward in this fashion, and orthogonality properties generalize from the classic ones known for the scalar harmonics. See [Martel and Poisson 2005; Nagar and Rezzolla 2005; Price 2007] for more technical details on this.

Then, the basic idea of the spherical harmonic approach to Schwarzschild-Droste perturbations is to expand the components of the time-radial sector of $\delta\boldsymbol{g}$ in $\{Y^{lm}\}$ (as we have argued that we can), the two-sphere projection of $\delta\boldsymbol{g}$—in this case, as a full $(0, 2)$-tensor—on the two-spheres in the tensor harmonic basis $\mathscr{B}_{\mathrm{TH}}$ (3.3.10), and finally the time-angular and radial-angular components of $\delta\boldsymbol{g}$ ("lost" in the full two-sphere projection, and reminiscent of the shift vector in the canonical spacetime splitting), as vectors on the two spheres in the vector harmonic basis $\mathscr{B}_{\mathrm{VH}}$ (3.3.9).

Initial investigations in this direction were pioneered by [Regge and Wheeler 1957], who began by considering only the odd-parity modes. Supposing we write $\delta\boldsymbol{g} = \delta\boldsymbol{g}_{\mathrm{even}} + \delta\boldsymbol{g}_{\mathrm{odd}}$ to indicate the total splitting of the metric perturbation into even and odd parity spherical harmonic series expansions, the idea was thus to consider first $\delta\boldsymbol{g}_{\mathrm{odd}}$ since clearly its general form is simpler (completely lacking any contributions from the time-radial sector) compared with that of $\delta\boldsymbol{g}_{\mathrm{even}}$.

Regge and Wheeler actually found a perturbative gauge transformation—now eponymously named—that reduces the number of odd-parity degrees of freedom even further, showing that all of the information about $\delta g^{\mathrm{odd}}_{ab}$ can be reconstructed from a single, gauge-invariant *master function*, also known eponymously as the *Regge-Wheeler function*, which we denote as $\Phi_{(-)}(t,r,\theta,\phi)$. To state the exact relationship in this language requires a few further definitions which we prefer to omit here. Nevertheless, we shall see explicitly in the following subsection an approach to this problem from a canonical point of view, where the definition of this master function is possible to state quite simply in canonical language.

In fact, the exact choice of definition of this master function is not completely unique. There is freedom in its normalization, and often it is more convenient to work with what effectively amounts to the time integral of the Regge-Wheeler function $\Phi_{(-)}$, called the Cunningham-Price-Moncrief function [Cunningham et al. 1978]. The variety of master functions and normalization conventions which are today often worked with is detailed, e.g., in [Nagar and Rezzolla 2005].

In any case, the harmonic modes $\Psi^{lm}_{(-)}(t,r)$ of the Regge-Wheeler master function $\Phi_{(-)}(t,r,\theta,\phi)$ (or of the Cunningham-Price-Moncrief function) completely decouple and satisfy a $(1+1)$-dimensional wave equation with a radially-dependent potential. It has since been known as the *Regge-Wheeler equation*. (We shall see it explicitly in the next subsection.) Remarkably, some years later when the analysis of the even-parity part of the metric perturbation $\delta \boldsymbol{g}_{\text{even}}$ was also fully carried out in this style by [Vishveshwara 1970] and [Zerilli 1970], it was similarly found that all the components thereof can also be reconstructed from a gauge-invariant master function, in this case called the *Zerilli-Moncrief function*, and denoted here as $\Phi_{(+)}(t,r,\theta,\phi)$. Moreover, the equation satisfied by the modes $\Psi^{lm}_{(+)}(t,r)$ of this function, known as the *Zerilli equation*, is the same as the one for the odd-parity modes $\Psi^{lm}_{(-)}(t,r)$, just with a different (slightly more complicated) radial potential. We now develop this in detail, following a canonical approach, in the next subsection.

3.3.3. The Regge-Wheeler and Zerilli equations via canonical methods. Here we summarize a derivation of the Schwarzschild-Droste perturbation equations—the Regge-Wheeler and Zerilli equations—put forward by [Jezierski 1999]. In particular, this approach is based on canonical methods, and will thus reveal a useful synthesis of many of the main ideas we have developed in this thesis so far. Furthermore, it will also prove very helpful in our work on entropy in Chapter 4.

First, in order to avoid notational confusion, let $\delta_{\mathscr{P}}$ denote for this subsection the functional exterior derivative on the phase space $\mathscr{P}$ (so as not to confuse it with the use of "δ" in our perturbative notation). In [Jezierski 1999], the reduced symplectic form of GR for the perturbed Schwarzschild-Droste spacetime is computed—that is, the pullback of the symplectic form $\boldsymbol{\omega} = \int_{\mathring{\Sigma}} \mathbf{e}\, \delta_{\mathscr{P}} \pi^{ab} \wedge \delta_{\mathscr{P}} h_{ab}$ to the reduced phase space $\mathscr{S}$. (See Chapter 2.)

In this case, both the three-metric and its canonical momentum are perturbation series in λ (themselves transports onto $\mathring{\Sigma}$ in $\mathring{\mathscr{M}}$ from a perturbed Cauchy surface of $\mathscr{M}_{(\lambda)}$), i.e.

$\boldsymbol{h} = \mathring{\boldsymbol{h}} + \lambda\delta\boldsymbol{h} + \mathcal{O}(\lambda^2)$ (with $\mathring{\boldsymbol{h}}$ given by (3.3.4)) and $\boldsymbol{\pi} = \lambda\delta\boldsymbol{\pi} + \mathcal{O}(\lambda^2)$ (since here $\mathring{\boldsymbol{\pi}} = 0$). Then, one can decompose all of these quantities into "radial" and "angular" parts according to the $(2+1)$ spatial decomposition. Doing this turns out to naturally isolate the components which are pure gauge degrees of freedom. One then factors these out, and the final result [Jezierski 1999] can be expressed solely in terms of two canonical pairs $(\Phi_{(\pm)}(t,r,\theta,\pi), \Pi_{(\pm)}(t,r,\theta,\pi))$, all simple functions on $\mathring{\mathscr{M}}$, as:

$$\boldsymbol{\omega}|_{\mathscr{S}} = \sum_{\varsigma=\pm} \int_{\mathring{\Sigma}} \mathbf{e}\, \delta_{\mathscr{P}}\Pi_{(\pm)} \wedge \mathbb{D}\delta_{\mathscr{P}}\Phi_{(\pm)}\,, \tag{3.3.11}$$

where the operator $\mathbb{D} = \triangle_{\mathbb{S}^2}^{-1}(\triangle_{\mathbb{S}^2}+2)^{-1}$ is formed from the unit two-sphere Laplacian $\triangle_{\mathbb{S}^2} = \boldsymbol{\nabla}_{\mathbb{S}^2}\cdot\boldsymbol{\nabla}_{\mathbb{S}^2}$.

The two configuration variables $\Phi_{(\pm)}$ are in fact, as notationally anticipated, precisely the Zerilli and Regge-Wheeler master functions respectively, discussed in the previous subsection. In terms of the (dynamical) perturbations of the canonical variables $(\delta\boldsymbol{h}, \delta\boldsymbol{\pi})$, these are given by:

$$\Phi_{(-)} = \frac{2r^2}{\mathring{N}\sqrt{\mathring{h}}}\boldsymbol{\epsilon} : \left(\boldsymbol{\nabla}_{\mathbb{S}^2}\left(\boldsymbol{r}\cdot\delta\boldsymbol{\pi}\right)\right)\,, \tag{3.3.12}$$

$$\begin{aligned}\Phi_{(+)} = {} & r^2\boldsymbol{\nabla}_{\mathbb{S}^2}\cdot\left(\boldsymbol{\nabla}_{\mathbb{S}^2}\cdot\delta\boldsymbol{h}\right) - \left(\triangle_{\mathbb{S}^2}+1\right)\mathrm{tr}_{\mathbb{S}^2_r}\left(\delta\boldsymbol{h}\right) \\ & + \triangle_{(+)}\left(2\left(\boldsymbol{r} + r\boldsymbol{\nabla}_{\mathbb{S}^2}\right)\cdot\left(\boldsymbol{r}\cdot\delta\boldsymbol{h}\right) - rf\left(r\right)\partial_r \mathrm{tr}_{\mathbb{S}^2_r}\left(\delta\boldsymbol{h}\right)\right)\,,\end{aligned} \tag{3.3.13}$$

where ϵ_{ij} is the Levi-Civita symbol, and the last line is written using the operator $\triangle_{(+)} = (\triangle_{\mathbb{S}^2}+2)(\triangle_{\mathbb{S}^2}+2-\frac{6M}{r})^{-1}$. Their conjugate momenta are given by

$$\Pi_{(\pm)} = \frac{\mathring{N}}{\sqrt{\mathring{h}}}\dot{\Phi}_{(\pm)}\,, \tag{3.3.14}$$

with the factor $\mathring{N}/\sqrt{\mathring{h}} = r^2\sin\theta/f(r)$ in Schwarzschild coordinates.

The full set of $\mathcal{O}(\lambda)$ canonical equations of motion for $(\delta\boldsymbol{h}, \delta\boldsymbol{\pi})$ can then be shown to reduce to a set of two simple systems of canonical equations for $(\Phi_{(\pm)}, \Pi_{(\pm)})$. Combining these into second-order equations, they can be written together compactly as:

$$\left(\mathring{\Box} + \frac{8M}{r^3}\mathfrak{Q}_{(\pm)}\right)\Phi_{(\pm)} = 0\,, \tag{3.3.15}$$

where $\mathring{\Box} = \mathring{\boldsymbol{\nabla}}\cdot\mathring{\boldsymbol{\nabla}}$ is the wave operator on $\mathring{\mathscr{M}}$ and

$$\mathfrak{Q}_{(-)} = 1\,, \tag{3.3.16}$$

$$\mathfrak{Q}_{(+)} = \left(\triangle_{\mathbb{S}^2}-1\right)\left(\triangle_{\mathbb{S}^2}+2-\frac{3M}{r}\right)\left(\triangle_{\mathbb{S}^2}+2-\frac{6M}{r}\right)^{-2}\,. \tag{3.3.17}$$

Now to finally simplify (3.3.15) into the classic Regge-Wheeler and Zerilli equations, we can introduce the spherical harmonic decomposition of the master functions (with a conventional $1/r$ factor), i.e. expand them as series in $\{Y^{lm}\}$:

$$\Phi_{(\pm)}\left(t,r,\theta,\phi\right)=\frac{1}{r}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}Y^{lm}\left(\theta,\phi\right)\Psi^{lm}_{(\pm)}(t,r)\,. \tag{3.3.18}$$

Inserting this into (3.3.15) yields the $(1+1)$-dimensional wave equation with a radially-dependent potential:

$$\left(\partial_t^2-\partial_{r_*}^2+V^l_{(\pm)}\left(r\right)\right)\Psi^{lm}_{(\pm)}\left(t,r\right)=0\,. \tag{3.3.19}$$

The wave operator is here written in terms of the tortoise coordinate $r_*=r+2M\ln(\frac{r}{2M}-1)$. The potentials are explicitly

$$V^l_{(-)}\left(r\right)=f\left(r\right)\frac{\Lambda r-6M}{r^3}\,, \tag{3.3.20}$$

$$V^l_{(+)}\left(r\right)=f\left(r\right)\frac{\left(\Lambda-2\right)^2\left(\Lambda r^3+6Mr^2\right)+36\left(\Lambda-2\right)M^2r+72M^3}{r^3\left(\left(\Lambda-2\right)r+6M\right)^2}\,, \tag{3.3.21}$$

where $\Lambda=l\left(l+1\right)$ is minus the eigenvalue of Y^{lm}.

3.4. Perturbations of the Kerr spacetime

3.4.1. The Kerr spacetime and the Newman-Penrose formalism. The metric of the Kerr spacetime $(\mathring{\mathscr{M}},\mathring{\boldsymbol{g}},\mathring{\boldsymbol{\nabla}})$, in Boyer-Lindquist coordinates $\{t,r,\theta,\phi\}$, is given by:

$$\begin{aligned}\mathring{\boldsymbol{g}}=&-\frac{\Delta\left(r\right)}{\Sigma\left(r,\theta\right)}\left(\mathrm{d}t-a\sin^2\theta\mathrm{d}\phi\right)^2+\frac{\sin^2\theta}{\Sigma\left(r,\theta\right)}\left(\left(r^2+a^2\right)\mathrm{d}\varphi-a\mathrm{d}t\right)^2\\&+\frac{\Sigma\left(r,\theta\right)}{\Delta\left(r\right)}\mathrm{d}r^2+\Sigma\left(r,\theta\right)\mathrm{d}\theta^2\,,\end{aligned} \tag{3.4.1}$$

where

$$\Delta\left(r\right)=r^2-2Mr+a^2\,, \tag{3.4.2}$$

$$\Sigma\left(r,\theta\right)=r^2+a^2\cos^2\theta\,. \tag{3.4.3}$$

This metric was discovered by [Kerr 1963]. See [Teukolsky 2015] for more details on the history.

Unlike the Schwarzschild-Droste spacetime, there is no spherical symmetry here to permit tackling the perturbation problem via methods such as the $(2+1+1)$ decomposition and spherical harmonic expansions, as described in the previous section.

Instead, an approach which has proven very useful in this case is the *Newman-Penrose formalism* [E. Newman and Penrose 1962]. This is an approach that can be formulated more generally and powerfully from the point of view of spinor methods (see Chapter 13 of [Wald 1984], Chapter 2 of [J. Stewart 1993], or the lecture notes [Andersson et al. 2016]), but we describe it here more accessibly as a variant of the tetrad method introduced during our discussion on quantum gravity in Chapter 2 (Section 2.5). See Chapter 1 of [Chandrasekhar 1998] for a detailed exposition from this point of view.

The idea is to introduce what is referred to as a *null tetrad*, typically denoted as $\{\mathfrak{e}_I^a\}_{I=0}^3 = \{l^a, n^a, m^a, \bar{m}^a\}$, where the vectors $\boldsymbol{l}$ and $\boldsymbol{n}$ are real, while $\boldsymbol{m}$ and $\bar{\boldsymbol{m}}$ are complex conjugates. Such a tetrad is defined by replacing the flat metric in internal coordinates (the condition $g_{ab}\mathfrak{e}_I^a\mathfrak{e}_J^b = \eta_{IJ}$ used earlier to define tetrads in Section 2.5) instead with the following:

$$g_{ab}\mathfrak{e}_I^a\mathfrak{e}_J^b = \begin{bmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{3.4.4}$$

In other words, $\boldsymbol{l}\cdot\boldsymbol{n} = -1 = -\boldsymbol{m}\cdot\bar{\boldsymbol{m}}$, with all other inner products vanishing. The latter means that all vectors are null, and moreover that $\boldsymbol{m}$ and $\bar{\boldsymbol{m}}$ are orthogonal to $\boldsymbol{l}$ and $\boldsymbol{n}$. Locally, the complex vectors $\boldsymbol{m}$ and $\bar{\boldsymbol{m}}$ can be regarded as complex combinations of two orthonormal spacelike (real) vectors X^a and Y^a which are both orthogonal to the two real null vectors $\boldsymbol{l}$ and $\boldsymbol{n}$; in particular, $\boldsymbol{m} = \frac{1}{\sqrt{2}}(\boldsymbol{X} + \mathrm{i}\boldsymbol{Y})$.

Now consider the Weyl tensor C_{abcd}, defined to be the trace-free part of the Riemann tensor R_{abcd}. (See e.g. Chapter 13 of [Misner et al. 1973]). In dimensions lower than four, C_{abcd} is actually zero (so the Ricci tensor completely determines the Riemann tensor). In four dimensions, it is given by:

$$C^{ab}{}_{cd} = R^{ab}{}_{cd} - 2\delta^{[a}{}_{[c}R^{b]}{}_{d]} + \frac{1}{3}\delta^{[a}{}_{[c}\delta^{b]}{}_{d]}R\,. \tag{3.4.5}$$

The Riemann tensor has twenty independent components, ten of which are accounted for by the Ricci scalar and the other ten by the Weyl tensor. In vacuum spacetimes, the Weyl tensor is thus the same as the Riemann tensor, and is for this reason that it is often said to represent the “purely gravitational degrees of freedom” of GR.

The ten independent components of C_{abcd} can be shown to have a one-to-one correspondence with the following five complex scalars [Chandrasekhar 1998]:

$$\psi_0 = C_{abcd}l^a m^b l^c m^d\,, \tag{3.4.6}$$

$$\psi_1 = C_{abcd}l^a n^b l^c m^d\,, \tag{3.4.7}$$

$$\psi_2 = C_{abcd}l^a m^b \bar{m}^c n^d\,, \tag{3.4.8}$$

$$\psi_3 = C_{abcd}l^a n^b \bar{m}^c n^d\,, \tag{3.4.9}$$

$$\psi_4 = C_{abcd}n^a \bar{m}^b n^c \bar{m}^d\,. \tag{3.4.10}$$

As a concrete example before moving on to discussing perturbations in this context, a commonly used null tetrad which is especially useful for calculations involving the Kerr spacetime is the Kinnersley tetrad [Kinnersley 1969]. In Boyer-Lindquist coordinates,

$$\boldsymbol{l} = \frac{1}{\Delta(r)}\left[\left(r^2+a^2\right)\partial_t + \Delta(r)\,\partial_r + a\partial_\phi\right] , \tag{3.4.11}$$

$$\boldsymbol{n} = \frac{1}{2\Sigma^2(r,\theta)}\left[\left(r^2+a^2\right)\partial_t - \Delta(r)\,\partial_r + a\partial_\phi\right] , \tag{3.4.12}$$

$$\boldsymbol{m} = \frac{1}{\sqrt{2}\,(r+\mathrm{i}a\cos\theta)}\left[\mathrm{i}a\sin\theta\partial_t + \partial_\theta + \frac{\mathrm{i}}{\sin\theta}\partial_\phi\right] . \tag{3.4.13}$$

Using this null tetrad in (3.4.6)-(3.4.10), one finds that all complex scalars vanish except for ψ_2, which is given by:

$$\psi_2 = \frac{M}{(r-\mathrm{i}a\cos\theta)^3} . \tag{3.4.14}$$

3.4.2. The Teukolsky equation. Consider an asymptotically flat background spacetime (not necessarily Kerr) perturbed by a plane gravitational wave that is outgoing near future null infinity. Then one can show that the perturbation to ψ_4, also typically denoted ψ_4 (and we concordantly abuse our notation, where it should be called $\delta\psi_4$), is related to the $+$ and $\times$ (independent) wave polarization modes h_+ and $h_\times$ respectively, according to:

$$\lim_{r\to\infty}\psi_4 = \frac{1}{2}\left(\ddot{h}_+ - \mathrm{i}\ddot{h}_\times\right) . \tag{3.4.15}$$

See e.g. the review [Sasaki and Tagoshi 2003]. In this way, the complex scalar ψ_4 is regarded as describing all pertinent information about outgoing radiation. Hence a general equation for (perturbations of) ψ_4 is of interest for the study of gravitational waveforms.

The idea of the Newman-Penrose approach to Kerr perturbations, then, is to develop evolution equations for the complex scalars $\psi_0,\ldots,\psi_4$ rather than using the (linearized) Einstein equation [Eq. (3.2.15)] for the metric perturbation $\delta\boldsymbol{g}$ directly. This issue presents a number of mathematical subtleties vis-á-vis accounting for the correspondence of the various degrees of freedom; we do not wish to enter deeply into this here, but the basic idea is to use the Einstein equation in the definition of the Weyl tensor (3.4.5), written in terms of the complex scalars, and then to treat the equations for the connection one-forms (see Section 2.5) essentially as the equations of motion.

It turns out that the equation for ψ_4 is non-separable. However, Teukolsky discovered that the equation for a re-scaled perturbation variable,

$$\psi(t,r,\theta,\phi) = \frac{\psi_4(t,r,\theta,\phi)}{\rho^4(\theta)} \quad \text{where} \quad \rho(\theta) = -(r-\mathrm{i}a\cos\theta)^{-1} \tag{3.4.16}$$

led to a separable equation now eponymously named [Teukolsky 1973, 1972]. In vacuum, it reads:

$$
\begin{aligned}
0 = &\left(\frac{\left(r^2+a^2\right)^2}{\Delta\left(r\right)} - a^2\sin^2\theta\right)\partial_t^2\psi + 4\frac{Mar}{\Delta\left(r\right)}\partial_t\partial_\phi\psi + \left(\frac{a^2}{\Delta\left(r\right)} - \frac{1}{\sin^2\theta}\right)\partial_\phi^2\psi \\
&- \frac{1}{\Delta^2\left(r\right)}\partial_r\left(\Delta^3\left(r\right)\partial_r\psi\right) - \frac{1}{\sin\theta}\partial_\theta\left(\sin\theta\partial_\theta\psi\right) \\
&- 4\left(\frac{a\left(r-M\right)}{\Delta\left(r\right)} + \mathrm{i}\frac{\cos\theta}{\sin^2\theta}\right)\partial_\phi\psi - 4\left(\frac{M\left(r^2-a^2\right)}{\Delta\left(r\right)} - r - \mathrm{i}a\cos\theta\right)\partial_t\psi \\
&+ \left(4\cot^2\theta - 2\right)\psi\,.
\end{aligned}
\tag{3.4.17}
$$

(There is a very similar generalization of this equation for scalar, neutrino, and electromagnetic perturbations). See also [Teukolsky 2015] for more technical as well as historical details.

Part II

Novel Contributions:
Entropy, Motion and Self-Force in General Relativity

CHAPTER 4

Entropy Theorems and the Two-Body Problem

Chapter summary. This chapter is based on the publication [Oltean, Bonetti, et al. 2016].

In general Hamiltonian theories, entropy may be understood either as a statistical property of canonical systems (attributable to epistemic ignorance), or as a mechanical property (that is, as a monotonic function on the phase space along trajectories). In classical mechanics, various theorems have been proposed for proving the nonexistence of entropy in the latter sense. Here we explicate, clarify, and extend the proofs of these theorems to some standard matter (scalar and electromagnetic) field theories in curved spacetime, and then we show why these proofs fail in general relativity. As a concrete application, we focus on the consequences of these results for the gravitational two-body problem.

In Section 4.1, we provide a historical overview of these issues following the development of statistical mechanics at the end of the 19th century. We formulate more exactly the problem of explaining the second law of thermodynamics for entropy in the two—statistical and mechanical—senses mentioned above. For the remainder of this chapter, we treat the notion of entropy in the latter (mechanical) sense.

In Section 4.2, we carry out a proof for the nonexistence of entropy in classical mechanics following an idea briefly sketched by Poincaré (and previously never developed into a full proof), following a perturbative approach based on Taylor expansions of Poisson brackets about a hypothetical "thermodynamic equilibrium" point. We also discuss an alternative, topological approach developed by Olsen.

The aim of section 4.3 is then to examine to what extent these proofs carry over to general relativity. We show that the perturbative approach can be used to prove the nonexistence of entropy in standard non-gravitational (scalar and electromagnetic) field theories on curved spacetime, but fails to apply to general relativity itself. We also discuss the topological approach and its failure to prove nonexistence of entropy in general relativity due to the non-compactness of the phase space of the theory.

In Section 4.4, we focus on the gravitational two-body problem in light of these ideas, and in particular, we prove the non-compactness of the phase space of perturbed Schwarzschild-Droste spacetimes. We thus identify the lack of recurring orbits in phase space as a distinct sign of dissipation and hence entropy production.

Section 4.5 offers some concluding remarks.

Teoremes d'entropia i el problema de dos cossos (chapter summary translation in Catalan). Aquest capítol es basa en la publicació [Oltean, Bonetti, et al. 2016].

En teories hamiltonianes generals, l'entropia es pot entendre o bé com una propietat estadística dels sistemes canònics (atribuïble a la ignorància epistèmica), o com a propietat mecànica (és a dir, com a funció monotònica en l'espai de fase al llarg de les trajectòries). En la mecànica clàssica, s'han proposat diversos teoremes per demostrar la inexistència d'entropia en aquest darrer sentit. Aquí expliquem, aclarim i estenem les proves d'aquests teoremes a algunes teories de camps de matèria estàndard (escalar i electromagnètica) en l'espai-temps corbat, i després mostrem per què aquestes proves fracassen en la relativitat general. Com a aplicació concreta, ens centrem en les conseqüències d'aquests resultats sobre el problema gravitatori de dos cossos.

A la secció 4.1, proporcionem una panoràmica històrica d'aquestes qüestions després del desenvolupament de la mecànica estadística a finals del segle XIX. Formulem més exactament el problema d'explicar la segona llei de la termodinàmica per a l'entropia en els dos sentits - estadístic i mecànic - mencionats anteriorment. Per a la resta d'aquest capítol, tractem la noció d'entropia en el darrer sentit (mecànic).

A la secció 4.2, realitzem una prova de la inexistència d'entropia en la mecànica clàssica seguint una idea breument esbossada per Poincaré (i mai abans desenvolupada en una prova completa), seguint un mètode pertorbatiu basat en les expansions de Taylor dels claudàtors de Poisson al voltant d'un hipotètic "equilibri termodinàmic". També discutirem un mètode topològic alternatiu desenvolupat per Olsen.

L'objectiu de la secció 4.3 és examinar fins a quin punt aquestes proves es traslladen a la relativitat general. Mostrem que el mètode pertorbatiu es pot utilitzar per demostrar la inexistència d'entropia en teories estàndard de camp no gravitacionals (escalars i electromagnètiques) en l'espai-temps corbat, però no s'aplica a la mateixa relativitat general. També discutirem el mètode topològic i el seu impossibilitat de demostrar la inexistència d'entropia en la relativitat general a causa de la no compactitat de l'espai de fase de la teoria.

A la secció 4.4, ens centrem en el problema gravitatori de dos cossos a la vista d'aquestes idees i, en particular, demostrem la no compactitat de l'espai de fase dels espais-temps de Schwarzschild-Droste pertorbats. Així identifiquem la manca d'òrbites recurrents en l'espai de fase com a signe distint de dissipació i per tant de producció d'entropia.

La secció 4.5 ofereix algunes observacions finals.

Théorèmes d'entropie et le problème à deux corps (chapter summary translation in French). Ce chapitre est basé sur la publication [Oltean, Bonetti, et al. 2016].

Dans les théories hamiltoniennes générales, l'entropie peut être comprise soit comme une propriété statistique des systèmes canoniques (imputable à l'ignorance épistémique), soit comme une propriété mécanique (c'est-à-dire comme une fonction monotone sur l'espace des phases le long des trajectoires). Dans la mécanique classique, différents théorèmes ont été proposés pour démontrer l'absence d'entropie dans ce dernier sens. Ici, nous expliquons, clarifions et étendons les démonstrations de ces théorèmes à certaines théories de champ standard (scalaire et électromagnétique) dans l'espace-temps courbé, puis nous montrons pourquoi ces démonstrations échouent en relativité générale. Comme application concrète, nous nous concentrons sur les conséquences de ces résultats sur le problème gravitationnel à deux corps.

Dans la section 4.1, nous fournissons un aperçu historique de ces questions suite au développement de la mécanique statistique à la fin du XIXe siècle. Nous formulons plus précisément le problème de l'explication de la deuxième loi de la thermodynamique pour l'entropie dans les deux sens - statistique et mécanique - mentionnés ci-dessus. Pour la suite de ce chapitre, nous traitons la notion d'entropie dans ce dernier sens (mécanique).

Dans la section 4.2, nous effectuons une démonstration de l'absence d'entropie dans la mécanique classique en suivant une idée brièvement exposée par Poincaré (que n'a jamais été transformée en une démonstration complète), en suivant une méthode perturbative basée sur expansions de Taylor, des crochets de Poisson autour d'une hypothétique « équilibre thermodynamique ». Nous discutons également d'une approche topologique alternative développée par Olsen.

Le but de la section 4.3 est alors d'examiner dans quelle mesure ces démonstrations se reportent à la relativité générale. Nous montrons que l'approche perturbative peut être utilisée pour prouver l'absence d'entropie dans les théories standard des champs non gravitationnels (scalaires et électromagnétiques) sur l'espace-temps courbé, mais elle ne s'applique pas à la relativité générale elle-même. Nous discutons également de l'approche topologique et de son incapacité à démontrer l'absence d'entropie dans la relativité générale à cause de la non compacité de l'espace des phases de la théorie.

Dans la section 4.4, nous nous concentrons sur le problème gravitationnel à deux corps a vu de ces idées et en particulier, nous prouvons la non compacité de l'espace des phases d'espaces-temps de Schwarzschild-Droste perturbés. Nous identifions ainsi le manque d'orbites récurrentes dans l'espace des phases comme un signe distinct de dissipation et donc de production d'entropie.

La section 4.5 offre quelques remarques de conclusion.

4.1. Introduction

The problem of reconciling the second law of thermodynamics[1] with classical (deterministic) Hamiltonian evolution is among the oldest in fundamental physics [H. R. Brown et al. 2009; Davies 1977; Sklar 1995]. In the context of classical mechanics (CM), this question motivated much of the development of statistical thermodynamics in the second half of the 19th century. In the context of general relativity (GR), thermodynamic ideas have occupied—and, very likely, will continue to occupy—a central role in our understanding of black holes and efforts to develop a theory of quantum gravity. Indeed, much work in recent years has been expended relating GR and thermodynamics [Rovelli 2012], be it in the form of "entropic gravity" proposals [Carroll and Remmen 2016; Putten 2012; Verlinde 2010] (which derive the Einstein equation from entropy formulas), or gravity-thermodynamics correspondences [Freidel 2015; Freidel and Yokokura 2015] (wherein entropy production in GR is derived from conservation equations, in analogy with classical fluid dynamics). And yet, there is presently little consensus on the general meaning of "the entropy of a gravitational system", and still less on the question of why—purely as a consequence of the dynamical (Hamiltonian) equations of motion—such an entropy should (strictly) monotonically increase in time, *i.e.* obey the second law of thermodynamics.

However one wishes to approach the issue of defining it, gravitational entropy should in some sense emerge from suitably defined (micro-)states associated with the degrees of freedom *not* of any matter content in spacetime, but of the gravitational field itself—which, in GR, means the spacetime geometry—or statistical properties thereof. Of course, we know of restricted situations in GR where we not only have entropy definitions which make sense, but which also manifestly obey the second law—that is, in black hole thermodynamics[2]. In particular, the black hole entropy is identified (up to proportionality) with its area, and hence, we have that the total entropy increases when, say, two initially separated black holes merge—a process resulting, indeed, as a direct consequence of standard evolution of the equations of motion. What is noteworthy about this is that black hole entropy is thus understood not as a statistical idea, but directly as a functional on the phase space of GR (comprising degrees of freedom which are subject to deterministic canonical evolution).

In CM, the question of the statistical nature of entropy dominated many of the early debates on the origin of the second law of thermodynamics during the development of the kinetic theory of gases [Sklar 1995]. Initial hopes, especially by [Boltzmann 1872],

[1] "It is the only physical theory of universal content concerning which I am convinced that, within the framework of applicability of its basic concepts, it will never be overthrown." [Einstein 1949]
[2] This field was initially pioneered by [Bekenstein 1973] and [Hawking 1975]. See [Wald 2001] for a review.

were that entropy could in fact be understood as a (strictly monotonic) function on classical phase space. However, many objections soon appeared which rendered this view problematic—the two most famous being the reversibility argument of [Loschmidt 1876] and the recurrence theorem of [Poincaré 1890].

The Loschmidt reversibility argument, in essence, hinges upon the time-reversal symmetry of the canonical equations of motion, and hence, the ostensibly equal expectation of evolution towards or away from equilibrium. Yet, arguably, this is something which may be circumvented via a sufficiently convincing proposition for identifying the directionality of (some sort of) arrow of time—and in fact, recent work [Barbour, Koslowski, et al. 2013, 2014] shows how this can actually be done in the Newtonian N-body problem, leading in this context to a clearly defined "gravitational" arrow of time. For related work in a cosmological context, see [Sahni, Shtanov, et al. 2015; Sahni and Toporensky 2012].

The Poincaré recurrence argument, on the other hand, relies on a proof that any canonical system in a bounded phase space will always return arbitrarily close to its initial state (and moreover it will do so an unbounded number of times) [Arnold 1997; Luis Barreira 2006]. As the only other assumption needed for this proof is Liouville's theorem (which asserts that, in any Hamiltonian theory, the probability measure for a system to be found in an infinitesimal phase space volume is time independent), the only way for it to be potentially countered is by positing an unbounded phase space for all systems—which clearly is not the case for situations such as an ideal gas in a box.

Such objections impelled the creators of kinetic theory, Maxwell and Boltzmann in particular, to abandon the attempt to understand entropy—in what we may accordingly call a *mechanical* sense—as a phase space function, and instead to conceive of it as a statistical notion the origin of which is *epistemic ignorance*, *i.e.* observational uncertainty of the underlying (deterministic) dynamics. The famous *H-Theorem* of [Boltzmann 1872], which was in fact initially put forth for the purposes of expounding the former, became reinterpreted and propounded in the light of the latter.

Of course, later such a statistical conception of entropy came to be understood in the context of quantum mechanics via the von Neumann entropy (defined in terms of the density matrix of a quantum system) and also in the context of information theory via the Shannon entropy (defined in terms of probabilities of a generic random variable) [Nielsen and Chuang 2010]. Indeed, the meaning of the word "entropy" is nowadays often taken to reflect an observer's knowledge (or ignorance) about the microstates of a system.

Thus, the question of why the second law of thermodynamics should hold in a Hamiltonian system may be construed within two possible formulations—on the one hand, a *mechanical* (or *ontological*), and on the other, a *statistical* (or *epistemic*) point of view. Respectively, we can state these as follows.

Problem I (*mechanical/ontological*): Does there exist a function (or functional, if we are dealing with a field theory) on phase space which monotonically increases along the orbits of the Hamiltonian flow?

Problem II (*statistical/epistemic*): Does there exist a function of time, defined in a suitable way in terms of a probability density on phase space, which always has a non-negative time derivative in a Hamiltonian system?

In the language and notation that we have established in Chapter 2 for describing general canonical systems, we can state these questions more precisely:

Problem I (*mechanical/ontological*): Does there exist any function on the (reduced, if there are constraints) phase space $S : \mathscr{S} \to \mathbb{R}$ which monotonically increases along the orbits of (the Hamiltonian flow) Φ_t?

Problem II (*statistical/epistemic*): Does there exist any function of time $S : \mathscr{T} \to \mathbb{R}$, on some time interval $\mathscr{T} \subseteq \mathbb{R}$, defined in a suitable way in terms of a probability density $\rho : \mathscr{S} \times \mathscr{T} \to [0, 1]$ on the (reduced) phase space, satisfying $\mathrm{d}S/\mathrm{d}t \geq 0$ in a Hamiltonian system? [Traditionally, the definition taken here for entropy is (a coarse-grained version of) $S(t) = -\int_{\mathscr{P}} \boldsymbol{\Omega}\, \rho \ln \rho$, or its appropriate reduction to $\mathscr{S}$ if there are constraints.]

In CM, it is Problem II that has received the most attention since the end of the 19th century. In fact, there has been significant work in recent years by mathematicians [Villani 2008; Yau 2011] aimed at placing the statistical formulation of the H-Theorem on more rigorous footing, and thus at proving more persuasively that, using appropriate assumptions, the answer to Problem II is in fact *yes*. In contrast, after the early Loschmidt reversibility and Poincaré recurrence arguments, Problem I has received some less well-known responses to the effect of demonstrating (even more convincingly) that the answer to it *under certain conditions* (to be carefully elaborated) is actually *no*. In this chapter, we will concern ourselves with two such types of responses to Problem I: first, what we call the *perturbative* approach, also proposed by [Poincaré 1889]; and second, what we call the *topological* approach, due to [Olsen 1993] and related to the recurrence theorem. In the former, one tries to Taylor expand the time derivative of a phase space function, computed via the Poisson bracket, about a hypothetical equilibrium point in phase space, and one obtains contradictions with its strict positivity away from equilibrium. We revisit the original paper of Poincaré, clarify the assumptions of the argument, and carefully carry out the proof which is—excepting a sketch which makes it seem more trivial than it actually turns out to be—omitted therein. We furthermore extend this theorem to matter fields—in particular, a scalar and electromagnetic field—in curved spacetime. In the topological approach, on the other hand, one uses topological properties of the phase space itself to prove non-existence of monotonic functions. We review the proof of Olsen, and discuss its connections with the recurrence theorem and more recent periodicity theorems in Hamiltonian systems from symplectic geometry.

In GR, one may consider similar lines of reasoning as in CM to attempt to answer Problems I and II. Naively, one might expect the same answer to Problem II, namely *yes*—however, as we will argue later in greater detail, there are nontrivial mathematical issues that need to be circumvented here even in formulating it. For Problem I, as discussed, one might confidently expect the answer to also be *yes*. Therefore, although we do not yet know how to define entropy in GR with complete generality, we *can* at least ask why the proofs that furnish a negative answer to Problem I in CM fail here, and perhaps thereby gain fruitful insight into the essential features we should expect of such a definition.

Following the perturbative approach, we will show that a Taylor-expanded Poisson bracket does not contain terms which satisfy definite inequalities (as they do in CM). The reason, as we will see, is that the second functional derivatives of the gravitational Hamiltonian can (unlike in CM) be both positive and negative, and so its curvature in phase space cannot be used to constrain (functionals of) the orbits; no contradiction arises here with the second law of thermodynamics.

Following the topological approach, there are two points of view which may explicate why the proofs in CM do not carry over to GR. Firstly, it is believed that, in general, the phase space of GR is non-compact [Schiffrin and Wald 2012]. Of course, this assertion depends on the nature of the degrees of freedom thought to be available in the spacetime under consideration, but even in very simple situations (such as cosmological spacetimes), it has been shown explicitly that the total phase space measure diverges. Physically, what this non-compactness implies is the freedom of a gravitational system to explore phase space unboundedly, without having to return (again and again) to its initial state. This leads us to the second (related) point of view as to why the topological proofs in CM fail in GR: namely, the non-recurrence of phase space orbits. Aside from trivial situations, solutions to the canonical equations of GR are typically non-cyclic (*i.e.* they do not close in phase space) permitting the existence of functionals which may thus increase along the Hamiltonian flow. In fact, to counter the Poincaré recurrence theorem in CM, there even exists a “no-return” theorem in GR [R. P. A. C. Newman 1986; Tipler 1980, 1979] for spacetimes which admit compact Cauchy surfaces and satisfy suitable energy and genericity conditions; it broadly states that the spacetime cannot return, even arbitrarily close, to a previously occupied state. One might nonetheless expect non-recurrence to be a completely general feature of all (nontrivial) gravitational systems, including spacetimes with non-compact Cauchy surfaces.

A setting of particular interest for this discussion is the gravitational two-body problem. With the recent detections and ongoing efforts towards further observations of gravitational waves from two-body systems, the emission of which ought to be closely related to entropy production, a precise understanding and quantification of the latter is becoming more and more salient. In the CM two-body (*i.e.* Kepler) problem, the consideration of Problem I clearly explains the lack of entropy production due to phase space compactness (for a given finite range of initial conditions). In the Newtonian N-body problem, where (as we will elaborate) neither the perturbative nor the topological proofs are applicable, the answer to Problem I was actually shown to be *yes* in [Barbour, Koslowski, et al. 2013, 2014]. In GR, the two-body problem may be considered in the context of perturbed Schwarzschild-Droste (SD) spacetimes (as is relevant, for instance, in the context of extreme-mass-ratio inspirals). Here, the phase space volume (symplectic) form has been explicitly computed in [Jezierski 1999]. We will use this in this chapter to show that in such spacetimes, the phase space is non-compact; hence there are no contradictions with non-recurrence or entropy production.

4.2. Entropy theorems in classical mechanics

4.2.1. Setup. Classical particle mechanics with N degrees of freedom [Arnold 1997] can be formulated as a Lagrangian theory with an N-dimensional configuration space $\mathscr{Q}$. This means that we will have a canonical theory[3] on a $2N$-dimensional phase space $\mathscr{P}$. We can choose canonical coordinates $(q_1, ..., q_N)$ with conjugate momenta $(p_1, ..., p_N)$ such that the symplectic form on $\mathscr{P}$ is given by

$$\boldsymbol{\omega} = \sum_{j=1}^{N} \mathrm{d}p_j \wedge \mathrm{d}q_j. \tag{4.2.1}$$

Then, the volume form on $\mathscr{P}$ is simply the N-th exterior power of the symplectic form, in particular $\boldsymbol{\Omega} = [(-1)^{N(N-1)/2}/N!]\boldsymbol{\omega}^{\wedge N}$, and $\boldsymbol{X}_H$ is here given in coordinates by

$$\boldsymbol{X}_H = \sum_{j=1}^{N} \left(\frac{\partial H}{\partial p_j}\frac{\partial}{\partial q_j} - \frac{\partial H}{\partial q_j}\frac{\partial}{\partial p_j} \right). \tag{4.2.2}$$

[3] Note that the formulation here proceeds essentially exactly as in the case of field theories, elaborated at length in Chapter 2. In fact, the treatment of particle mechanics can be regarded mathematically as just a special case of the general treatment of field theories, in particular by using a collection of (N) fields which are all distributional (Dirac delta functions in three-dimensional space) and supported at the respective particle locations.

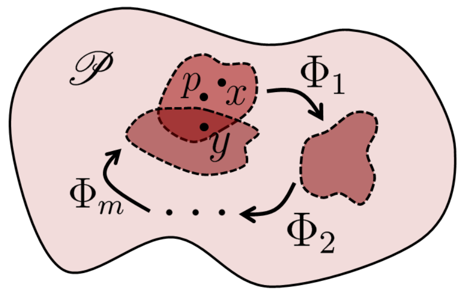

FIGURE 4.1. The idea of the proof for the Poincaré recurrence theorem.

The action of $\boldsymbol{X}_H$ on any phase space function $F : \mathscr{P} \to \mathbb{R}$, called the Poisson bracket, gives its time derivative:

$$\dot{F} = \frac{\mathrm{d}F}{\mathrm{d}t} = \boldsymbol{X}_H(F) = \{F, H\}\,. \tag{4.2.3}$$

We obtain from this $\dot{q}_j = \{q_j, H\} = \partial H/\partial p_j$ and $\dot{p}_j = \{p_j, H\} = -\partial H/\partial q_j$, which are the canonical equations of motion. Moreover, we have that the symplectic form of $\mathscr{P}$, and hence its volume form, are preserved along Φ_t; in other words, we have $\mathcal{L}_{\boldsymbol{X}_H}\boldsymbol{\omega} = 0 = \mathcal{L}_{\boldsymbol{X}_H}\boldsymbol{\Omega}$, which is known as Liouville's theorem.

We now turn to addressing Problem I in CM—that is, the question of whether there exists a function $S : \mathscr{P} \to \mathbb{R}$ that behaves like entropy in a classical Hamiltonian system. Possibly the most well-known answer given to this is the Poincaré recurrence theorem. We can easily offer a proof of this, shown pictorially in Figure 4.1 (see section 16 of [Arnold 1997]): Assume $\mathscr{P}$ is compact and $\Phi_t(\mathscr{P}) = \mathscr{P}$. Let $\mathscr{U} \subset \mathscr{P}$ be the neighborhood of any point $p \in \mathscr{P}$, and consider the sequence of images $\{\Phi_n(\mathscr{U})\}_{n=0}^{\infty}$. Each $\Phi_n(\mathscr{U})$ has the same measure $\int_{\Phi_n(\mathscr{U})}\boldsymbol{\Omega}$ (because of Liouville's theorem), so if they never intersected, $\mathscr{P}$ would have infinite measure. Therefore there exist k, l with $k > l$ such that $\Phi_k(\mathscr{U}) \cap \Phi_l(\mathscr{U}) \neq \varnothing$, implying $\Phi_m(\mathscr{U}) \cap \mathscr{U} \neq \varnothing$ where $m = k - l$. For any $y \in \Phi_m(\mathscr{U}) \cap \mathscr{U}$, there exists an $x \in \mathscr{U}$ such that $y = \Phi_m(x)$. Thus, any point returns arbitrarily close to the initial conditions in a compact and invariant phase space.

Let us now discuss, in turn, the perturbative and topological approaches to this problem.

4.2.2. Perturbative approach. We revisit and carefully explicate, in this subsection, the argument given by [Poincaré 1889] to the effect that an entropy function $S : \mathscr{P} \to \mathbb{R}$ does not exist. First, we will clarify the assumptions that need to go into it, *i.e.* the conditions we must impose both on the entropy S as well as on the Hamiltonian H, and then we will supply a rigorous proof.

4.2.2.1. *Review of Poincaré's idea for a proof.* In his original paper [Poincaré 1889] (translated into English in [Olsen 1993]), the argument given by Poincaré (expressed using the contemporary notation of this paper) for the non-existence of such a function $S : \mathscr{P} \to \mathbb{R}$ is the following: if S behaves indeed like entropy, it should satisfy

$$\dot{S} = \{S, H\} = \sum_{k=1}^{N} \left(\frac{\partial H}{\partial p_k} \frac{\partial S}{\partial q_k} - \frac{\partial H}{\partial q_k} \frac{\partial S}{\partial p_k} \right) > 0 \tag{4.2.4}$$

around a hypothetical equilibrium point in $\mathscr{P}$. Taylor expanding each term and assuming all first partials of S and H vanish at this equilibrium, we obtain a quadratic form (in the distances away from equilibrium) plus higher-order terms. If we are "sufficiently close" to equilibrium, we may ignore the higher-order terms and simply consider the quadratic form, which thus needs to be positive definite for the above inequality [Eq. (4.2.4)] to hold. But here Poincaré, without presenting any further explicit computations, simply asserts that "it is easy to satisfy oneself that this is impossible if one or the other of the two *quadratic forms S and H* is definite, which is the case here." (Our modern language modification is in italic.)

Neither the casual dismissal of the higher-order terms, nor, even more crucially, the fact that "it is easy to satisfy oneself" of the impossibility of this quadratic form to be positive definite is immediately apparent from this discussion. In fact, all of the points in this line of reasoning require a careful statement of the necessary assumptions as well as some rather non-trivial details of the argumentation required to obtain the conclusion (that $\dot{S} = 0$).

In what follows, we undertake precisely that. First we look at the assumptions needed for this method to yield a useful proof, and then we carry out the proof in full detail and rigor.

4.2.2.2. *Entropy conditions.* A function $S : \mathscr{P} \to \mathbb{R}$ can be said to behave like entropy insofar as it satisfies the laws of thermodynamics. In particular, it should conform to two assumptions: first, that it should have an equilibrium point, and second, that it should obey the second law of thermodynamics—which heuristically states that it should be increasing in time everywhere except at the equilibrium point, where it should cease to change. We state these explicitly as follows:

S1 *(Existence of equilibrium)***:** We assume there exists a point in phase space, $x_0 \in \mathscr{P}$, henceforth referred to as the "equilibrium" configuration of the system, which is a stationary point of the entropy S, *i.e.* all first partials thereof should vanish when evaluated there:

$$\left(\frac{\partial S}{\partial q_j} \right)_0 = 0 = \left(\frac{\partial S}{\partial p_j} \right)_0 , \tag{4.2.5}$$

where, for convenience, we use the notation $(\cdot)_0 = (\cdot)|_{x_0}$ to indicate quantities evaluated at equilibrium. Note that by the definition of the Poisson bracket [Eq. (4.2.3)], this implies $(\dot{S})_0 = 0$.

S2 *(Second law of thermodynamics)***:** A common formulation of the second law asserts that the entropy S is always increasing in time when the system is away from equilibrium (*i.e.* $\dot{S} > 0$ everywhere in $\mathscr{P}\backslash x_0$), and attains its maximum value at equilibrium, where it ceases to change in time (*i.e.* $\dot{S} = 0$ at x_0, as implied by the first condition). We need to work, however, with a slightly stronger version of the second law: namely, the requirement that the Hessian matrix of $\dot{S}$,

$$\mathbf{Hess}(\dot{S}) = \left[\begin{array}{c|c} \dfrac{\partial^2 \dot{S}}{\partial q_i \partial q_j} & \dfrac{\partial^2 \dot{S}}{\partial q_i \partial p_j} \\ \hline \dfrac{\partial^2 \dot{S}}{\partial p_i \partial q_j} & \dfrac{\partial^2 \dot{S}}{\partial p_i \partial p_j} \end{array}\right], \tag{4.2.6}$$

is positive definite when evaluated at equilibrium, *i.e.* $(\mathbf{Hess}(\dot{S}))_0 \succ 0$.

We make now a few remarks about these assumptions. Firstly, S2 is a sufficient—though not strictly necessary—condition to guarantee $\dot{S} > 0$ in $\mathscr{P}\backslash x_0$ and $\dot{S} = 0$ at x_0. However, the assumption of positive definiteness of the Hessian of the entropy S itself at equilibrium is often used in statistical mechanics [Abad 2012], and so it may not be objectionable to extend this supposition to $\dot{S}$ as well. (In any case, this leaves out only special situations where higher-order derivative tests are needed to certify the global minimization of $\dot{S}$ at equilibrium, which arguably are more of mathematical rather than physical interest; we may reasonably expect the entropy as well as its time derivative to be quadratic in the phase space variables as a consequence of its ordinary statistical mechanics definitions in terms of energy.)

Secondly, the above two conditions omit the consideration of functions on $\mathscr{P}$ which are *everywhere* strictly monotonically increasing in time, *i.e.* the time derivative of which is always positive with no equilibrium point. The topological approaches to Problem I, which we will turn to in the next subsection, do accommodate the possibility such functions.

Thirdly, the equilibrium point $x_0 \in \mathscr{P}$, though usually (physically) expected to be unique, need not be for the purposes of what follows, so long as it obeys the two conditions S1 and S2. In other words, it suffices that there exists at least one such point in $\mathscr{P}$.

Fourthly, there is no topological requirement being imposed on the phase space $\mathscr{P}$. It is possible, in other words, for its total measure $\mu(\mathscr{P}) = \int_{\mathscr{P}} \mathbf{\Omega}$ to diverge. This means that the theorem applies to systems which can, in principle, explore phase space unboundedly, without any limits being imposed (either physically or mathematically) thereon.

4.2.2.3. *Hamiltonian conditions.* Next, we make a few assumptions about the Hamiltonian $H : \mathscr{P} \to \mathbb{R}$ which we need to impose in order to carry out our proof. The first two assumptions are reasonable for any typical Hamiltonian in classical mechanics, as we will discuss. The third, however, is stronger than necessary to account for all Hamiltonians in general—and indeed, as we will see, unfortunately leaves out certain classes of Hamiltonians of interest. However, we regard it as a necessary assumption which we cannot relax in order to formulate the proof according to this approach. Our assumptions on H are thus as follows:

H1 *(Kinetic terms)***:** With regards to the second partials of H with respect to the momentum variables, we assume the following:

(a) We can make a choice of coordinates so as to diagonalize (*i.e.* decouple) the kinetic terms. In other words, we can choose to write H in such a form that we have:

$$\frac{\partial^2 H}{\partial p_i \partial p_j} = \delta_{ij} \frac{\partial^2 H}{\partial p_j^2} \,. \tag{4.2.7}$$

(b) Additionally, the second partials of H with respect to each momentum variable, representing the coefficients of the kinetic terms, should be non-negative:

$$\frac{\partial^2 H}{\partial p_j^2} \geq 0 \,. \tag{4.2.8}$$

H2 *(Mixed terms)***:** We assume that we can decouple the terms that mix kinetic and coordinate degrees of freedom (via performing integrations by parts, if necessary, in the action out of which the Hamiltonian is constructed), such that H can be written in a form where:

$$\frac{\partial^2 H}{\partial p_i \partial q_j} = 0 \,. \tag{4.2.9}$$

H3 *(Potential terms)***:** We need to restrict our consideration to Hamiltonians the partial Hessian of which, with respect to the coordinate variables, is positive semidefinite at the point of equilibrium (assuming it exists), *i.e.* $[\partial^2 H/\partial q_i \partial q_j]_0 \succeq 0$. In fact we need to impose a slightly stronger (sufficient, though not strictly necessary) condition: that any of the row sums of $[\partial^2 H/\partial q_i \partial q_j]_0$ are non-negative. That is to say, we assume:

$$\sum_{i=1}^{N} \left(\frac{\partial^2 H}{\partial q_i \partial q_j} \right)_0 \geq 0. \tag{4.2.10}$$

We can make a few remarks about these assumptions. Firstly, H1 and H2 are manifestly satisfied for the most typically-encountered form of the Hamiltonian in CM,

$$H = \sum_{j=1}^{N} \frac{p_j^2}{2m_j} + V\left(q_1, ..., q_N\right), \tag{4.2.11}$$

where m_j are the masses associated with each degree of freedom and V is the potential (a function of only the configuration variables, and not the momenta). Indeed, H1(a) is satisfied since we have $\partial^2 H/\partial p_i \partial p_j = 0$ unless $i = j$, regardless of V. For H1(b), we clearly have $\partial^2 H/\partial p_j^2 = 1/m_j > 0$ assuming masses are positive. Finally, H2 holds as $\partial^2 H/\partial p_i \partial q_j = 0$ is satisfied by construction.

Secondly, For typical Hamiltonians [Eq. (4.2.11)], H3 translates into a condition on the potential V, *i.e.* the requirement that $\sum_{i=1}^{N}(\partial^2 V/\partial q_i \partial q_j)_0 \geq 0$. This is not necessarily satisfied in general in CM, though it is for many systems. For example, when we have just one degree of freedom, $N = 1$, this simply means that the potential $V(q)$ is concave upward at the point of equilibrium (thus regarded as a *stable* equilibrium), *i.e.* $(\mathrm{d}^2 V(q)/\mathrm{d}q^2)_0 \geq 0$, which is reasonable to assume. As another example, for a system of harmonic oscillators with no interactions, $V = \frac{1}{2}\sum_{j=1}^{N} m_j \omega_j^2 q_j^2$, we clearly have $\sum_{i=1}^{N} \partial^2 V/\partial q_i \partial q_j = m_j \omega_j^2 > 0$ for positive masses. Indeed, even introducing interactions does not change this so long as the couplings are mostly non-negative. (In other words, if the negative couplings do not dominate in strength over the positive ones.) Higher (positive) powers of the q_j variables in V are also admissible under a similar argument. However, we can see that condition H3 [Eq. (4.2.10)] excludes certain classes of inverse-power potentials. Most notably, it excludes the Kepler (gravitational two-body) Hamiltonian, $H = (1/2m)\,(p_1^2 + p_2^2) - GMm/(q_1^2 + q_2^2)^{1/2}$, where q_j are the Cartesian coordinates in the orbital plane, and p_j the associated momenta. In this case, we have $\det([\partial^2 H/\partial q_i \partial q_j]) = -2\,(GMm)^2/(q_1^2 + q_2^2)^3 < 0$, hence $[\partial^2 H/\partial q_i \partial q_j]$ is negative definite everywhere and therefore cannot satisfy H3 [Eq. (4.2.10)].

4.2.2.4. *Our proof.* We will now show that there cannot exist a function $S : \mathscr{P} \to \mathbb{R}$ satisfying the assumptions S1-S2 of subsubsection 4.2.2.2 in a Hamiltonian system that obeys the assumptions H1-H3 of subsubsection 4.2.2.3 on $H : \mathscr{P} \to \mathbb{R}$. We do this by simply assuming that such a function exists, and we will show that this implies a contradiction. For a pictorial representation, see Figure 4.2.

$\boldsymbol{N = 1}$**:** Let us first carry out the proof for $N = 1$ degree of freedom so as to make the argument for general N easier to follow. Let $S : \mathscr{P} \to \mathbb{R}$ be any function on the configuration space $\mathscr{P}$ satisfying assumptions S1-S2 of subsubsection 4.2.2.2, *i.e.* it has an equilibrium point and the Hessian of its time derivative is positive definite there. We know that its time derivative at any point $x = (q,p) \in \mathscr{P}$ can be evaluated, as discussed in subsection 4.2.1, via the Poisson bracket:

$$\dot{S} = \frac{\partial H}{\partial p}\frac{\partial S}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial S}{\partial p}. \tag{4.2.12}$$

Let us now insert into this the Taylor series for each term expanded about the equilibrium point $x_0 = (q_0, p_0)$. Denoting $\Delta q = q - q_0$ and $\Delta p = p - p_0$, and using $\mathcal{O}(\Delta^n)$ to

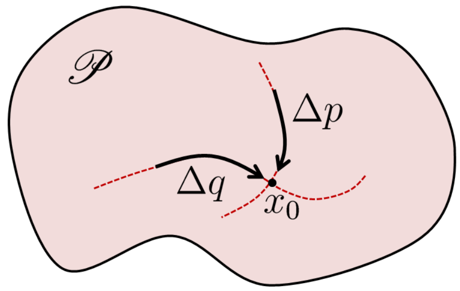

FIGURE 4.2. The idea of the perturbative approach is to evaluate $\dot{S}$ along different directions in phase space away from equilibrium, and arrive at a contradiction with its strict positivity.

represent n-th order terms in products of Δq and Δp, we have:

$$\frac{\partial H}{\partial q} = \left(\frac{\partial H}{\partial q}\right)_0 + \left(\frac{\partial^2 H}{\partial q^2}\right)_0 \Delta q + \left(\frac{\partial^2 H}{\partial p \partial q}\right)_0 \Delta p + \mathcal{O}\left(\Delta^2\right), \tag{4.2.13}$$

and similarly for the p partial of H, while

$$\frac{\partial S}{\partial q} = \left(\frac{\partial^2 S}{\partial q^2}\right)_0 \Delta q + \left(\frac{\partial^2 S}{\partial p \partial q}\right)_0 \Delta p + \mathcal{O}\left(\Delta^2\right), \tag{4.2.14}$$

and similarly for the p partial of S, where we have used the condition S1 [Eq. (4.2.5)] which entails that the zero-order term vanishes. Inserting all Taylor series into the Poisson bracket [Eq. (4.2.12)] and collecting terms, we obtain the following result:

$$\dot{S} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \Delta q \\ \Delta p \end{bmatrix} + \begin{bmatrix} \Delta q & \Delta p \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} \Delta q \\ \Delta p \end{bmatrix} + \mathcal{O}\left(\Delta^3\right), \tag{4.2.15}$$

where:

$$a = \left(\frac{\partial H}{\partial p}\right)_0 \left(\frac{\partial^2 S}{\partial q^2}\right)_0 - \left(\frac{\partial H}{\partial q}\right)_0 \left(\frac{\partial^2 S}{\partial q \partial p}\right)_0, \tag{4.2.16}$$

$$b = \left(\frac{\partial H}{\partial p}\right)_0 \left(\frac{\partial^2 S}{\partial p \partial q}\right)_0 - \left(\frac{\partial H}{\partial q}\right)_0 \left(\frac{\partial^2 S}{\partial p^2}\right)_0, \tag{4.2.17}$$

and:

$$A = \left(\frac{\partial^2 H}{\partial q \partial p}\right)_0 \left(\frac{\partial^2 S}{\partial q^2}\right)_0 - \left(\frac{\partial^2 H}{\partial q^2}\right)_0 \left(\frac{\partial^2 S}{\partial q \partial p}\right)_0, \tag{4.2.18}$$

$$B = \frac{1}{2}\left[\left(\frac{\partial^2 H}{\partial p^2}\right)_0 \left(\frac{\partial^2 S}{\partial q^2}\right)_0 - \left(\frac{\partial^2 H}{\partial q^2}\right)_0 \left(\frac{\partial^2 S}{\partial p^2}\right)_0\right], \tag{4.2.19}$$

$$C = \left(\frac{\partial^2 H}{\partial p^2}\right)_0 \left(\frac{\partial^2 S}{\partial p \partial q}\right)_0 - \left(\frac{\partial^2 H}{\partial p \partial q}\right)_0 \left(\frac{\partial^2 S}{\partial p^2}\right)_0. \tag{4.2.20}$$

By assumption S2, we have that $\dot{S}$ as given above [Eq. (4.2.15)] is strictly positive for any $x \neq x_0$ in $\mathscr{P}$. In particular, let $\delta > 0$ and let us consider $\dot{S}$ [Eq. (4.2.15)] evaluated at the sequence of points $\{x_n^{\pm}\}_{n=1}^{\infty}$, where $x_n^{\pm} = (q_0 \pm \delta/n, p_0)$, such that the only deviation away from equilibrium is along the direction $\Delta q = \pm\delta/n$, with all other Δp vanishing. Then, for any n, we must have according to our expression for $\dot{S}$ [Eq. (4.2.15)]:

$$\dot{S}\left(x_n^{+}\right) = a\frac{\delta}{n} + A\frac{\delta^2}{n^2} + \mathcal{O}\left(\frac{\delta^3}{n^3}\right) > 0, \tag{4.2.21}$$

$$\dot{S}\left(x_n^{-}\right) = -a\frac{\delta}{n} + A\frac{\delta^2}{n^2} + \mathcal{O}\left(\frac{\delta^3}{n^3}\right) > 0. \tag{4.2.22}$$

Taking the $n \to \infty$ limit of the first inequality implies $a \geq 0$, while doing the same for the second inequality implies $a \leq 0$. Hence $a = 0$. A similar argument (using $\Delta p = \pm\delta/n$) implies $b = 0$. Thus, S needs to satisfy the constraints

$$0 = \left(\frac{\partial H}{\partial p}\right)_0 \left(\frac{\partial^2 S}{\partial q^2}\right)_0 - \left(\frac{\partial H}{\partial q}\right)_0 \left(\frac{\partial^2 S}{\partial q \partial p}\right)_0, \tag{4.2.23}$$

$$0 = \left(\frac{\partial H}{\partial p}\right)_0 \left(\frac{\partial^2 S}{\partial p \partial q}\right)_0 - \left(\frac{\partial H}{\partial q}\right)_0 \left(\frac{\partial^2 S}{\partial p^2}\right)_0, \tag{4.2.24}$$

and this leaves us with

$$\dot{S} = \begin{bmatrix} \Delta q & \Delta p \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} \Delta q \\ \Delta p \end{bmatrix} + \mathcal{O}\left(\Delta^3\right). \tag{4.2.25}$$

Now, imposing the Hamiltonian assumptions H1(a) and H2 [eqs. (4.2.7) and (4.2.9) respectively] simplifies A and C, from the above [eqs. (4.2.18) and (4.2.19) respectively] to:

$$A = -\left(\frac{\partial^2 H}{\partial q^2}\right)_0 \left(\frac{\partial^2 S}{\partial q \partial p}\right)_0, \tag{4.2.26}$$

$$C = \left(\frac{\partial^2 H}{\partial p^2}\right)_0 \left(\frac{\partial^2 S}{\partial p \partial q}\right)_0. \tag{4.2.27}$$

Positive-definiteness of $(\mathbf{Hess}(\dot{S}))_0$ (assumption S2) implies that the quadratic form above [Eq. (4.2.25)] should be positive definite. This means that we cannot have $(\partial^2 H/\partial p^2)_0 = 0$, since then C would not be strictly positive and we would get a contradiction. This, combined with assumption H1(b) [Eq. (4.2.8)], implies that $(\partial^2 H/\partial p^2)_0 > 0$. This in combination with $C > 0$ means that $(\partial^2 S/\partial p \partial q)_0 > 0$. But $A > 0$ also, in order to have positive-definiteness of the quadratic form [Eq. (4.2.25)], and this combined with assumption H3 [Eq. (4.2.10)], *i.e.* $(\partial^2 H/\partial q^2)_0 \geq 0$, implies $(\partial^2 S/\partial p \partial q)_0 < 0$. Thus we get a contradiction, and so no such function S exists.

General N: The extension of the proof to general N follows similar lines, though with a few added subtleties. Let us now proceed with it. As before, suppose $S : \mathscr{P} \to \mathbb{R}$ is any function on $\mathscr{P}$ satisfying S1-S2. Its time derivative at any point $x = (q_1, ..., q_N, p_1, ..., p_N) \in \mathscr{P}$ can be evaluated via the Poisson bracket:

$$\dot{S} = \sum_{k=1}^{N} \left(\frac{\partial H}{\partial p_k} \frac{\partial S}{\partial q_k} - \frac{\partial H}{\partial q_k} \frac{\partial S}{\partial p_k} \right). \tag{4.2.28}$$

Let us now insert into this the Taylor series for each term expanded about the equilibrium point $x_0 = ((q_0)_1, ..., (q_0)_N, (p_0)_1, ..., (p_0)_1)$. Denoting $\Delta q_i = q_i - (q_0)_i$ and $\Delta p_i = p_i - (p_0)_i$, and using $\mathcal{O}(\Delta^n)$ to represent n-th order terms in products of Δq_i and Δp_i, we have:

$$\frac{\partial H}{\partial q_k} = \left(\frac{\partial H}{\partial q_k}\right)_0 + \sum_{i=1}^{N} \left[\left(\frac{\partial^2 H}{\partial q_i \partial q_k}\right)_0 \Delta q_i + \left(\frac{\partial^2 H}{\partial p_i \partial q_k}\right)_0 \Delta p_i \right] + \mathcal{O}\left(\Delta^2\right), \tag{4.2.29}$$

and similarly for the p_k partial of H, while

$$\frac{\partial S}{\partial q_k} = \sum_{i=1}^{N} \left[\left(\frac{\partial^2 S}{\partial q_i \partial q_k}\right)_0 \Delta q_i + \left(\frac{\partial^2 S}{\partial p_i \partial q_k}\right)_0 \Delta p_i \right] + \mathcal{O}\left(\Delta^2\right), \tag{4.2.30}$$

and similarly for the p_k partial of S, where we have used the condition S1 [Eq. (4.2.5)] which entails that the zero-order term vanishes. Inserting all Taylor series into the Poisson bracket [Eq. (4.2.28)] and collecting terms, we obtain the following result:

$$\dot{S} = \begin{bmatrix} \mathbf{a}^{\mathrm{T}} & \mathbf{b}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \Delta q_1 \\ \vdots \\ \Delta p_N \end{bmatrix} + \begin{bmatrix} \Delta q_1 & \cdots & \Delta p_N \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \Delta q_1 \\ \vdots \\ \Delta p_N \end{bmatrix} + \mathcal{O}\left(\Delta^3\right), \tag{4.2.31}$$

where we have the following components for the N-dimensional vectors:

$$a_i = \sum_{k=1}^{N} \left[\left(\frac{\partial H}{\partial p_k} \right)_0 \left(\frac{\partial^2 S}{\partial q_i \partial q_k} \right)_0 - \left(\frac{\partial H}{\partial q_k} \right)_0 \left(\frac{\partial^2 S}{\partial q_i \partial p_k} \right)_0 \right], \tag{4.2.32}$$

and

$$b_i = \sum_{k=1}^{N} \left[\left(\frac{\partial H}{\partial p_k} \right)_0 \left(\frac{\partial^2 S}{\partial p_i \partial q_k} \right)_0 - \left(\frac{\partial H}{\partial q_k} \right)_0 \left(\frac{\partial^2 S}{\partial p_i \partial p_k} \right)_0 \right], \tag{4.2.33}$$

and for the $N \times N$ matrices:

$$\begin{aligned} A_{ij} =& \frac{1}{2} \sum_{k=1}^{N} \left[\left(\frac{\partial^2 H}{\partial q_i \partial p_k} \right)_0 \left(\frac{\partial^2 S}{\partial q_j \partial q_k} \right)_0 + \left(\frac{\partial^2 H}{\partial q_j \partial p_k} \right)_0 \left(\frac{\partial^2 S}{\partial q_i \partial q_k} \right)_0 \right. \\ & \left. - \left(\frac{\partial^2 H}{\partial q_i \partial q_k} \right)_0 \left(\frac{\partial^2 S}{\partial q_j \partial p_k} \right)_0 - \left(\frac{\partial^2 H}{\partial q_j \partial q_k} \right)_0 \left(\frac{\partial^2 S}{\partial q_i \partial p_k} \right)_0 \right], \end{aligned} \tag{4.2.34}$$

$$\begin{aligned} B_{ij} =& \frac{1}{2} \sum_{k=1}^{N} \left[\left(\frac{\partial^2 H}{\partial q_i \partial p_k} \right)_0 \left(\frac{\partial^2 S}{\partial p_j \partial q_k} \right)_0 + \left(\frac{\partial^2 H}{\partial p_j \partial p_k} \right)_0 \left(\frac{\partial^2 S}{\partial q_i \partial q_k} \right)_0 \right. \\ & \left. - \left(\frac{\partial^2 H}{\partial q_i \partial q_k} \right)_0 \left(\frac{\partial^2 S}{\partial p_j \partial p_k} \right)_0 - \left(\frac{\partial^2 H}{\partial p_j \partial q_k} \right)_0 \left(\frac{\partial^2 S}{\partial q_i \partial p_k} \right)_0 \right], \end{aligned} \tag{4.2.35}$$

$$\begin{aligned} C_{ij} =& \frac{1}{2} \sum_{k=1}^{N} \left[\left(\frac{\partial^2 H}{\partial p_i \partial p_k} \right)_0 \left(\frac{\partial^2 S}{\partial p_j \partial q_k} \right)_0 + \left(\frac{\partial^2 H}{\partial p_j \partial p_k} \right)_0 \left(\frac{\partial^2 S}{\partial p_i \partial q_k} \right)_0 \right. \\ & \left. - \left(\frac{\partial^2 H}{\partial p_i \partial q_k} \right)_0 \left(\frac{\partial^2 S}{\partial p_j \partial p_k} \right)_0 - \left(\frac{\partial^2 H}{\partial p_j \partial q_k} \right)_0 \left(\frac{\partial^2 S}{\partial p_i \partial p_k} \right)_0 \right]. \end{aligned} \tag{4.2.36}$$

By assumption S2, we have that $\dot{S}$ as given above [Eq. (4.2.31)] is strictly positive for any $x \neq x_0$ in $\mathscr{P}$. In particular, let $\delta > 0$ and let us consider $\dot{S}$ [Eq. (4.2.31)] evaluated at the sequence of points $\{x_n^{\pm}\}_{n=1}^{\infty}$, where $x_n^{\pm} = ((q_0)_1, ..., (q_0)_l \pm \delta/n, ..., (q_0)_N, (p_0)_1, ..., (p_0)_1)$, for any l, such that the only deviation away from equilibrium is along the direction $\Delta q_l = \pm\delta/n$, with all other Δq_i and Δp_i vanishing. Then, for any n, we must have according to the above expression for $\dot{S}$ [Eq. (4.2.31)]:

$$\dot{S}\left(x_n^+\right) = a_l \frac{\delta}{n} + A_{ll} \frac{\delta^2}{n^2} + \mathcal{O}\left(\frac{\delta^3}{n^3}\right) > 0, \tag{4.2.37}$$

$$\dot{S}\left(x_n^-\right) = -a_l \frac{\delta}{n} + A_{ll} \frac{\delta^2}{n^2} + \mathcal{O}\left(\frac{\delta^3}{n^3}\right) > 0. \tag{4.2.38}$$

Taking the $n \to \infty$ limit of the first inequality implies $a_l \geq 0$, while doing the same for the second inequality implies $a_l \leq 0$. Hence $a_l = 0$. Since l is arbitrary, this means that $a_i = 0, \forall i$. A similar argument (using $\Delta p_l = \pm\delta/n$) implies $b_i = 0, \forall i$. Thus, S needs to satisfy the constraints

$$0 = \sum_{k=1}^{N} \left[\left(\frac{\partial H}{\partial p_k}\right)_0 \left(\frac{\partial^2 S}{\partial q_i \partial q_k}\right)_0 - \left(\frac{\partial H}{\partial q_k}\right)_0 \left(\frac{\partial^2 S}{\partial q_i \partial p_k}\right)_0 \right], \tag{4.2.39}$$

$$0 = \sum_{k=1}^{N} \left[\left(\frac{\partial H}{\partial p_k}\right)_0 \left(\frac{\partial^2 S}{\partial p_i \partial q_k}\right)_0 - \left(\frac{\partial H}{\partial q_k}\right)_0 \left(\frac{\partial^2 S}{\partial p_i \partial p_k}\right)_0 \right], \tag{4.2.40}$$

and this leaves us with

$$\dot{S} = \begin{bmatrix} \Delta q_1 & \cdots & \Delta p_N \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \Delta q_1 \\ \vdots \\ \Delta p_N \end{bmatrix} + \mathcal{O}\left(\Delta^3\right). \tag{4.2.41}$$

Now, imposing the Hamiltonian assumptions H1(a) and H2 [eqs. (4.2.7) and (4.2.9) respectively] simplifies $\mathbf{A}$ and $\mathbf{C}$, from the above [eqs. (4.2.34) and (4.2.36) respectively] to:

$$A_{ij} = -\frac{1}{2} \sum_{k=1}^{N} \left[\left(\frac{\partial^2 H}{\partial q_i \partial q_k}\right)_0 \left(\frac{\partial^2 S}{\partial q_j \partial p_k}\right)_0 + \left(\frac{\partial^2 H}{\partial q_j \partial q_k}\right)_0 \left(\frac{\partial^2 S}{\partial q_i \partial p_k}\right)_0 \right], \tag{4.2.42}$$

$$C_{ij} = \frac{1}{2} \left[\left(\frac{\partial^2 H}{\partial p_i^2}\right)_0 \left(\frac{\partial^2 S}{\partial p_j \partial q_i}\right)_0 + \left(\frac{\partial^2 H}{\partial p_j^2}\right)_0 \left(\frac{\partial^2 S}{\partial p_i \partial q_j}\right)_0 \right]. \tag{4.2.43}$$

Positive-definiteness of $(\mathbf{Hess}(\dot{S}))_0$ implies that the quadratic form above [Eq. (4.2.41)] should be positive definite. This means that we cannot have $(\partial^2 H/\partial p_j^2)_0 = 0, \forall j$, since then $\mathbf{C}$ would not be positive definite and we would get a contradiction. This, combined with assumption H1(b) [Eq. (4.2.8)], implies that $(\partial^2 H/\partial p_j^2)_0 > 0, \forall j$. Moreover, we also have:

$$\sum_{i,j=1}^{N} C_{ij} = \sum_{i,j=1}^{N} \left(\frac{\partial^2 H}{\partial p_j^2}\right)_0 \left(\frac{\partial^2 S}{\partial p_i \partial q_j}\right)_0 > 0. \tag{4.2.44}$$

The reason for this is easily seen by noting that positive-definiteness of $\mathbf{C}$, by definition, means that its product with any nonzero vector and its transpose should be positive, *i.e.* $\mathbf{z}^{\mathrm{T}}\mathbf{C}\mathbf{z} > 0$ for any nonzero vector $\mathbf{z}$; in particular, $\mathbf{z} = (1, 1, ..., 1)^{\mathrm{T}}$ achieves the above inequality [Eq. (4.2.44)]. But then, let us consider $\sum_{i,j=1}^{N} A_{ij}$. Positive-definiteness of $(\mathbf{Hess}(\dot{S}))_0$ (*i.e.* of the quadratic form [Eq. (4.2.31)]) implies, just as in the case of $\mathbf{C}$, that $\sum_{i,j=1}^{N} A_{ij} > 0$, or

$$\sum_{i,j=1}^{N} (-A_{ij}) < 0. \tag{4.2.45}$$

At the same time, we have:

$$\sum_{i,j=1}^{N}(-A_{ij}) = \sum_{i,j,k=1}^{N}\left(\frac{\partial^2 H}{\partial q_i \partial q_k}\right)_0 \left(\frac{\partial^2 S}{\partial q_j \partial p_k}\right)_0 . \tag{4.2.46}$$

Taking the minimum over the k index in the term with the H partials,

$$\sum_{i,j=1}^{N}(-A_{ij}) \geq \sum_{i,j,k=1}^{N}\left[\min_{1\leq l\leq N}\left(\frac{\partial^2 H}{\partial q_i \partial q_l}\right)_0\right]\left(\frac{\partial^2 S}{\partial q_j \partial p_k}\right)_0 , \tag{4.2.47}$$

This means that the sums can be separated, and after relabelling, the above [Eq. (4.2.47)] becomes:

$$\sum_{i,j=1}^{N}(-A_{ij}) \geq \left[\min_{1\leq l\leq N}\sum_{k=1}^{N}\left(\frac{\partial^2 H}{\partial q_k \partial q_l}\right)_0\right]\sum_{i,j=1}^{N}\left(\frac{\partial^2 S}{\partial p_i \partial q_j}\right)_0 . \tag{4.2.48}$$

Now, insert the identity $1 = (\partial^2 H/\partial p_j^2)_0/(\partial^2 H/\partial p_j^2)_0$ into the i, j sum, and maximize over the denominator to get:

$$\sum_{i,j=1}^{N}(-A_{ij}) \geq \left[\min_{1\leq l\leq N}\sum_{k=1}^{N}\left(\frac{\partial^2 H}{\partial q_k \partial q_l}\right)_0\right]\sum_{i,j=1}^{N}\frac{\left(\partial^2 H/\partial p_j^2\right)_0}{\left(\partial^2 H/\partial p_j^2\right)_0}\left(\frac{\partial^2 S}{\partial p_i \partial q_j}\right)_0 \tag{4.2.49}$$

$$\geq \left[\min_{1\leq l\leq N}\sum_{k=1}^{N}\left(\frac{\partial^2 H}{\partial q_k \partial q_l}\right)_0\right]\sum_{i,j=1}^{N}\left[\max_{1\leq m\leq N}\left(\frac{\partial^2 H}{\partial p_m^2}\right)_0\right]^{-1}\left(\frac{\partial^2 H}{\partial p_j^2}\right)_0\left(\frac{\partial^2 S}{\partial p_i \partial q_j}\right)_0 \tag{4.2.50}$$

$$= \left\{\left[\min_{1\leq l\leq N}\sum_{k=1}^{N}\left(\frac{\partial^2 H}{\partial q_k \partial q_l}\right)_0\right]\left[\max_{1\leq m\leq N}\left(\frac{\partial^2 H}{\partial p_m^2}\right)_0\right]^{-1}\right\}\sum_{i,j=1}^{N} C_{ij} \tag{4.2.51}$$

$$\geq 0, \tag{4.2.52}$$

since the term in curly brackets is non-negative (because of assumption H3 on the Hamiltonian), and we had earlier $\sum_{i,j=1}^{N} C_{ij} > 0$. But we also had $\sum_{i,j=1}^{N}(-A_{ij}) < 0$. Hence we get a contradiction. Therefore, no such function S exists. This concludes our proof.

4.2.3. Topological approach. We now turn to the topological approach to answering Problem I in CM. First we review the basic ideas of Olsen's line of argumentation [Olsen 1993], then we discuss their connections with the periodicity of phase space orbits.

4.2.3.1. *Review of Olsen's proof.* The assumptions made on $S : \mathscr{P} \to \mathbb{R}$ are in this case not as strict as in the perturbative approach. See Figure 4.3 for a pictorial representation.

In effect, we simply need to assume that S is nondecreasing along trajectories, which are confined to an invariant closed space $\mathscr{P}$. Under these conditions, Olsen furnishes two proofs [Olsen 1993] for why S is necessarily a constant. In the first one, the essential idea

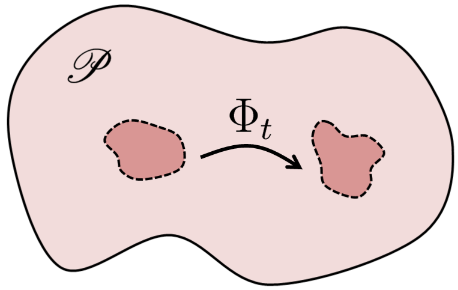

FIGURE 4.3. The topological approach relies on phase space compactness and Liouville's theorem, *i.e.* the fact that the Hamiltonian flow is volume-preserving.

is that the volume integral of S in $\mathscr{P}$ can be written after a change of variables as

$$\int_{\mathscr{P}} \boldsymbol{\Omega}\, S = \int_{\mathscr{P}} \boldsymbol{\Omega}\, (S \circ \Phi_t)\,, \tag{4.2.53}$$

owing to the fact that $\mathscr{P}$ is left invariant by the Hamiltonian flow Φ_t generated by the Hamiltonian vector field $\boldsymbol{X}_H$, and that $\mathcal{L}_{\boldsymbol{X}_H}\boldsymbol{\Omega} = 0$. Because the above expression [Eq. (4.2.53)] is time-independent, S must be time-independent, hence constant along trajectories. The second proof (based on the same assumptions) is rather more technical, but relies also basically on topological ideas; in fact, it is more related to the Poincaré recurrence property [Luis Barreira 2006].

We can make a few remarks. Firstly, there is in this case no requirement on the specific form of the Hamiltonian function $H : \mathscr{P} \to \mathbb{R}$. In fact, H can even contain explicit dependence on time and the proof still holds.

Secondly, the essential ingredient here is the compactness of the phase space $\mathscr{P}$. Indeed, even in Poincaré's original recurrence theorem [Poincaré 1890], as we saw in Section 4.2, the only necessary assumptions were also phase space compactness and invariance along with Liouville's theorem.

4.2.3.2. *Periodicity in phase space.* Even more can be said about the connection between phase space compactness and the recurrence of orbits than the Poincaré recurrence theorem. There are recent theorems in symplectic geometry which show that exact periodicity of orbits can exist in compact phase spaces.

For example, let us assume the Hamiltonian is of typical form [Eq. (4.2.11)]. Then, there is a theorem [Hofer and Zehnder 2011] which states that for a compact configuration space $\mathscr{Q}$, we have periodic solutions of $\boldsymbol{X}_H$. In fact, it was even shown [Suhr and Zehmisch 2016] that we have periodic solutions provided certain conditions on the potential V are satisfied and $\mathscr{Q}$ just needs to have bounded geometry (*i.e.* to be geodesically complete and to have the scalar curvature and derivative thereof bounded).

Thus, under the assumption of compactness or any other condition which entails closed orbits, we cannot have a function which behaves like entropy in this sense for a very simple reason. Assume $S : \mathscr{P} \to \mathbb{R}$ is nondecreasing along trajectories and let us consider an orbit $\gamma : \mathbb{R} \to \mathscr{P}$ in phase space (satisfying $\mathrm{d}\gamma\left(t\right)/\mathrm{d}t = \boldsymbol{X}_H(\gamma\left(t\right))$) which is closed. This means that for any $x \in \mathscr{P}$ on the orbit, there exist $t_0, T \in \mathbb{R}$ such that $x = \gamma(t_0) = \gamma(t_0 + T)$. Hence, we have $S(x) = S(\gamma(t_0)) = S(\gamma(t_0 + T)) = S(x)$, so S is constant along the orbit and therefore cannot behave like entropy.

4.3. Entropy theorems in general relativity

We now turn to addressing the question of why these theorems do not carry over from CM to GR. We follow the notation and general setup presented in Chapter 2. The only notable exception is that we write the general (possibly multi-) index A on the configuration variables φ as a subscript (φ_A) instead of a superscript (φ^A, as before), and we indicate summation over such indices explicitly. Moreover, we write integrals over Cauchy surfaces with respect to the flat volume form, $\mathbf{e} = \mathrm{d}^3x = \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3$.

4.3.1. Perturbative approach. We wish to investigate under what conditions the CM no-entropy proof of subsection 4.2.2 transfers over to field theories in curved spacetime. To this effect, we consider the equivalent setup: broadly speaking, we ask whether there exists a phase space functional $S : \mathscr{P} \to \mathbb{R}$ which is increasing in time everywhere except at an "equilibrium" configuration. In particular, we use the following two entropy conditions in analogy with those of subsubsection 4.2.2.3 in CM:

S1 *(Existence of equilibrium)***:** We assume there exists a point $x_0 = (\mathring{\varphi}, \mathring{\pi}) \in \mathscr{P}$, where S is stationary, and (to simplify the analysis) H is stationary as well:

$$\frac{\delta S\left[\mathring{\varphi},\mathring{\pi}\right]}{\delta\mathring{\varphi}_A\left(x\right)} = \frac{\delta S\left[\mathring{\varphi},\mathring{\pi}\right]}{\delta\mathring{\pi}_A\left(x\right)} = 0 = \frac{\delta H\left[\mathring{\varphi},\mathring{\pi}\right]}{\delta\mathring{\varphi}_A\left(x\right)} = \frac{\delta H\left[\mathring{\varphi},\mathring{\pi}\right]}{\delta\mathring{\pi}_A\left(x\right)}. \tag{4.3.1}$$

This implies $\dot{S}[\mathring{\varphi},\mathring{\pi}] = 0 = \dot{H}[\mathring{\varphi},\mathring{\pi}]$.

S2 *(Second law of thermodynamics)***:** We assume that the Hessian of $\dot{S}$ is positive definite at equilibrium, *i.e.* $\mathbf{Hess}(\dot{S}[\mathring{\varphi},\mathring{\pi}]) \succ 0$. This is a sufficient condition to ensure that $\dot{S} > 0$ in $\mathscr{P}\backslash x_0$, and $\dot{S} = 0$ at x_0.

We then follow the same procedure as in subsubsection 4.2.2.4: we insert into the Poisson bracket

$$\dot{S} = \int_\Sigma \mathrm{d}^3x \sum_A \left(\frac{\delta H[\varphi,\pi]}{\delta \pi_A(x)} \frac{\delta S[\varphi,\pi]}{\delta \varphi_A(x)} - \frac{\delta H[\varphi,\pi]}{\delta \varphi_A(x)} \frac{\delta S[\varphi,\pi]}{\delta \pi_A(x)} \right) \tag{4.3.2}$$

the functional Taylor series [Dreizler and Engel 2011] for each term about $(\mathring{\varphi}, \mathring{\pi})$, denoting $\Delta\varphi_A(x) = \varphi_A(x) - \mathring{\varphi}(x)$ and $\Delta\pi_A(x) = \pi_A(x) - \mathring{\pi}(x)$:

$$\begin{aligned} \frac{\delta H[\varphi,\pi]}{\delta \pi_A(x)} = \frac{\delta H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_A(x)} + \int_\Sigma \mathrm{d}^3y \sum_B \Bigg\{ & \frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\varphi}_B(y)\, \delta \mathring{\pi}_A(x)} \Delta\mathring{\varphi}_B(y) \\ & + \frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_B(y)\, \delta \mathring{\pi}_A(x)} \Delta\mathring{\pi}_B(y) \Bigg\} + \mathcal{O}\left(\Delta^2\right), \end{aligned} \tag{4.3.3}$$

and similarly for the other terms. Then we apply S1 in this case [Eq. (4.3.1)], which makes all zero-order terms vanish. Finally, the Poisson bracket in this case [Eq. (4.3.2)] becomes:

$$\begin{aligned} \dot{S} = \int_\Sigma \mathrm{d}^3x \int_\Sigma \mathrm{d}^3y \int_\Sigma \mathrm{d}^3z \sum_{A,B,C} \\ \Bigg\{ & \left[\frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\varphi}_B(y)\, \delta \mathring{\pi}_A(x)} \frac{\delta^2 S[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\varphi}_C(z)\, \delta \mathring{\varphi}_A(x)} - \frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\varphi}_B(y)\, \delta \mathring{\varphi}_A(x)} \frac{\delta^2 S[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\varphi}_C(z)\, \delta \mathring{\pi}_A(x)} \right] \\ & \times \Delta\varphi_B(y)\, \Delta\varphi_C(z) \\ + & \left[\frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\varphi}_B(y)\, \delta \mathring{\pi}_A(x)} \frac{\delta^2 S[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_C(z)\, \delta \mathring{\varphi}_A(x)} + \frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_C(z)\, \delta \mathring{\pi}_A(x)} \frac{\delta^2 S[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\varphi}_B(y)\, \delta \mathring{\varphi}_A(x)} \right. \\ & \left. - \frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\varphi}_B(y)\, \delta \mathring{\varphi}_A(x)} \frac{\delta^2 S[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_C(z)\, \delta \mathring{\pi}_A(x)} - \frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_C(z)\, \delta \mathring{\varphi}_A(x)} \frac{\delta^2 S[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\varphi}_B(y)\, \delta \mathring{\pi}_A(x)} \right] \\ & \times \Delta\varphi_B(y)\, \Delta\pi_C(z) \\ + & \left[\frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_B(y)\, \delta \mathring{\pi}_A(x)} \frac{\delta^2 S[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_C(z)\, \delta \mathring{\varphi}_A(x)} - \frac{\delta^2 H[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_B(y)\, \delta \mathring{\varphi}_A(x)} \frac{\delta^2 S[\mathring{\varphi},\mathring{\pi}]}{\delta \mathring{\pi}_C(z)\, \delta \mathring{\pi}_A(x)} \right] \\ & \times \Delta\pi_B(y)\, \Delta\pi_C(z) \Bigg\} + \mathcal{O}\left(\Delta^3\right). \end{aligned} \tag{4.3.4}$$

We compute this, in turn, for a scalar field in curved spacetime, for EM in curved spacetime, and for GR. We will show that no function S obeying the conditions S1-S2 given here exists in the case of the first two, but that the same cannot be said of the latter.

4.3.1.1. *Scalar field.* Let us consider a theory for a scalar field $\phi\left(x\right)$ in a potential $V\left[\phi\left(x\right)\right]$, defined by the Lagrangian

$$\mathcal{L}=\sqrt{-g}\left(-\frac{1}{2}g^{ab}\nabla_{a}\phi\nabla_{b}\phi-V\left[\phi\right]\right). \tag{4.3.5}$$

There are no constraints in this case. For turning the above [Eq. (4.3.5)] into a canonical theory, let us choose a foliation of $\mathscr{M}$ such that $\boldsymbol{N}=0$. The canonical measure [Crnković and Witten 1989] is then simply given by $\boldsymbol{\Omega}=\int_{\Sigma}\mathrm{d}^{3}x\,\delta\dot{\phi}\wedge\delta\phi$, and the Hamiltonian [Poisson 2007] is

$$H\left[\phi,\pi\right]=\int_{\Sigma}\mathrm{d}^{3}x\,N\left(\frac{\pi^{2}}{2\sqrt{h}}+\frac{\sqrt{h}}{2}h^{ab}\nabla_{a}\phi\nabla_{b}\phi+\sqrt{h}V\left[\phi\right]\right), \tag{4.3.6}$$

where $\pi=(\sqrt{h}/N)\dot{\phi}$ is the canonical momentum.

Let us compute the second functional derivatives of H. We have:

$$\begin{aligned}\frac{\delta^{2}H\left[\phi,\pi\right]}{\delta\phi\left(y\right)\delta\phi\left(x\right)}=&N\left(x\right)\sqrt{h\left(x\right)}V''\left[\phi\left(x\right)\right]\delta\left(x-y\right)\\&-\partial_{a}\left(N\left(x\right)\sqrt{h\left(x\right)}h^{ab}\left(x\right)\partial_{b}\delta\left(x-y\right)\right), \qquad (4.3.7)\\\frac{\delta^{2}H\left[\phi,\pi\right]}{\delta\pi\left(y\right)\delta\pi\left(x\right)}=&\frac{N\left(x\right)}{\sqrt{h\left(x\right)}}\delta\left(x-y\right), \qquad (4.3.8)\end{aligned}$$

and the mixed derivatives $\delta^{2}H\left[\phi,\pi\right]/\delta\pi\left(y\right)\delta\phi\left(x\right)$ vanish.

We now proceed as outlined above: We assume there exists an entropy function $S:\mathscr{P}\to\mathbb{R}$ obeying S1-S2 with an equilibrium field configuration $(\mathring{\phi},\mathring{\pi})$, and we will show that there is a contradiction with $\dot{S}>0$. Additionally, we assume that $V''[\mathring{\phi}]\geq0$; in other words, the equilibrium field configuration is one where the potential is concave upwards, *i.e.* it is a stable equilibrium.

According to the above expression for $\dot{S}$ [Eq. (4.3.4)], we have that entropy production in this case is given by

$$\begin{aligned}\dot{S}=\int_{\Sigma}\mathrm{d}^{3}x\,\mathrm{d}^{3}y\,\mathrm{d}^{3}z\Bigg\{&\left[-\frac{\delta^{2}H[\mathring{\phi},\mathring{\pi}]}{\delta\mathring{\phi}\left(y\right)\delta\mathring{\phi}\left(x\right)}\frac{\delta^{2}S[\mathring{\phi},\mathring{\pi}]}{\delta\mathring{\phi}\left(z\right)\delta\mathring{\pi}\left(x\right)}\right]\Delta\phi\left(y\right)\Delta\phi\left(z\right)\\&+\left[\frac{\delta^{2}H[\mathring{\phi},\mathring{\pi}]}{\delta\mathring{\pi}\left(z\right)\delta\mathring{\pi}\left(x\right)}\frac{\delta^{2}S[\mathring{\phi},\mathring{\pi}]}{\delta\mathring{\phi}\left(y\right)\delta\mathring{\phi}\left(x\right)}-\frac{\delta^{2}H[\mathring{\phi},\mathring{\pi}]}{\delta\mathring{\phi}\left(y\right)\delta\mathring{\phi}\left(x\right)}\frac{\delta^{2}S[\mathring{\phi},\mathring{\pi}]}{\delta\mathring{\pi}\left(z\right)\delta\mathring{\pi}\left(x\right)}\right]\Delta\phi\left(y\right)\Delta\pi\left(z\right)\\&+\left[\frac{\delta^{2}H[\mathring{\phi},\mathring{\pi}]}{\delta\mathring{\pi}\left(y\right)\delta\mathring{\pi}\left(x\right)}\frac{\delta^{2}S[\mathring{\phi},\mathring{\pi}]}{\delta\mathring{\pi}\left(z\right)\delta\mathring{\phi}\left(x\right)}\right]\Delta\pi\left(y\right)\Delta\pi\left(z\right)\Bigg\}+\mathcal{O}\left(\Delta^{3}\right),\end{aligned} \tag{4.3.9}$$

where we have used the fact that the mixed derivatives vanish. Let us now evaluate $\dot{S}$ along different directions in $\mathscr{P}$ away from $(\mathring{\phi},\mathring{\pi})$. Suppose $\Delta\pi$ is nonzero everywhere on Σ, and $\Delta\phi$ vanishes everywhere on Σ. Then, using the second momentum derivative of

H [Eq. (4.3.8)], $\dot{S}$ [Eq. (4.3.9)] becomes:

$$\dot{S} = \int_\Sigma \mathrm{d}^3x \, \mathrm{d}^3y \, \mathrm{d}^3z \, \frac{N(x)}{\sqrt{h(x)}} \delta(x-y) \times \frac{\delta^2 S[\mathring{\phi}, \mathring{\pi}]}{\delta\mathring{\pi}(z)\, \delta\mathring{\phi}(x)} \Delta\pi(y) \, \Delta\pi(z) + \mathcal{O}\left(\Delta^3\right) \tag{4.3.10}$$

$$= \int_\Sigma \mathrm{d}^3y \, \mathrm{d}^3z \, \frac{N(y)}{\sqrt{h(y)}} \frac{\delta^2 S[\mathring{\phi}, \mathring{\pi}]}{\delta\mathring{\pi}(z)\, \delta\mathring{\phi}(y)} \Delta\pi(y) \, \Delta\pi(z) + \mathcal{O}\left(\Delta^3\right) \tag{4.3.11}$$

$$\leq \left\{ \max_{x\in\Sigma} \frac{N(x)}{\sqrt{h(x)}} \left(\Delta\pi(x)\right)^2 \right\} \int_\Sigma \mathrm{d}^3y \, \mathrm{d}^3z \, \frac{\delta^2 S[\mathring{\phi}, \mathring{\pi}]}{\delta\mathring{\pi}(z)\, \delta\mathring{\phi}(y)} + \mathcal{O}\left(\Delta^3\right) . \tag{4.3.12}$$

The requirement that the LHS of the first line above [Eq. (4.3.10)] is strictly positive, combined with the strict positivity of the term in curly brackets in the third line [Eq. (4.3.12)] and the assumption (S2) of the definiteness of the Hessian of $\dot{S}$ at $(\mathring{\phi}, \mathring{\pi})$, altogether mean that the above [Eqs. (4.3.10)-(4.3.12)] imply:

$$\int_\Sigma \mathrm{d}^3y \, \mathrm{d}^3z \, \frac{\delta^2 S[\mathring{\phi}, \mathring{\pi}]}{\delta\mathring{\pi}(z)\, \delta\mathring{\phi}(y)} > 0. \tag{4.3.13}$$

Now let us evaluate $\dot{S}$ in a region of $\mathscr{P}$ where $\Delta\phi$ is nonzero everywhere on Σ, while $\Delta\pi$ vanishes everywhere on Σ. Then, using the second field derivative of H [Eq. (4.3.7)], the negative of the above expression for $\dot{S}$ [Eq. (4.3.9)] becomes:

$$\begin{aligned} -\dot{S} = \int_\Sigma \mathrm{d}^3x \, \mathrm{d}^3y \, \mathrm{d}^3z \, \Big\{ & N(x) \sqrt{h(x)} V''[\mathring{\phi}(x)] \delta(x-y) \\ & - \partial_a \left(N(x) \sqrt{h(x)} h^{ab}(x) \, \partial_b \delta(x-y) \right) \Big\} \\ & \times \frac{\delta^2 S[\mathring{\phi}, \mathring{\pi}]}{\delta\mathring{\phi}(z)\, \delta\mathring{\pi}(x)} \Delta\phi(y) \, \Delta\phi(z) + \mathcal{O}\left(\Delta^3\right) . \end{aligned} \tag{4.3.14}$$

Now, observe that

$$\begin{aligned} \int_\Sigma \mathrm{d}^3x \, \mathrm{d}^3y \, \mathrm{d}^3z \, & \left\{ \partial_a \left(N(x) \sqrt{h(x)} h^{ab}(x) \, \partial_b \delta(x-y) \right) \right\} \\ & \times \frac{\delta^2 S[\mathring{\phi}, \mathring{\pi}]}{\delta\mathring{\phi}(z)\, \delta\mathring{\pi}(x)} \Delta\phi(y) \, \Delta\phi(z) \end{aligned} \tag{4.3.15}$$

is simply a boundary term. This can be seen by integrating by parts until the derivative is removed from the delta distribution, the definition of the latter is applied to remove the x integration, and the result is a total derivative in the integrand. Assuming asymptotic decay properties sufficient to make this boundary term vanish, the above $-\dot{S}$ [Eq. (4.3.14)] simply becomes:

$$-\dot{S} = \int_{\Sigma} \mathrm{d}^3y\, \mathrm{d}^3z\, N(y)\, \sqrt{h(y)} V''[\mathring{\phi}(y)] \frac{\delta^2 S[\mathring{\phi}, \mathring{\pi}]}{\delta\mathring{\phi}(z)\, \delta\mathring{\pi}(y)} \Delta\phi(y)\, \Delta\phi(z) + \mathcal{O}\left(\Delta^3\right) \tag{4.3.16}$$

$$\geq \left\{\min_{x\in\Sigma} N(x)\, \sqrt{h(x)} V''[\mathring{\phi}(x)]\, (\Delta\phi(x))^2\right\} \int_{\Sigma} \mathrm{d}^3y\, \mathrm{d}^3z\, \frac{\delta^2 S[\mathring{\phi}, \mathring{\pi}]}{\delta\mathring{\phi}(z)\, \delta\mathring{\pi}(y)} + \mathcal{O}\left(\Delta^3\right). \tag{4.3.17}$$

The LHS of the first line [Eq. (4.3.16)] should be strictly negative, and the term in curly brackets in the second line [Eq. (4.3.17)] is strictly positive. Hence, owing to the definiteness of the Hessian of $\dot{S}$ at $(\mathring{\phi}, \mathring{\pi})$, and using the symmetry of the arguments in the integrand and equality of mixed derivatives, the above [Eqs. (4.3.16)-(4.3.17)] imply:

$$\int_{\Sigma} \mathrm{d}^3y\, \mathrm{d}^3z\, \frac{\delta^2 S[\mathring{\phi}, \mathring{\pi}]}{\delta\mathring{\pi}(z)\, \delta\mathring{\phi}(y)} < 0. \tag{4.3.18}$$

This is a contradiction with the inequality obtained previously [Eq. (4.3.13)]. Therefore, we have no function S for a scalar field theory that behaves like entropy according to assumptions S1-S2.

We remark that in this case, we get the conclusion $\dot{S} = 0$ using the perturbative approach despite the fact that the topological one would not work in the case of a non-compact Cauchy surface. The reason is that:

$$\mu(\mathscr{P}) = \int_{\mathscr{P}} \boldsymbol{\Omega} = \int_{\mathscr{P}} \int_{\Sigma} \mathrm{d}^3x\, \delta\dot{\phi}(x) \wedge \delta\phi(x) \tag{4.3.19}$$

$$\geq \int_{\mathscr{P}} \int_{\Sigma} \mathrm{d}^3x\, \min_{y\in\Sigma} \left[\delta\dot{\phi}(y) \wedge \delta\phi(y)\right] \tag{4.3.20}$$

$$= \int_{\mathscr{P}} \left\{\min_{y\in\Sigma} \left[\delta\dot{\phi}(y) \wedge \delta\phi(y)\right]\right\} \left[\int_{\Sigma} \mathrm{d}^3x\right], \tag{4.3.21}$$

which diverges if Σ is non-compact. (N.B. The reason why the term in curly brackets is finite but non-zero is that the field and its time derivative cannot be always vanishing at any given point, for if they were it would lead only to the trivial solution.) Thus, only the perturbative approach is useful here for deducing lack of entropy production for spacetimes with non-compact Cauchy surfaces.

4.3.1.2. *Electromagnetism.* Before we inspect EM in curved spacetime, let us carry out the analysis in flat spacetime ($N = 1$, $\boldsymbol{N} = \boldsymbol{0}$, and $\boldsymbol{h} = {}^{(3)}\boldsymbol{\delta} = \mathrm{diag}(0, 1, 1, 1)$), for massive (or de Broglie-Proca) EM [Prescod-Weinstein and Bertschinger 2014], defined by the Lagrangian

$$\mathcal{L} = -\frac{1}{4}\boldsymbol{F} : \boldsymbol{F} - \frac{1}{2}m^2 \boldsymbol{A} \cdot \boldsymbol{A} + \boldsymbol{A} \cdot \boldsymbol{J}, \tag{4.3.22}$$

where $F_{ab} = \partial_a A_b - \partial_b A_a$, A_a is the electromagnetic potential (Faraday tensor), and J_a is an external source.

We have a constrained Hamiltonian system in this case. In particular, the momentum canonically conjugate to $A_0 = V$ vanishes identically. This means that instead of A_a, we may take (its spatial part) $\mathcal{A}_a = {}^{(3)}\delta_{ab} A^b$ along with its conjugate momentum, $\pi^a = \dot{\mathcal{A}}^a - \partial^a V$, to be the phase space variables—while appending to the canonical equations of motion resulting from $H[\boldsymbol{\mathcal{A}}, \boldsymbol{\pi}]$ the constraint $0 = \delta H / \delta V$. In particular, we have [Prescod-Weinstein and Bertschinger 2014]:

$$H[\boldsymbol{\mathcal{A}}, \boldsymbol{\pi}] = \int_\Sigma \mathrm{d}^3 x \bigg(\frac{1}{4} \boldsymbol{\mathcal{F}} : \boldsymbol{\mathcal{F}} + \frac{m^2}{2} \left(\boldsymbol{\mathcal{A}} \cdot \boldsymbol{\mathcal{A}} - V^2 \right) - \boldsymbol{\mathcal{A}} \cdot \boldsymbol{\mathcal{J}} + \frac{1}{2} \boldsymbol{\pi} \cdot \boldsymbol{\pi} - \left(\partial_a \pi^a + \rho \right) V + \partial_a \left(V \pi^a \right) \bigg), \tag{4.3.23}$$

where $\mathcal{F}_{ab} = {}^{(3)}\delta_{ac} {}^{(3)}\delta_{bd} F^{cd}$, $\rho = J^0$ and $\mathcal{J}^a = {}^{(3)}\delta^{ab} J_b$.

The Poisson bracket [Eq. (4.3.4)] is, in this case:

$$\begin{aligned} \dot{S} = \int_\Sigma \mathrm{d}^3 x \, \mathrm{d}^3 y \, \mathrm{d}^3 z \Bigg\{ & \left[- \frac{\delta^2 H[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\mathcal{A}}_b(y) \, \delta \mathring{\mathcal{A}}_a(x)} \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\mathcal{A}}_c(z) \, \delta \mathring{\pi}^a(x)} \right] \Delta \mathcal{A}_b(y) \, \Delta \mathcal{A}_c(z) \\ & + \left[\frac{\delta^2 H[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\pi}_c(z) \, \delta \mathring{\pi}_a(x)} \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\mathcal{A}}_b(y) \, \delta \mathring{\mathcal{A}}^a(x)} \right. \\ & \left. - \frac{\delta^2 H[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\mathcal{A}}_b(y) \, \delta \mathring{\mathcal{A}}_a(x)} \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\pi}_c(z) \, \delta \mathring{\pi}^a(x)} \right] \Delta \mathcal{A}_b(y) \, \Delta \pi_c(z) \\ & + \left[\frac{\delta^2 H[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\pi}_b(y) \, \delta \mathring{\pi}_a(x)} \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\pi}_c(z) \, \delta \mathring{\mathcal{A}}^a(x)} \right] \Delta \pi_b(y) \, \Delta \pi_c(z) \Bigg\} + \mathcal{O}\left(\Delta^3 \right), \end{aligned} \tag{4.3.24}$$

where we have used the fact that the mixed derivatives of the Hamiltonian [Eq. (4.3.23)] vanish by inspection, and we compute the second field and momentum derivatives thereof to be, respectively:

$$\frac{\delta^2 H[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\mathcal{A}}_b(y) \, \delta \mathring{\mathcal{A}}_a(x)} = - \left\{ {}^{(3)}\delta^{ab} \partial^c \partial_c \delta(x - y) - \partial^b \partial^a \delta(x - y) \right\} + m^2 \left[{}^{(3)}\delta^{ab} \delta(x - y) \right], \tag{4.3.25}$$

$$\frac{\delta^2 H[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\pi}_b(y) \, \delta \mathring{\pi}_a(x)} = {}^{(3)}\delta^{ab} \delta(x - y) \, . \tag{4.3.26}$$

Analogously with our strategy in the scalar field case, let us evaluate $\dot{S}$ along different directions away from equilibrium. In particular, let us suppose $\Delta \pi_1$ is nonzero everywhere on Σ, and that $\Delta \pi_2$, $\Delta \pi_3$, and $\Delta \mathcal{A}_a$ all vanish everywhere on Σ. Then, using the second momentum derivative of H [Eq. (4.3.26)], $\dot{S}$ [Eq. (4.3.24)] becomes:

$$\dot{S} = \int_{\Sigma} \mathrm{d}^3x\, \mathrm{d}^3y\, \mathrm{d}^3z\, \delta\,(x-y)\, \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_1\,(z)\, \delta\mathring{\mathcal{A}}_1(x)} \Delta\pi_1(y)\, \Delta\pi_1\,(z) + \mathcal{O}\left(\Delta^3\right) \tag{4.3.27}$$

$$= \int_{\Sigma} \mathrm{d}^3y\, \mathrm{d}^3z\, \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_1\,(z)\, \delta\mathring{\mathcal{A}}_1\,(y)} \Delta\pi_1\,(y)\, \Delta\pi_1\,(z) + \mathcal{O}\left(\Delta^3\right) \tag{4.3.28}$$

$$\leq \left\{ \max_{x\in\Sigma} \left(\Delta\pi_1\,(x)\right)^2 \right\} \int_{\Sigma} \mathrm{d}^3y\, \mathrm{d}^3z\, \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_1\,(z)\, \delta\mathring{\mathcal{A}}_1\,(y)} + \mathcal{O}\left(\Delta^3\right). \tag{4.3.29}$$

The argument proceeds as before: the strict positivity of the LHS of the first line above [Eq. (4.3.27)], combined with that of the term in curly brackets in the third line [Eq. (4.3.29)] and the assumption (S2) of the definiteness of the Hessian of $\dot{S}$ at $(\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}})$, altogether mean that the above [Eqs. (4.3.27)-(4.3.29)] imply:

$$\int_{\Sigma} \mathrm{d}^3y\, \mathrm{d}^3z\, \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_1\,(z)\, \delta\mathring{\mathcal{A}}_1\,(y)} > 0. \tag{4.3.30}$$

Now let us evaluate $\dot{S}$ where $\Delta\mathcal{A}_1$ is nonzero everywhere on Σ, while $\Delta\mathcal{A}_2$, $\Delta\mathcal{A}_3$ and $\Delta\pi_a$ all vanish everywhere on Σ. Then, using the second field derivative of H [Eq. (4.3.25)], the negative of the above expression for $\dot{S}$ [Eq. (4.3.24)] becomes:

$$\begin{aligned} -\dot{S} = \int_{\Sigma} \mathrm{d}^3x\, \mathrm{d}^3y\, \mathrm{d}^3z \Bigg[\Bigg(- \Big\{ {}^{(3)}\delta^{ab} \partial^c \partial_c \delta\,(x-y) \\ - \partial^b \partial^a \delta\,(x-y) \Big\} + m^2 \left[{}^{(3)}\delta^{ab} \delta\,(x-y) \right] \Bigg) \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta\mathring{\mathcal{A}}_c\,(z)\, \delta\mathring{\pi}^a\,(x)} \Bigg] \\ \times \Delta\mathcal{A}_b\,(y)\, \Delta\mathcal{A}_c\,(z) + \mathcal{O}\left(\Delta^3\right). \end{aligned} \tag{4.3.31}$$

The term in curly brackets simply furnishes a (vanishing) boundary term (up to $\mathcal{O}(\Delta^3)$). Note that for $m=0$ (corresponding to Maxwellian EM in flat spacetime) we would thus get an indefinite Hessian of $\dot{S}$ at $(\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}})$, and hence no function S that behaves like entropy as per S1-S2. So let us assume $m^2 > 0$. Using the symmetry of the arguments in the integrand and equality of mixed derivatives, we are thus left with:

$$-\dot{S} = \int_{\Sigma} \mathrm{d}^3y\, \mathrm{d}^3z\, m^2 \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_1\,(z)\, \delta\mathring{\mathcal{A}}_1\,(y)} \Delta\mathcal{A}_1\,(y)\, \Delta\mathcal{A}_1\,(z) + \mathcal{O}\left(\Delta^3\right) \tag{4.3.32}$$

$$\geq \left\{ m^2 \min_{x\in\Sigma} \left(\Delta\mathcal{A}_1\,(x)\right)^2 \right\} \int_{\Sigma} \mathrm{d}^3y\, \mathrm{d}^3z\, \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_1\,(z)\, \delta\mathring{\mathcal{A}}_1\,(y)} + \mathcal{O}\left(\Delta^3\right). \tag{4.3.33}$$

The LHS of the first line [Eq. (4.3.32)] should be strictly negative, and the term in curly brackets in the second line [Eq. (4.3.33)] is strictly positive. Hence, owing to the definiteness of the Hessian of $\dot{S}$ at $(\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}})$, the above [Eqs. (4.3.32)-(4.3.33)] imply:

$$\int_\Sigma \mathrm{d}^3 y \, \mathrm{d}^3 z \, \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\pi}_1(z) \, \delta \mathring{\mathcal{A}}_1(y)} < 0. \tag{4.3.34}$$

This is a contradiction with the previous inequality on the same quantity [Eq. (4.3.30)]. Hence there is no function S that behaves like entropy (according to S1-S2) for a massive EM field in flat spacetime.

Let us now carry out the proof for a simple Maxwellian EM field in curved spacetime, defined by the Lagrangian

$$\mathcal{L} = -\frac{1}{4}\sqrt{-g}\boldsymbol{F} : \boldsymbol{F}, \tag{4.3.35}$$

where $F_{ab} = \nabla_a A_b - \nabla_b A_a$ and A_a is the electromagnetic potential. As in the scalar field case, we work with a spacetime foliation such that $\boldsymbol{N} = 0$.

As with EM in flat spacetime, this is a constrained Hamiltonian system: the momentum canonically conjugate to $A_0 = V$ vanishes identically, meaning again that instead of A_a, we may take (its spatial part) $\mathcal{A}_a = h_{ab}A^b$ along with its conjugate momentum, $\pi^a = (\sqrt{h}/N)h^{ab}(\dot{\mathcal{A}}_b - \partial_b V)$, to be the physical phase space variables—appending to the canonical equations of motion resulting from $H[\boldsymbol{\mathcal{A}}, \boldsymbol{\pi}]$ the constraint $0 = \delta H/\delta V = \partial_a \pi^a$ (which is simply Gauss' law). In particular, we have [Prescod-Weinstein and Bertschinger 2014]:

$$H[\boldsymbol{\mathcal{A}}, \boldsymbol{\pi}] = \int_\Sigma \mathrm{d}^3 x \left(\frac{1}{4} N\sqrt{h}\,\boldsymbol{\mathcal{F}} : \boldsymbol{\mathcal{F}} + \frac{N}{2\sqrt{h}} \boldsymbol{\pi}\cdot\boldsymbol{\pi} + \pi^a \partial_a V \right), \tag{4.3.36}$$

where $\mathcal{F}_{ab} = h_{ac}h_{bd}F^{cd} = \mathcal{D}_a\mathcal{A}_b - \mathcal{D}_b\mathcal{A}_a$, where $\boldsymbol{\mathcal{D}}$ is the derivative induced on Σ.

The Poisson bracket [Eq. (4.3.4)] is here given by the same expression as in flat spacetime [Eq. (4.3.9)], owing to the fact that the mixed derivatives of the Hamiltonian [Eq. (4.3.36)] vanish. Let us focus on regions in phase space where $\Delta\boldsymbol{\pi}$ vanishes everywhere on Σ, but $\Delta\boldsymbol{\mathcal{A}}$ is everywhere nonzero. There,

$$\dot{S} = \int_\Sigma \mathrm{d}^3 x \, \mathrm{d}^3 y \, \mathrm{d}^3 z \left[-\frac{\delta^2 H[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\mathcal{A}}_b(y) \, \delta \mathring{\mathcal{A}}_a(x)} \frac{\delta^2 S[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\mathcal{A}}_c(z) \, \delta \mathring{\pi}^a(x)} \right] \times \Delta\mathcal{A}_b(y)\,\Delta\mathcal{A}_c(z) + \mathcal{O}\left(\Delta^3\right). \tag{4.3.37}$$

We compute:

$$\frac{\delta^2 H[\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}}]}{\delta \mathring{\mathcal{A}}_b(y) \, \delta \mathring{\mathcal{A}}_a(x)} = -\sqrt{h(x)} \left\{ \mathcal{D}^c \left[N(x)\, h^{ab}(x)\, \mathcal{D}_c \delta(x-y) \right] - \mathcal{D}^b \left[N(x)\, \mathcal{D}^a \delta(x-y) \right] \right\}. \tag{4.3.38}$$

Inserting this into the above expression for $\dot{S}$ [Eq. (4.3.37)], we simply get a (vanishing) boundary term (up to $\mathcal{O}(\Delta^3)$). We conclude that we have an indefinite Hessian of $\dot{S}$ at $(\mathring{\boldsymbol{\mathcal{A}}}, \mathring{\boldsymbol{\pi}})$, and hence no function S that behaves like entropy as per S1-S2.

4.3.1.3. *Gravity.* Here we use the basic notation and phase space construction of Section 2.4. We briefly remind the reader that the Hamiltonian of general relativity is given by

$$H = \frac{1}{2\kappa}\int_\Sigma \boldsymbol{\epsilon}_\Sigma \left[NC - 2\boldsymbol{N}\cdot\boldsymbol{C} + 2\boldsymbol{\mathcal{D}}\cdot\left(\frac{\boldsymbol{N}\cdot\boldsymbol{\pi}}{\sqrt{h}}\right)\right], \tag{4.3.39}$$

where N and $\boldsymbol{N}$ are the lapse and shift, C and $\boldsymbol{C}$ are the secondary constraints, and $\boldsymbol{h}$ and $\boldsymbol{\pi}$ are the induced three-metric on Σ and its canonical momentum respectively. See Chapter 2 for full details.

Following the same procedure as before for a hypothetical entropy functional $S[\boldsymbol{h},\boldsymbol{\pi}]$ and an equilibrium configuration $(\mathring{\boldsymbol{h}}, \mathring{\boldsymbol{\pi}})$ in phase space, we see that the Poisson bracket [Eq. (4.3.4)] in this case has the following form:

$$\begin{aligned}
\dot{S} = \int_\Sigma & \mathrm{d}^3x\,\mathrm{d}^3y\,\mathrm{d}^3z \\
& \Bigg\{ \left[\frac{\delta^2 H[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{h}^{cd}(y)\,\delta\mathring{\pi}^{ab}(x)}\frac{\delta^2 S[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{h}_{ef}(z)\,\delta\mathring{h}_{ab}(x)} - \frac{\delta^2 H[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{h}^{cd}(y)\,\delta\mathring{h}^{ab}(x)}\frac{\delta^2 S[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{h}_{ef}(z)\,\delta\mathring{\pi}_{ab}(x)}\right] \\
&\qquad \times \Delta h^{cd}(y)\,\Delta h_{ef}(z) \\
&+ \left[\frac{\delta^2 H[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{h}^{cd}(y)\,\delta\mathring{\pi}^{ab}(x)}\frac{\delta^2 S[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_{ef}(z)\,\delta\mathring{h}_{ab}(x)} + \frac{\delta^2 H[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_{ef}(z)\,\delta\mathring{\pi}^{ab}(x)}\frac{\delta^2 S[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{h}^{cd}(y)\,\delta\mathring{h}_{ab}(x)}\right. \\
&\quad \left. - \frac{\delta^2 H[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{h}^{cd}(y)\,\delta\mathring{h}^{ab}(x)}\frac{\delta^2 S[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_{ef}(z)\,\delta\mathring{\pi}_{ab}(x)} - \frac{\delta^2 H[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_{ef}(z)\,\delta\mathring{h}^{ab}(x)}\frac{\delta^2 S[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{h}^{cd}(y)\,\delta\mathring{\pi}_{ab}(x)}\right] \\
&\qquad \times \Delta h^{cd}(y)\,\Delta \pi_{ef}(z) \\
&+ \left[\frac{\delta^2 H[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}^{cd}(y)\,\delta\mathring{\pi}^{ab}(x)}\frac{\delta^2 S[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_{ef}(z)\,\delta\mathring{h}_{ab}(x)} - \frac{\delta^2 H[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}^{cd}(y)\,\delta\mathring{h}^{ab}(x)}\frac{\delta^2 S[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_{ef}(z)\,\delta\mathring{\pi}_{ab}(x)}\right] \\
&\qquad \times \Delta \pi^{cd}(y)\,\Delta \pi_{ef}(z) \Bigg\} \\
&+\mathcal{O}\left(\Delta^3\right).
\end{aligned} \tag{4.3.40}$$

The difference with the previous cases is that here, in general, none of the second derivatives of the Hamiltonian vanish, and crucially, they do not have a definite sign. For example, let us compute the second derivative of H with respect to the canonical momentum:

$$\frac{\delta^2 H[\mathring{\boldsymbol{h}},\mathring{\boldsymbol{\pi}}]}{\delta\mathring{\pi}_{cd}(y)\,\delta\mathring{\pi}^{ab}(x)} = \frac{2\mathring{N}(x)}{\sqrt{\mathring{h}(x)}}\left(\delta^c{}_{(a}\delta^d{}_{b)} - \frac{1}{2}\mathring{h}_{ab}(x)\,\delta^{cd}\right)\delta(x-y)\,. \tag{4.3.41}$$

In CM or the examples of field theories in curved spacetime we have considered, the second derivative of H with respect to the momentum had a definite sign (by virtue of its association with the positivity of kinetic-type terms). In this case, however, this second derivative [Eq. (4.3.41)] is neither always positive nor always negative. Thus an argument similar to the previous proofs cannot work here: the gravitational Hamiltonian [Eq. (2.4.15)] is of such a nature that its concavity in phase space components (as is, for example, its concavity in the canonical momentum components [Eq. (4.3.41)]) is not independent of the phase space variables themselves, and cannot be ascribed a definite (positive or negative) sign. And so, a contradiction cannot arise with the Poisson bracket of a phase space functional (such as the gravitational entropy) being non-zero (and, in particular, positive).

4.3.2. Topological approach. As discussed in subsection 4.2.3, the topological proofs of Olsen for the non-existence of entropy production in CM rely crucially on the assumption that the phase space $\mathscr{P}$ is compact. In such a situation, a system has a finite measure of phase space $\mu(\mathscr{P})$ available to explore, and there cannot exist a function which continually increases along orbits.

By contrast, in GR, it is believed that the (reduced) phase space $\mathscr{S}$ is generically non-compact [Schiffrin and Wald 2012]. That is to say, the measure $\mu(\mathscr{S}) = \int_{\mathscr{S}} \boldsymbol{\Omega}|_{\mathscr{S}}$ in general diverges, where $\boldsymbol{\Omega}|_{\mathscr{S}}$ is (using the notation of Chapter 2) the pullback of the symplectic form of GR,

$$\boldsymbol{\Omega} = \int_{\Sigma} \mathrm{d}^3x\, \delta\pi^{ab} \wedge \delta h_{ab}\,, \tag{4.3.42}$$

to $\mathscr{S}$. (See Section 2.4 for more details.) This means that the same methods of proof as in CM (subsection 4.2.3) cannot be applied.

The connection between a (monotonically increasing) entropy function in GR and the divergence of its (reduced) phase space measure warrants some discussion. The latter, it may be noted, is arguably not completely inevitable. In other words, one may well imagine a space of admissible solutions to the Einstein equations (or equivalently, the canonical gravitational equations) the effective degrees of freedom of which are such that they form a finite-measure phase space. Dynamically-trivial examples of this might be SD black holes. Thus the assertion that $\mu(\mathscr{S})$ diverges hinges on the nature of the degrees of freedom believed to be available in the spacetimes under consideration. However, it has been explicitly shown [Schiffrin and Wald 2012] that even in very basic dynamically-nontrivial situations, such as simple cosmological spacetimes, $\mu(\mathscr{S})$ does indeed diverge. In fact, the proof found in [Schiffrin and Wald 2012] is carried out for compact Cauchy surfaces, and the conclusion is therefore in concordance with the no-return theorem [Tipler 1980, 1979] which also assumes compact Cauchy surfaces. In the following section, we will show that this happens for perturbed SD spacetimes as well (where the Cauchy surface is non-compact).

The generic divergence of $\mu(\mathscr{S})$ entails that a gravitational system has an unbounded region of phase space available to explore. In other words, it is not confined to a finite region where it would have to eventually return to a configuration from which it started (which would make a monotonically increasing entropy function impossible).

It is moreover worth remarking that this situation creates nontrivial problems for a statistical (*i.e.* probability-based) general-relativistic definition of entropy, $S(t)$ (as described in Section 4.1)—which, indeed, one may also ultimately desire to work with and relate to the mechanical meaning of entropy mainly discussed in this paper. Naively, one might think of defining such a statistical entropy function as something along the lines of $S = -\sum_X P(X) \ln P(X)$, where X denotes a physical property of interest and $P(X)$ its probability. In turn, the latter might be understood as the relative size of the phase space region $\mathscr{S}_X \subset \mathscr{S}$ possessing the property X, *i.e.* $P(X) = \mu(\mathscr{S}_X)/\mu(\mathscr{S})$. In this case, we either have [Schiffrin and Wald 2012]: $P(X) = 0$ if $\mu(\mathscr{S}_X)$ is finite, $P(X) = 1$ if $\mu(\mathscr{S}\backslash\mathscr{S}_X)$ is finite, or $P(X)$ is ill-defined otherwise. Ostensibly, one would need to invoke a regularization procedure in order to obtain finite probabilities (in general) according to this. However, different regularization procedures that have been applied (mainly in the context of cosmology) have proven to yield widely different results depending on the method of the procedure being used [Schiffrin and Wald 2012]. Alternatively, a statistical general-relativistic definition of S in terms of a probability density $\rho : \mathscr{S} \times \mathscr{T} \to [0,1]$ (similarly to CM) as $S = -\int_{\mathscr{S}} \mathbf{\Omega}|_{\mathscr{S}} \rho \ln \rho$ would likewise face divergence issues. Therefore, any future attempt to define gravitational entropy in such a context will have to either devise an unambiguous and well-defined regularization procedure (for obtaining finite probabilities), or implement a well-justified cutoff of the (reduced) phase space measure.

We now turn to discussing these issues in a context where we expect an intuitive illustration of gravitational entropy production—the two-body problem.

4.4. Entropy in the gravitational two-body problem

One of the most elementary situations in GR in which we expect the manifestation of a phenomenon such as entropy production is the gravitational two-body problem.

In CM, the two-body (or Kepler) problem manifestly involves no increase in the entropy of a system. The perturbative approach, as discussed in subsubsection 4.2.2.3, involves assumptions on the nature of the Hamiltonian which preclude any conclusions from it in this regard. However, the topological approach, elaborated in subsection 4.2.3, is applicable: assuming that Keplerian orbits are bounded, the configuration space $\mathscr{Q}$ can be considered to be compact, and therefore the phase space $\mathscr{P}$ obtained from it (involving finite conjugate momenta) is compact as well. Concordant with the topological proofs, then, we will have no entropy production in such a situation. The case of the N-body problem however is, as alluded to earlier, not the same: neither the assumptions of the perturbative approach, not of the topological approach (specifically, a compact phase space) are applicable, and it has been shown that a monotonically increasing function on phase space does in fact exist [Barbour, Koslowski, et al. 2013, 2014], and hence, a gravitational arrow of time (and entropy production) associated with it.

In GR, we know the two-body problem involves energy loss and therefore should implicate an associated production of entropy. The no-return theorem [Tipler 1980, 1979] is inapplicable here because this problem does not involve a compact Cauchy surface. The perturbative approach here fails to disprove the second law (as discussed in subsubsection 4.3.1.3), and we will now show that so too does the topological approach.

The two-body problem in GR where one small body orbits a much larger body of mass M can be modeled in the context of perturbations to the SD metric,

$$g_{ab}\mathrm{d}x^a\mathrm{d}x^b = -f\left(r\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{f\left(r\right)} + r^2\sigma_{ab}\mathrm{d}x^a\mathrm{d}x^b, \tag{4.4.1}$$

where $f\left(r\right) = 1 - 2M/r$ and $\boldsymbol{\sigma} = \mathrm{diag}(0, 0, 1, \sin^2\theta)$ is the metric of the two-sphere $\mathbb{S}^2$. According to standard black hole perturbation theory (see, for example, [Chandrasekhar 1998; Frolov and Novikov 1998; Price 2007]), and as developed at greater length in Section 3.3, it is possible to choose a gauge so that the polar and axial parts of perturbations to this metric are encoded in a single gauge-invariant variable each. In particular, they are given respectively by

$$\Phi_{(\pm)} = \frac{1}{r}\sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y^{lm}\left(\theta,\phi\right)\Psi^{lm}_{(\pm)}\left(t,r\right), \tag{4.4.2}$$

where Y^{lm} are spherical harmonics and $\Psi^{lm}_{(\pm)}$ are called, respectively, the Zerilli and Regge-Wheeler master functions, which satisfy known wave-like equations and from which the perturbations to $\boldsymbol{g}$ can be reconstructed. In [Jezierski 1999], the symplectic form of the reduced phase space $\mathscr{S}$ for such spacetimes is computed:

$$\boldsymbol{\Omega}|_{\mathscr{S}} = \sum_{\varsigma=\pm}\int_{\Sigma}\mathrm{d}^3x\,\delta\Upsilon_{(\varsigma)}\wedge\mathbb{D}\delta\Phi_{(\varsigma)}, \tag{4.4.3}$$

where $\Upsilon_{(\pm)} = [r^2\sin\theta/f(r)]\dot{\Phi}_{(\pm)}$, and $\mathbb{D} = \Delta_{\boldsymbol{\sigma}}^{-1}(\Delta_{\boldsymbol{\sigma}} + 2)^{-1}$ where $\Delta_{\boldsymbol{\sigma}}$ is the Laplace operator on $\mathbb{S}^2$. See Section 3.3. for more details.

The work [Jezierski 1999] where this symplectic form [Eq. (4.4.3)] was derived simply uses it to define and formulate conservation laws for energy and angular momentum in perturbed SD spacetimes. It does not, however, address the question of the total measure of $\mathscr{S}$. We will now show that the (reduced) phase space measure $\mu\left(\mathscr{S}\right) = \int_{\mathscr{S}}\boldsymbol{\Omega}|_{\mathscr{S}}$ for such spacetimes in fact diverges, preventing any argument based on phase space compactness for the non-existence of entropy production.

Inserting the definitions of the different variables and suppressing for the moment the coordinate dependence of the spherical harmonics and master functions, we have

$$\mu(\mathscr{S}) = \sum_{\varsigma=\pm} \int_{\mathscr{S}} \int_{\Sigma} \mathrm{d}^3x \, \delta\Upsilon_{(\varsigma)} \wedge \mathbb{D}\delta\Phi_{(\varsigma)} \tag{4.4.4}$$

$$= \sum_{\varsigma=\pm} \int_{\mathscr{S}} \int_{\Sigma} \mathrm{d}^3x \, \delta \left(\frac{r^2 \sin\theta}{f(r)} \dot{\Phi}_{(\varsigma)} \right) \wedge \mathbb{D}\delta\Phi_{(\varsigma)} \tag{4.4.5}$$

$$= \sum_{\varsigma=\pm} \int_{\mathscr{S}} \int_{\Sigma} \mathrm{d}^3x \, \delta \left(\frac{r \sin\theta}{f(r)} \sum_{l,m} Y^{lm} \dot{\Psi}^{lm}_{(\varsigma)} \right) \wedge \mathbb{D}\delta \left(\frac{1}{r} \sum_{l',m'} Y^{l'm'} \Psi^{l'm'}_{(\varsigma)} \right). \tag{4.4.6}$$

Now using the fact that the functional exterior derivative acts only on the master functions and the operator $\mathbb{D}$ only on the spherical harmonics, we can write this as

$$\mu(\mathscr{S}) = \sum_{\varsigma=\pm} \int_{\mathscr{S}} \int_{\Sigma} \mathrm{d}^3x \, \left(\frac{r \sin\theta}{f(r)} \sum_{l,m} Y^{lm} \delta\dot{\Psi}^{lm}_{(\varsigma)} \right) \wedge \left(\frac{1}{r} \sum_{l',m'} \left(\mathbb{D}Y^{l'm'} \right) \delta\Psi^{l'm'}_{(\varsigma)} \right) \tag{4.4.7}$$

$$= \sum_{\varsigma=\pm} \sum_{l,l',m,m'} \int_{\mathscr{S}} \int_{\Sigma} \mathrm{d}^3x \, \left(\frac{r \sin\theta}{f(r)} Y^{lm} \frac{1}{r} \mathbb{D}Y^{l'm'} \right) \delta\dot{\Psi}^{lm}_{(\varsigma)} \wedge \delta\Psi^{l'm'}_{(\varsigma)}. \tag{4.4.8}$$

Writing the Cauchy surface integral in terms of coordinates and collecting terms,

$$\mu(\mathscr{S}) = \sum_{\varsigma=\pm} \sum_{l,l',m,m'} \int_{\mathscr{S}} \int_{2M}^{\infty} \mathrm{d}r \int_{\mathbb{S}^2} \mathrm{d}\theta \mathrm{d}\phi \, \frac{1}{f(r)} \left[(\sin\theta) \, Y^{lm} \mathbb{D} Y^{l'm'} \right] \delta\dot{\Psi}^{lm}_{(\varsigma)} \wedge \delta\Psi^{l'm'}_{(\varsigma)} \tag{4.4.9}$$

$$= \sum_{\varsigma=\pm} \sum_{l,l',m,m'} \int_{\mathscr{S}} \left[\int_{\mathbb{S}^2} \mathrm{d}\theta \mathrm{d}\phi \, (\sin\theta) \, Y^{lm} \mathbb{D} Y^{l'm'} \right] \int_{2M}^{\infty} \frac{\mathrm{d}r}{f(r)} \delta\dot{\Psi}^{lm}_{(\varsigma)} \wedge \delta\Psi^{l'm'}_{(\varsigma)} \tag{4.4.10}$$

$$= \sum_{\varsigma=\pm} \sum_{l,l',m,m'} A^{ll'mm'} \int_{\mathscr{S}} \int_{2M}^{\infty} \frac{\mathrm{d}r}{f(r)} \delta\dot{\Psi}^{lm}_{(\varsigma)} \wedge \delta\Psi^{l'm'}_{(\varsigma)}, \tag{4.4.11}$$

where $A^{ll'mm'} = \int_{\mathbb{S}^2} \mathrm{d}\theta \mathrm{d}\phi \, (\sin\theta) Y^{lm} \mathbb{D} Y^{l'm'}$ is a finite integral involving only the spherical harmonics. Restoring the arguments of the master functions, and recalling that the meaning of $\delta f(t, r)$ (for any function f) is simply that of a one-form on the phase space at (t, r) in spacetime, we can write from the above [Eq. (4.4.11)]:

$$\mu(\mathscr{S}) = \sum_{\varsigma=\pm}\sum_{l,l',m,m'} A^{ll'mm'} \int_{\mathscr{S}} \int_{2M}^{\infty} \frac{\mathrm{d}r}{f(r)} \left[\delta\dot{\Psi}^{lm}_{(\varsigma)}(t,r) \wedge \delta\Psi^{l'm'}_{(\varsigma)}(t,r)\right] \tag{4.4.12}$$

$$\geq \sum_{\varsigma=\pm}\sum_{l,l',m,m'} A^{ll'mm'} \int_{\mathscr{S}} \int_{2M}^{\infty} \frac{\mathrm{d}r}{f(r)} \left[\min_{\bar{r}\in[2M,\infty)} \delta\dot{\Psi}^{lm}_{(\varsigma)}(t,\bar{r}) \wedge \delta\Psi^{l'm'}_{(\varsigma)}(t,\bar{r})\right] \tag{4.4.13}$$

$$= \sum_{\varsigma=\pm}\sum_{l,l',m,m'} A^{ll'mm'} \left[\int_{\mathscr{S}} \min_{\bar{r}\in[2M,\infty)} \delta\dot{\Psi}^{lm}_{(\varsigma)}(t,\bar{r}) \wedge \delta\Psi^{l'm'}_{(\varsigma)}(t,\bar{r})\right] \int_{2M}^{\infty} \frac{\mathrm{d}r}{f(r)} \tag{4.4.14}$$

$$= \left\{\sum_{\varsigma=\pm}\sum_{l,l',m,m'} A^{ll'mm'} \left[\min_{\bar{r}\in[2M,\infty)} \int_{\mathscr{S}} \delta\dot{\Psi}^{lm}_{(\varsigma)}(t,\bar{r}) \wedge \delta\Psi^{l'm'}_{(\varsigma)}(t,\bar{r})\right]\right\} \left[\int_{2M}^{\infty} \frac{\mathrm{d}r}{f(r)}\right]. \tag{4.4.15}$$

The phase space integral $\int_{\mathscr{S}} \delta\dot{\Psi}^{lm}_{(\varsigma)}(t,\bar{r}) \wedge \delta\Psi^{l'm'}_{(\varsigma)}(t,\bar{r})$ is finite but nonzero even when minimised over $\bar{r}$, because for any nontrivial solutions of the master functions, there will be no point in spacetime where they will always be vanishing (for all time). Thus (assuming that the l, l', m, m' sums are convergent), everything in the curly bracket in the last line above [Eq. (4.4.15)] is nonzero but finite. However, it multiplies $\int_{2M}^{\infty} \mathrm{d}r/f(r)$ which diverges (at both integration limits). Hence, $\mu(\mathscr{S})$ diverges for such spacetimes.

We can make a few remarks. Firstly, one might be concerned in the above argument, specifically in the last line [Eq. (4.4.15)], about what might happen in the asymptotic limit of the phase space integral: in other words, it maybe the case (i) that $\min_{\bar{r}\in[2M,\infty)} \int_{\mathscr{S}} \delta\dot{\Psi}^{lm}_{(\varsigma)}(t,\bar{r}) \wedge \delta\Psi^{l'm'}_{(\varsigma)}(t,\bar{r})$ could turn out to be $\lim_{r\to\infty} \int_{\mathscr{S}} \delta\dot{\Psi}^{lm}_{(\varsigma)}(t,r) \wedge \delta\Psi^{l'm'}_{(\varsigma)}(t,r)$; and, if so, one might naively worry (ii) that the latter vanishes due to asymptotic decay properties of the master functions. This will actually not happen. To see why, suppose (i) is true. The master functions must obey outgoing boundary conditions at spatial infinity, *i.e.* $0 = [\partial_t + f(r)\partial_r]\Psi^{lm}_{(\varsigma)}$ as $r \to \infty$. Hence we have

$$\begin{aligned} &\lim_{r\to\infty} \int_{\mathscr{S}} \delta\dot{\Psi}^{lm}_{(\varsigma)} \wedge \delta\Psi^{l'm'}_{(\varsigma)} \\ &= \lim_{r\to\infty} \int_{\mathscr{S}} \delta\left(-v\partial_r\Psi^{lm}_{(\varsigma)}\right) \wedge \delta\Psi^{l'm'}_{(\varsigma)} \\ &= -\int_{\mathscr{S}} \lim_{r\to\infty} \delta\left(\partial_r\Psi^{lm}_{(\varsigma)}\right) \wedge \delta\Psi^{l'm'}_{(\varsigma)}, \end{aligned} \tag{4.4.16}$$

which is nonzero, because the vanishing of the master functions and their radial partials at spatial infinity for all time corresponds only to trivial solutions. Therefore, we have that $\min_{\bar{r}\in[2M,\infty)} \int_{\mathscr{S}} \delta\dot{\Psi}^{lm}_{(\varsigma)}(t,\bar{r}) \wedge \delta\Psi^{l'm'}_{(\varsigma)}(t,\bar{r})$ is always nonzero for nontrivial solutions.

Secondly, if the two-body system in this framework is an extreme-mass-ratio inspiral, *i.e.* the mass of the orbiting body, or "particle", is orders of magnitude smaller than that of the larger one, and the former is modeled using a stress-energy-momentum tensor with support only on its worldline, then it is known that $\Psi^{lm}_{(\pm)}(t,r)$ has a discontinuity at the particle location, and thus, $\dot{\Psi}^{lm}_{(\pm)}(t,r)$ has a divergence there. Hence, the integral over $\mathscr{S}$ even before our inequality above [Eq. (4.4.12)] is already divergent due to the divergence of $\dot{\Psi}^{lm}_{(\pm)}(t,r)$ in the integrand. However, given that such an approach to describing these systems (*i.e.* having a stress-energy-momentum tensor of the particle with a delta distribution) is only an idealization, we regard the conclusion that $\mu(\mathscr{S})$ diverges as more convincing based on our earlier argument, which is valid in general—that is, even for possible descriptions of the smaller body that may be more realistic than that using delta distributions.

4.5. Conclusions

We have proven that there does not exist a monotonically increasing function of phase space—which may be identified as (what we have referred to as a "mechanical" notion of) entropy—in classical mechanics with N degrees of freedom for certain classes of Hamiltonians, as well as in some (classical) matter field theories in curved (nondynamical) spacetime, *viz.* for standard scalar and electromagnetic fields. To do this, we have followed the procedure for the proof sketched by Poincaré [Poincaré 1889] (what we have dubbed the *perturbative* approach), and we have here carried it out in full rigour for classical mechanics and extended it via similar techniques to field theories. What is noteworthy about this perturbative proof—counter (to our knowledge) to all other well-known proofs for the non-existence of entropy (in the "mechanical" sense) in classical canonical theories—is that it assumes nothing about the topology of the phase space; in other words, the phase space can be *non-compact.* Essentially, it relies only on curvature properties (in phase space) of the Hamiltonian of the canonical theory being considered. We have explicated these properties in the case of classical mechanics, and have assumed standard ones for the particular (curved spacetime) matter field theories we have investigated. It would be of interest for future work to determine, in the case of the former, whether they can be made less restrictive (than what we have required for our proof, which thus omits some classes of Hamiltonians of interest such as that for the the gravitational two-body problem), and in the case of the latter, whether they can be generalized or extended to broader classes of field theories. Indeed, it would be in general an interesting question to determine not only the necessary but also—*if possible*—the sufficient conditions that a Hamiltonian of a generic canonical theory needs to satisfy in order for this theorem to be applicable, *i.e.* in order to preclude "mechanical" entropy production. We have seen that it is precisely the curvature properties of the vacuum Hamiltonian of general relativity that prevent this method of proof from being extended thereto, where in fact one does expect (some version of) the second law of thermodynamics to hold.

Topological properties of the phase space can also entail the non-existence of "mechanical" entropy, as per the more standard and already well-understood proofs in classical mechanics where the phase space is assumed to be compact [Olsen 1993; Poincaré 1890]. However, even for non-gravitational canonical theories this assumption might be too restrictive, and for general relativity, it is believed that in general it is not the case. This renders any of these topological proofs inapplicable in the case of the latter, and moreover, it also significantly complicates any attempt to formulate a sensible "statistical" notion of (gravitational) entropy due to the concordant problems in working with finite probabilities (of phase space properties). These must ultimately be overcome (via some regularization procedure or cutoff argument) for establishing a connection between a "statistical" and "mechanical" entropy in general relativity. While we still lack any consensus on how to define the latter, it may be hoped that in the future, the generic validity of a (general relativistic) second law may be demonstrated on the basis of (perhaps curvature related) properties of the gravitational Hamiltonian—which in turn may enter into a statistical mechanics type definition of gravitational entropy in terms of some suitably defined partition function. In this regard, older work based on field-theoretic approaches [Horwitz 1983] and more recent developments such as proposals to relate entropy with a Noether charge (specifically, the Noether invariant associated with an infinitesimal time translation) in classical mechanics [Sasa and Yokokura 2016] may provide fruitful hints.

A clear situation in which we anticipate entropy production in general relativity, unlike in classical mechanics, is the gravitational two-body problem. For the latter, as we have discussed, the N-body problem actually does also exhibit features of entropy production. We have here shown explicitly that the phase space of perturbed Schwarzschild-Droste spacetimes is non-compact (even without the assumption of self-force). This means that the topological proofs are here inapplicable (but also, on the other hand, so is the "no-return" theorem for compact Cauchy surfaces, which by itself cannot be used in this case to understand the non-recurrence of phase space orbits). It is hoped that once a generally agreed upon definition of gravitational entropy is established, one would not only be able to use it to compute the entropy of two-body systems, but also to demonstrate that it should obey the second law (*i.e.* that it should be monotonically increasing in time). In the long run, an interesting problem to investigate is whether an entropy change, once defined and associated to motion in a Lagrangian formulation, could determine the trajectory of a massive and radiating body, moving in a gravitational field.

CHAPTER 5

The Motion of Localized Sources in General Relativity: Gravitational Self-Force from Quasilocal Conservation Laws

Chapter summary. This chapter is based on the preprint [Oltean, Epp, Sopuerta, et al. 2019].

An idealized "test" object in general relativity moves along a geodesic. However, if the object has a finite mass, this will create additional curvature in the spacetime, causing it to deviate from geodesic motion. If the mass is nonetheless sufficiently small, such an effect is usually treated perturbatively and is known as the gravitational self-force due to the object. This issue is still an open problem in gravitational physics today, motivated not only by basic foundational interest, but also by the need for its direct application in gravitational-wave astronomy. In particular, the observation of extreme-mass-ratio inspirals by the future space-based detector LISA will rely crucially on an accurate modeling of the self-force driving the orbital evolution and gravitational wave emission of such systems.

In this chapter, we present a novel derivation, based on conservation laws, of the basic equations of motion for this problem. They are formulated with the use of a quasilocal (rather than matter) stress-energy-momentum tensor—in particular, the Brown-York tensor—so as to capture gravitational effects in the momentum flux of the object, including the self-force. Our formulation and resulting equations of motion are independent of the choice of the perturbative gauge. We show that, in addition to the usual gravitational self-force term, they also lead to an additional "self-pressure" force not found in previous analyses, and also that our results correctly recover known formulas under appropriate conditions. Our approach thus offers a fresh geometrical picture from which to understand the self-force fundamentally, and potentially useful new avenues for computing it practically.

We begin in Section 5.1 with a brief introductory discussion on the idea of using conservation law approaches for the self-force problem generally, that is of understanding and computing the self-force as a momentum change or flux. While this has proven successful in the past for the electromagnetic self-force problem, such an approach has, up to this work, not been attempted in general in the gravitational case. This mainly has to do with the subtleties involved in properly defining notions of gravitational energy-momentum. These are concepts which do not make sense locally in relativistic physics (*i.e.* as volume densities), and so the typical solution—as in canonical general relativity—is to define them quasilocally (*i.e.* as boundary densities).

In Section 5.2, we review the general quasilocal energy-momentum conservation laws for general relativity used in this chapter. These laws have been obtained in recent work based on the Brown-York tensor, account for both gravitational as well as matter fluxes, and are valid in any arbitrary spacetime. They are constructed with the use of a concept called a quasilocal frame: a topological two-sphere of observers tracing out the worldtube boundary of the history of a finite spatial volume (that is, the finite system the fluxes of which we are studying).

In Section 5.3, we prove that the correction to the momentum flux of any small spatial region due to any metric perturbations in any spacetime in general relativity always contains the known form of the gravitational self-force. Our analysis also reveals a new term, not found in previous analyses and in principle equally dominant in general, namely one arising from a "self-pressure" effect with no analogy in Newtonian gravity. The appearance of these terms as corrections to the motion is independent of what actually sources the metric perturbations upon which they depend; rather than the "mass" of the small moving object itself, what seems to be fundamentally responsible for self-force effects in our analysis is the mass (or energy) and pressure of the spacetime vacuum.

In Section 5.4 we proceed to apply our analysis to a concrete self-force analysis actually used for computations, that is a specific choice of a perturbative family of spacetimes designed to describe the correction to the motion of a small object. We work with the rigorous approach of Gralla and Wald, and we show how under appropriate conditions our analysis recovers their equations of motion.

Finally, Section 5.5 offers some conclusions and outlook to future work.

La moció de les fonts localitzades en la relativitat general (chapter summary translation in Catalan). Aquest capítol es basa en el preprint [Oltean, Epp, Sopuerta, et al. 2019].

Un objecte de "prova" idealitzat en la relativitat general es mou al llarg d'una geodèsica. Tanmateix, si l'objecte té una massa finita, això crearà una corbatura addicional en l'espai-temps, fent que es desviï del moviment geodèsic. Si la massa és tot i això prou petita, aquest efecte se sol tractar de manera pertorbadora i es coneix com a força pròpia gravitacional a causa de l'objecte. Aquesta qüestió continua sent un problema obert en la física gravitatòria actual, motivada no només per l'interès fonamental bàsic, sinó també per la necessitat de la seva aplicació directa en l'astronomia d'ones gravitacionals. En particular, l'observació de caigudes en espiral amb raó de masses extrema per part del futur detector LISA basat en l'espai es basarà crucialment en un modelat precís de la força pròpia impulsant l'evolució orbital i l'emissió d'ones gravitacionals (crec que aquesta és la forma més correcta, tot in que són sinònimes, juntament amb gravitatòries) d'aquests sistemes.

En aquest capítol, es presenta una nova derivació, basada en lleis de conservació, de les equacions bàsiques de moviment d'aquest problema. Es formulen amb l'ús d'un tensor de tensió-energia quasilocal (en lloc de material), en particular, el tensor de Brown-York, per tal de captar efectes gravitacionals en el flux de moment de l'objecte, inclòs la força pròpia. La nostra formulació i les equacions de moviment resultants són independents de l'elecció de la mesura pertorbativa. Mostrem que, a més del terme de la força pròpia gravitacional habitual, també condueixen a una força de "pressió pròpia" addicional que no es va trobar en anàlisis anteriors, i també que els nostres resultats recuperen correctament les fórmules conegudes en condicions adequades. El nostre treball ofereix així una nova imatge geomètrica a partir de la qual es pot entendre fonamentalment la força pròpia, i possibles noves vies potencialment útils per a computar-la pràcticament.

Comencem a la secció 5.1 amb una breu discussió introductòria sobre la idea d'utilitzar els mètodes de lleis de conservació per al problema de la força pròpia en general, és a dir, comprendre i calcular la força pròpia com a canvi o flux d'impuls. Si bé en el passat això ha tingut èxit pel problema de la força pròpia electromagnètica, un anàlisi d'aquest tipus no s'ha intentat, fins a aquest treball, en general en el cas gravitatori. Això té a veure principalment amb les subtileses relacionades amb la definició adequada de les nocions d'energia-moment gravitatòria. Es tracta de conceptes que no tenen sentit localment en la física relativista (és a dir, com a densitats de volum), i per tant, la solució típica - com en la relativitat general canònica - és definir-los de forma quasilocal (és a dir, com a densitats de frontera).

A la secció 5.2, revisem les lleis quasilocals generals de conservació d'energia-moment per a la relativitat general utilitzades en aquest capítol. Aquestes lleis s'han obtingut en treballs recents basats en el tensor de Brown-York, tant per a fluxos gravitacionals com per a matèries, i són vàlids en qualsevol espai-temps arbitrari. Es construeixen amb l'ús d'un concepte conegut com a sistema de referència quasilocal: una esfera topològica bidimensional d'observadors que traça la frontera de la història d'un volum espacial finit (és a dir, el sistema finit els fluxos del qual estem estudiant).

A la secció 5.3, demostrem que la correcció al flux d'impuls de qualsevol petita regió espacial a causa de pertorbacions mètriques en qualsevol espai-temps en la relativitat general sempre conté la forma coneguda de la força pròpia gravitacional. La nostra anàlisi també revela un nou terme, que no es troba en anàlisis anteriors i, en principi, és igualment dominant en general, és a dir, un efecte de de "pressió pròpia" sense analogia en la gravetat newtoniana. L'aparició d'aquests termes com a correccions al moviment és independent del que realment provoca les pertorbacions mètriques de les quals depenen; més que la "massa" del petit objecte en moviment propi, el que sembla ser fonamentalment responsable dels efectes de la força pròpia en la nostra anàlisi és la massa (o energia) i la pressió del buit de l'espai-temps.

A la secció 5.4 procedim a aplicar la nostra anàlisi a una anàlisi concreta de força pròpia utilitzada realment per a càlculs, és a dir, una elecció específica d'una família pertorbadora d'espais-temps dissenyada per descriure la correcció al moviment d'un objecte petit. En particular, treballem amb el formalisme rigorós de Gralla i Wald, i mostrem com en condicions adequades la nostra anàlisi recupera les seves equacions de moviment.

Finalment, la secció 5.5 ofereix algunes conclusions i perspectives per a futurs treballs.

Le mouvement des sources localisées dans la relativité générale (chapter summary translation in French). Ce chapitre est basé sur le pré-impression [Oltean, Epp, Sopuerta, et al. 2019].

Un objet de « test » idéalisé dans la relativité générale se déplace le long d'une géodésique. Cependant, si l'objet a une masse finie, cela créera une courbure supplémentaire dans l'espace-temps, le faisant s'écarter du mouvement géodésique. Si la masse est néanmoins suffisamment petite, un tel effet est généralement traité de manière perturbative et il est connu comme la force propre gravitationnelle à cause de l'objet. Cette question est toujours un problème ouvert dans la physique gravitationnelle aujourd'hui, motivée non seulement par l'intérêt fondamental, mais également par la nécessité de son application directe dans l'astronomie des ondes gravitationnelles. En particulier, l'observation d'inspirals avec quotients extrêmes des masses par le futur détecteur spatial LISA reposera de manière cruciale sur une modélisation précise de la force propre à l'origine de l'évolution orbitale et de l'émission des ondes gravitationnelles de tels systèmes.

Dans ce chapitre, nous présentons une nouvelle dérivation, basée sur des lois de conservation, des équations de base du mouvement pour ce problème. Ils sont formulés avec l'utilisation d'un tenseur énergie-impulsion quasi-local (plutôt que de la matière) - en particulier, le tenseur de Brown-York - afin de capturer les effets gravitationnels dans le flux d'impulsion de l'objet, y compris la force propre. Notre formulation et les équations de mouvement résultantes sont indépendantes du choix de la jauge perturbative. Nous montrons que, en plus du terme habituel de la force propre gravitationnelle, ils conduisent également à une force de « pression propre » supplémentaire, pas trouvée dans les analyses précédentes et que nos résultats récupèrent correctement les formules connues dans des conditions appropriées. Notre analyse offre donc une nouvelle image géométrique à partir de laquelle on peut comprendre fondamentalement la force propre et de nouvelles voies potentiellement utiles pour la calculer de manière pratique.

Nous commençons à la section 5.1 par une brève discussion introductive sur l'idée d'appliquer les lois de conservation au problème de la force propre en général, c'est-à-dire de la compréhension et du calcul de la force propre comme un changement ou un flux en la quantité de mouvement. Bien que cela eût du succès au passé pour le problème de la force propre électromagnétique, une telle analyse n'a jusqu'à présent pas été tentée de manière générale dans le cas de la gravitation. Cela concerne principalement les subtilités impliquées dans la définition correcte des notions d'énergie-impulsion gravitationnelle. Ce sont des concepts qui n'ont pas de sens local dans la physique relativiste (c'est-à-dire, comme densités de volume) et la solution typique - comme dans la relativité générale canonique - consiste à les définir de manière quasilocale (c'est-à-dire, comme densités de frontière).

Dans la section 5.2, nous passons en revue les lois générales quasilocales de conservation d'énergie-impulsion pour la relativité générale utilisées dans ce chapitre. Ces lois ont été obtenues dans des travaux récents basés sur le tenseur de Brown-York, tiennent compte à la fois des flux gravitationnels et des flux de matière, et sont valables dans tout espace-temps arbitraire. Ils sont construits à l'aide d'un concept appelé le référentiel quasilocal (*quasilocal frame*): une sphère topologique bidimensionelle d'observateurs traçant la frontière de l'histoire d'un volume spatial fini (c'est-à-dire du système fini dont nous étudions les flux).

Dans la section 5.3, nous montrons que la correction du flux de la quantité de mouvement de toute petite région spatiale à cause des perturbations métriques dans quelque espace-temps de la relativité générale contient toujours la forme connue de la force propre gravitationnelle. Notre analyse révèle également un nouveau terme, pas retrouvé dans les analyses précédentes et en principe tout aussi dominant en général, à savoir un effet de « pression propre » sans analogie dans la gravité newtonienne. L'apparition de ces termes en tant que corrections du mouvement est indépendante de la source des perturbations métriques dont ils dépendent. Plutôt que la « masse » du petit objet en mouvement lui-même, ce qui semble être fondamentalement responsable pour les effets de force propre, dans notre analyse est la masse (ou énergie) et la pression du vide de l'espace-temps.

Dans la section 5.4, nous appliquons notre analyse à une formulation concrète de la force propre réellement utilisée pour les calculs, c'est-à-dire un choix spécifique d'une famille des espaces-temps perturbatifs conçue pour décrire la correction du mouvement d'un petit objet. Nous travaillons en particulier avec le formalisme rigoureux de Gralla et Wald et nous montrons comment, dans des conditions appropriées, notre analyse récupère leurs équations de mouvement.

Enfin, la section 5.5 propose quelques conclusions et perspectives pour les travaux futurs.

5.1. Introduction: the self-force problem via conservation laws

The idea of using conservation laws for tackling the self-force problem was appreciated and promptly exploited quite early on for the electromagnetic self-force. In the 1930s, [Dirac 1938] was the first to put forward such an analysis in flat spacetime, and later on

[DeWitt and Brehme 1960] extended it to non-dynamically curved spacetimes[1]. In such approaches, it can be shown[2] that the EoM for the electromagnetic self-force follows from local conservation expressions of the form

$$\Delta P^a = \int_{\Delta\mathscr{B}} \boldsymbol{\epsilon}_{\mathscr{B}} T^{ab} n_b \,, \tag{5.1.1}$$

where the LHS expresses the flux of *matter* four-momentum P^a between the "caps" of (*i.e.* closed spatial two-surfaces delimiting) a portion (or "time interval") of a thin worldtube boundary $\mathscr{B}$ (topologically $\mathbb{R} \times \mathbb{S}^2$), with natural volume form $\boldsymbol{\epsilon}_{\mathscr{B}}$ and (outward-directed) unit normal n^a (see Figure 5.1). In particular, one takes a time derivative of (5.1.1) to obtain an EoM expressing the time rate of change of momentum in the form of a closed spatial two-surface integral (by differentiating the worldtube boundary integral). For the electromagnetic self-force problem, the introduction of an appropriate matter stress-energy-momentum tensor T_{ab} into Eq. (5.1.1) and a bit of subsequent argumentation reduces the integral expression to the famous Lorentz-Dirac equation; on a spatial three-slice in a Lorentz frame and in the absence of external forces, for example, this simply reduces to $\dot{P}^i = \frac{2}{3} q^2 \dot{a}^i$ for a charge q. Formulations of the scalar and electromagnetic self-forces using generalized Killing fields have more recently been put forward in [Harte 2008; Harte 2009].

The success of conservation law approaches for formulating the electromagnetic self-force in itself inspires hope that the same may be done in the case of the gravitational self-force (GSF) problem. In particular, Gralla's formula in Eq. (1.5.7) strongly hints at the possibility of understanding the RHS not just as a mathematical ("angle averaging") device, but as a *true, physical flux of gravitational momentum* arising from a consideration of conservation expressions.

Nevertheless, to our knowledge, there has thus far been no proposed general treatment of the GSF following such an approach. This may, in large part, be conceivably attributed to the notorious conceptual difficulties surrounding the very question of the basic formulation of conservation laws in GR. Local conservation laws, along the lines of Eq. (5.1.1) that can readily be used for electromagnetism, no longer make sense once gravity is treated as dynamical. The reason has a simple explanation in the equivalence principle [Misner et al. 1973]: one can always find a local frame of reference with a vanishing local "gravitational field" (metric connection coefficients), and hence a vanishing local "gravitational energy-momentum", irrespective of how one might feel inclined to define the latter.

[1] By this, we mean spacetimes with non-flat but fixed metrics, which do not evolve dynamically (gravitationally) in response to the matter stress-energy-momentum present therein.
[2] See [Poisson 1999] for a basic and more contemporary presentation.

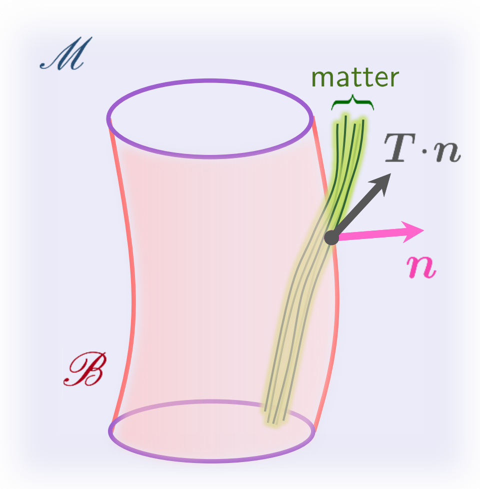

FIGURE 5.1. A worldtube boundary $\mathscr{B}$ (topologically $\mathbb{R} \times \mathbb{S}^2$) in $\mathscr{M}$, with (outward-directed) unit normal n^a. The change in matter four-momentum between two constant time slices of this worldtube is given by the flux of the normal projection (in one index) of the matter stress-energy-momentum tensor T_{ab} through the portion of $\mathscr{B}$ bounded thereby.

A wide variety of approaches have been taken over the decades towards formulating sensible notions of gravitational energy-momentum, with still no general consensus among relativists today on which to qualify as "the best" [Jaramillo and Gourgoulhon 2011; Szabados 2004]. Often the preference for employing certain definitions over others may simply come down to context or convenience, but in any case, there exist agreements between the most typical definitions in various limits. A very common feature among them is the idea of replacing a local notion of gravitational energy-momentum, *i.e.* energy-momentum as a volume density, with what is referred to as a *quasilocal* energy-momentum, *i.e.* energy-momentum as a boundary density. The typical Hamiltonian definitions of the (total) gravitational energy-momentum for an asymptotically-flat spacetime, for example, are of such a form. Among the most commonly used generalizations of these definitions to arbitrary (finite) spacetime regions was proposed in the early 1990s by [J. D. Brown and York 1993], and follow from what is now eponymously known as the Brown-York stress-energy-momentum tensor. It is a quasilocal tensor, meaning it is only defined on the boundary of an arbitrary spacetime region. For example, using this, the total (matter plus gravitational) energy inside a spatial volume is given up to a constant factor by the closed two-surface (boundary) integral of the trace of the boundary extrinsic curvature—precisely in agreement with the Hamiltonian definition of energy for the entire spacetime in the appropriate limit (where the closed two-surface approaches a two-sphere at asymptotically-flat spatial infinity) but, in principle, applicable to any region in any spacetime.

The formulation of general energy-momentum conservation laws in GR from the Brown-York tensor has been achieved with the use of a construction called *quasilocal frames* [Epp, P. L. McGrath, et al. 2013], a concept first proposed in [Epp, Mann, et al. 2009]. Essentially, the idea is that it does not here suffice to merely specify, as in the local matter conservation laws of the form of Eq. (5.1.1), a worldtube boundary $\mathscr{B}$ (as an embedded submanifold of $\mathscr{M}$) the interior of which contains the system of interest, and through which to measure the flux of gravitational energy-momentum. What is in fact required is the specification of a congruence making up this worldtube boundary, *i.e.* a two-parameter family of timelike worldlines with some chosen four-velocity field representing the motion of a topological two-sphere worth of *quasilocal observers*. We will motivate this construction in greater amplitude shortly, but the reason for needing it is basically to be able to meaningfully define "time-time" and "time-space" directions on $\mathscr{B}$ for our conservation laws. A congruence of this sort is what is meant by a quasilocal frame.

The enormous advantage in using these quasilocal conservation laws over other approaches lies in the fact that they hold in any arbitrary spacetime. Thus the existence of Killing vector fields—a typical requirement in other conservation law formulations—is in no way needed here.

This idea has been used successfully in a number of applications so far [Epp, Mann, et al. 2012; Epp, P. L. McGrath, et al. 2013; P. McGrath 2014; P. L. McGrath, Chanona, et al. 2014; P. L. McGrath, Epp, et al. 2012; Oltean, Epp, P. L. McGrath, et al. 2017, 2016]. These include the resolution of a variation of Bell's spaceship paradox[3] in which a box accelerates rigidly in a transverse, uniform electric field [P. L. McGrath, Epp, et al. 2012], recovering under appropriate conditions the typical (but more limited) local matter conservation expressions of the form of Eq. (5.1.1) from the quasilocal ones [Epp, P. L. McGrath, et al. 2013], application to post-Newtonian theory [P. L. McGrath, Chanona, et al. 2014] and to relativistic geodesy [Oltean, Epp, P. L. McGrath, et al. 2017, 2016].

A similar idea to quasilocal frames, called "gravitational screens," was proposed more recently in Refs. [Freidel 2015; Freidel and Yokokura 2015]. There, the authors also make use of quasilocal ideas to develop conservation laws very similar in style and form to those obtained via quasilocal frames. A detailed comparison between these two approaches has thus far not been carried out, but it would be very interesting to do so in future work. In

[3] Proposed initially by [Dewan and Beran 1959] and later made popular by J.S. Bell's version [Bell 1976].

particular, the notion of gravitational screens has been motivated more from thermodynamic considerations, and similarly casting quasilocal frames in this language could prove quite fruitful. For example, just as these approaches have given us operational definitions of concepts like the "energy in an arbitrary spacetime region" (and not just for special cases such as an entire spacetime), they may help to do the same for concepts like "entropy in an arbitrary spacetime region" (and not just for known special cases such as a black hole).

5.2. Setup: quasilocal conservation laws

Let $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ be any $(3+1)$-dimensional spacetime such that, given any matter stress-energy-momentum tensor T_{ab}, the Einstein equation,

$$\boldsymbol{G} = \kappa \boldsymbol{T} \;\; \text{in} \;\; \mathscr{M} \,, \tag{5.2.1}$$

holds. In what follows, we introduce the concept of quasilocal frames [Epp, Mann, et al. 2012, 2009; Epp, P. L. McGrath, et al. 2013; P. McGrath 2014; P. L. McGrath, Chanona, et al. 2014; P. L. McGrath, Epp, et al. 2012; Oltean, Epp, P. L. McGrath, et al. 2017, 2016] and describe the basic steps for their construction, as well as the energy and momentum conservation laws associated therewith. In Subsection 5.2.1 we offer an heuristic idea of quasilocal frames before proceeding in Subsection 5.2.2 to present the full mathematical construction. Then in Subsection 5.2.3 we motivate and discuss the quasilocal stress-energy-momentum tensor used in this work, in particular the Brown-York tensor. Finally in Subsection 5.2.4 we review the formulation of quasilocal conservation laws using these ingredients.

5.2.1. Quasilocal frames: heuristic idea. Before we enter into the technical details, we would like to offer a heuristic picture and motivation for defining the concept of quasilocal frames.

We would like to show how the GSF arises from general-relativistic conservation laws. For this, we require first the embedding into our spacetime $\mathscr{M}$ of a worldtube boundary $\mathscr{B} \simeq \mathbb{R} \times \mathbb{S}^2$. The interior of $\mathscr{B}$ contains the system the dynamics of which we are interested in describing. In principle, such a $\mathscr{B}$ can be completely specified by choosing an appropriate "radial function" $r(x)$ on $\mathscr{M}$ and setting it equal to a non-negative constant (such that the $r(x) = \text{const.} \geq 0$ Lorentzian slices of $\mathscr{M}$ have topology $\mathbb{R} \times \mathbb{S}^2$). This would be analogous to defining a (Riemannian, with topology $\mathbb{R}^3$) Cauchy surface by the constancy of a "time function" $t(x)$ on $\mathscr{M}$.

However, this does not quite suffice. As we have briefly argued in the introduction (and will shortly elaborate upon in greater technicality), the conservation laws appropriate to GR ought to be quasilocal in form, that is, involving stress-energy-momentum as boundary (not volume) densities. One may readily assume that the latter are defined by a quasilocal stress-energy-momentum tensor living on $\mathscr{B}$, which we denote—for the moment, generally—by τ_{ab}. (Later we give an explicit definition, namely that of the Brown-York tensor, for $\boldsymbol{\tau}$.)

To construct conservation laws, then, one would need to project this $\boldsymbol{\tau}$ into directions on $\mathscr{B}$, giving quantities such as energy or momenta, and then to consider their flux through a portion of $\mathscr{B}$ (an interval of time along the worldtube boundary). But in this case, we have to make clear what is meant by the energy ("time-time") and momenta ("time-space") components of $\boldsymbol{\tau}$ within $\mathscr{B}$, the changes in which we are interested in studying. For this reason, additional constructions are required.

In particular, what we need is a *congruence* of observers with respect to which projections of $\boldsymbol{\tau}$ yield stress-energy-momentum quantities. Since $\boldsymbol{\tau}$ is only defined on $\mathscr{B}$, this therefore needs to be a two-parameter family of (timelike) worldlines the union of which is $\mathscr{B}$ itself. This is analogous to how the integral curves of a "time flow" vector field (as in canonical GR) altogether constitute ("fill up") the entire spacetime $\mathscr{M}$, except in that case we are dealing with a three- (rather than two-) parameter family of timelike worldlines.

We refer to any set of observers, the worldlines of which form a two-parameter family constituting $\mathscr{B} \simeq \mathbb{R} \times \mathbb{S}^2$, as *quasilocal* observers. A specification of such a 2-parameter family, equivalent to specifying the unit four-velocity $u^a \in T\mathscr{B}$ of these observers (the integral curves of which "trace out" $\mathscr{B}$), is what is meant by a quasilocal frame.

With this, we can now meaningfully talk about projections of $\boldsymbol{\tau}$ into directions on $\mathscr{B}$ as stress-energy-momentum quantities. For example, $\tau_{\boldsymbol{uu}}$ may appear immediately suggestible as a definition for the (boundary) energy density. Indeed, later we take precisely this definition, and we will furthermore see how momenta (the basis of the GSF problem) can be defined as well.

5.2.2. Quasilocal frames: mathematical construction. Concordant with our discussion in the previous subsection, a quasilocal frame (see Figure 5.2 for a graphical illustration of the construction) is defined as a two-parameter family of timelike worldlines constituting the worldtube boundary (topologically $\mathbb{R} \times \mathbb{S}^2$) of the history of a finite (closed) spatial three-volume in $\mathscr{M}$. Let u^a denote the timelike unit vector field tangent to these worldlines. Such a congruence constitutes a submanifold of $\mathscr{M}$ that we call $\mathscr{B} \simeq \mathbb{R} \times \mathbb{S}^2$. Let n^a be the outward-pointing unit vector field normal to $\mathscr{B}$; note that $\boldsymbol{n}$ is uniquely fixed once $\mathscr{B}$ is specified. There is thus a Lorentzian metric $\boldsymbol{\gamma}$ (of signature $(-,+,+)$) induced on $\mathscr{B}$, the components of which are given by

$$\gamma_{ab} = g_{ab} - n_a n_b \,. \tag{5.2.2}$$

We denote the induced derivative operator compatible therewith by $\boldsymbol{\mathcal{D}}$. To indicate that a topologically $\mathbb{R} \times \mathbb{S}^2$ submanifold $(\mathscr{B}, \boldsymbol{\gamma}, \boldsymbol{\mathcal{D}})$ of $\mathscr{M}$ is a quasilocal frame (that is to say, defined as a particular congruence with four-velocity $\boldsymbol{u}$ as detailed above, and not just as an embedded submanifold) in $\mathscr{M}$, we write $(\mathscr{B}, \boldsymbol{\gamma}, \boldsymbol{\mathcal{D}}; \boldsymbol{u})$ or simply $(\mathscr{B}; \boldsymbol{u})$.

Let $\mathscr{H}$ be the two-dimensional subspace of $T\mathscr{B}$ consisting of the "spatial" vectors orthogonal to $\boldsymbol{u}$. Let $\boldsymbol{\sigma}$ denote the two-dimensional (spatial) Riemannian metric (of signature $(+,+)$) that projects tensor indices into $\mathscr{H}$, and is induced on $\mathscr{B}$ by the choice of $\boldsymbol{u}$ (and thus also $\boldsymbol{n}$), given by

$$\sigma_{ab} = \gamma_{ab} + u_a u_b = g_{ab} - n_a n_b + u_a u_b \,. \tag{5.2.3}$$

The induced derivative operator compatible with $\boldsymbol{\sigma}$ is denoted by $\boldsymbol{D}$. Let $\{x^{\mathfrak{i}}\}_{\mathfrak{i}=1}^{2}$ (written using Fraktur indices from the middle third of the Latin alphabet) be "spatial" coordinates on $\mathscr{B}$ that label the worldlines of the observers, and let t be a "time" coordinate on $\mathscr{B}$ such that surfaces of constant t, to which there exists a unit normal vector that we denote by $\tilde{u}^a$, foliate $\mathscr{B}$ by closed spatial two-surfaces $\mathscr{S}$ (with topology $\mathbb{S}^2$). Letting N denote the lapse function of $\boldsymbol{g}$, we have $\boldsymbol{u} = N^{-1}\partial/\partial t$.

Note that in general, $\mathscr{H}$ need not coincide with the constant time slices $\mathscr{S}$. Equivalently, $\boldsymbol{u}$ need not coincide with $\tilde{\boldsymbol{u}}$. In general, there will be a shift between them, such that

$$\tilde{\boldsymbol{u}} = \tilde{\gamma}(\boldsymbol{u} + \boldsymbol{v}) \,, \tag{5.2.4}$$

where v^a represents the spatial two-velocity of fiducial observers that are at rest with respect to $\mathscr{S}$ as measured by our congruence of quasilocal observers (the four-velocity of which is $\boldsymbol{u}$), and $\tilde{\gamma} = 1/\sqrt{1 - \boldsymbol{v}\cdot\boldsymbol{v}}$ is the Lorentz factor.

The specification of a quasilocal frame is thus equivalent to making a particular choice of a two-parameter family of timelike worldlines comprising $\mathscr{B}$. There are, a priori, three degrees of freedom (DoFs) available to us for doing this. Heuristically, these can be regarded as corresponding to the three DoFs in choosing the direction of $\boldsymbol{u}$—from which $\boldsymbol{n}$ and all induced quantities are then computable. (Note that $\boldsymbol{u}$ has four components, but one of the four is fixed by the normalization requirement $\boldsymbol{u}\cdot\boldsymbol{u} = -1$, leaving three independent direction DoFs.) Equivalently, we are in principle free to pick any three geometrical conditions (along the congruence) to fix a quasilocal frame. In practice, usually it is physically more natural, as well as mathematically easier, to work with geometric quantities other than $\boldsymbol{u}$ itself to achieve this.

Yet, it is worth remarking that simply writing down three desired equations (or conditions) to be satisfied by geometrical quantities on $\mathscr{B}$ does not itself guarantee that, in general, a submanifold $(\mathscr{B}, \boldsymbol{\gamma}, \boldsymbol{\mathcal{D}})$ obeying those three particular equations will always exist—and, if it does, that it will be the unique such submanifold—in an arbitrary $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$. Nevertheless, one choice of quasilocal frame that is known to always exist (a claim we will qualify more carefully in a moment) is that where the two-metric $\boldsymbol{\sigma}$ on $\mathscr{H}$ is "rigid" (or "time" independent)—these are called *rigid quasilocal frames.*

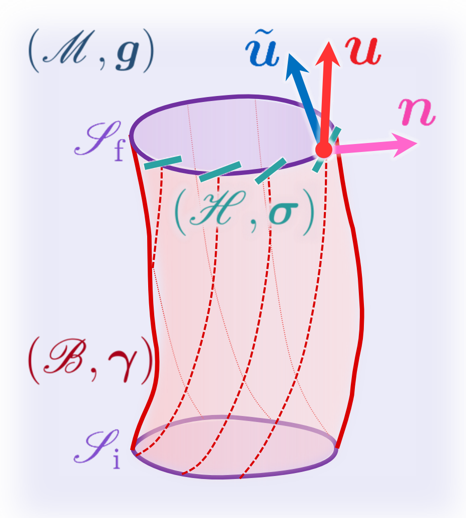

FIGURE 5.2. A portion of a quasilocal frame $(\mathscr{B}; \boldsymbol{u})$ in a spacetime $\mathscr{M}$, bounded by constant t two-surfaces $\mathscr{S}_{\rm i}$ and $\mathscr{S}_{\rm f}$. In particular, $\mathscr{B} \simeq \mathbb{R} \times \mathbb{S}^2$ is the union of all integral curves (two-parameter family of timelike worldlines), depicted in the figure as dotted red lines, of the vector field $\boldsymbol{u} \in T\mathscr{B}$ which represents the unit four-velocity of quasilocal observers making up the congruence. The unit normal to $\mathscr{B}$ (in $\mathscr{M}$) is $\boldsymbol{n}$ and the normal to each constant t slice $\mathscr{S}$ of $\mathscr{B}$ is $\tilde{\boldsymbol{u}}$ (not necessarily coincidental with $\boldsymbol{u}$). Finally, $\mathscr{H}$ (with induced metric $\boldsymbol{\sigma}$) is the two-dimensional subspace of $T\mathscr{B}$ consisting of the "spatial" vectors orthogonal to $\boldsymbol{u}$. Note that unlike $\mathscr{S}$, $\mathscr{H}$ need not be integrable (indicated in the figure by the failure of $\mathscr{H}$ to make a closed two-surface).

Most of the past work that has been done with quasilocal frames has in fact been done in the rigid case [Epp, Mann, et al. 2012, 2009; Epp, P. L. McGrath, et al. 2013; P. McGrath 2014; P. L. McGrath, Chanona, et al. 2014; P. L. McGrath, Epp, et al. 2012]. We know however that other quasilocal frame choices are also possible, such as *geoids*—dubbed geoid quasilocal frames [Oltean, Epp, P. L. McGrath, et al. 2017, 2016]: these are the general-relativistic generalization of "constant gravitational potential" surfaces in Newtonian gravity. Regardless, the quasilocal frame choice that we will mainly consider in this paper is the rigid one (and we will be clear when this choice is explicitly enacted).

Intuitively, the reason for this preference is that imposing in this way the condition of "spatial rigidity" on $(\mathscr{B}; \boldsymbol{u})$—a two-dimensional (boundary) rigidity requirement, which unlike three-dimensional rigidity, is permissible in GR—eliminates from the description of the system any effects arising simply from the motion of the quasilocal observers relative to each other. Thus, the physics of what is going on inside the system (worldtube) is essentially all that affects its dynamics.

Technically, there is a further reason: a proof of the existence of solutions—i.e. the existence of a submanifold $\mathscr{B} \simeq \mathbb{R} \times \mathbb{S}^2$ in $\mathscr{M}$ that is also a quasilocal frame $(\mathscr{B}; \boldsymbol{u})$—for any spacetime $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ has up to now only been fully carried out for rigid quasilocal frames[4]. While, as we have commented, other quasilocal frame choices may be generally possible in principle (and may be shown to be possible to construct, case-by-case, in specific spacetimes—as we have done, e.g., with geoid quasilocal frames [Oltean, Epp, P. L. McGrath, et al. 2017, 2016]), they are as yet not rigorously guaranteed to exist in arbitrary spacetimes.

The quasilocal rigidity conditions can be stated in a number of ways. Most generally, defining

$$\theta_{ab} = \sigma_{ac}\sigma_{bd}\nabla^c u^d \tag{5.2.5}$$

to be the *strain rate tensor* of the congruence, they amount to the requirement of vanishing expansion $\theta = \mathrm{tr}(\boldsymbol{\theta})$ and shear $\theta_{\langle ab\rangle} = \theta_{(ab)} - \frac{1}{2}\theta\sigma_{ab}$, *i.e.*

$$\theta = 0 = \theta_{\langle ab\rangle} \Leftrightarrow 0 = \theta_{(ab)}\,. \tag{5.2.6}$$

In the adapted coordinates, these three conditions are expressible as the vanishing of the time derivative of the two-metric on $\mathscr{H}$, i.e. $0 = \partial_t \boldsymbol{\sigma}$. Both of these two equivalent mathematical conditions, $\theta_{(ab)} = 0 = \partial_t \boldsymbol{\sigma}$, capture physically the meaning of the quasilocal observers moving "rigidly" with respect to each other (i.e. the "radar-ranging" distances between them does not change in time).

5.2.3. The quasilocal stress-energy-momentum tensor. Before we consider the formulation of conservation laws with the use of quasilocal frames (from which our analysis of the GSF will eventually emerge), we wish to address in a bit more detail an even more fundamental question: what are conservation laws in GR actually supposed to be about? At the most basic level, they should express changes (over time) in some appropriately defined notion of energy-momentum. As we are interested in gravitational systems (and specifically, those driven by the effect of the GSF), this energy-momentum must include that of the gravitational field, in addition to that of any matter fields if present.

Hence, we may assert from the outset that it does not make much sense in GR to seek conservation laws based solely on the matter stress-energy-momentum tensor $\boldsymbol{T}$, such as Eq. (5.1.1). It is evident that these would, by construction, account for matter only—leaving out gravitational effects in general (which could exist in the complete absence

[4] The idea of the proof is to explicitly construct the solutions order-by-order in an expansion in the areal radius around an arbitrary worldline in an arbitrary spacetime [Epp, Mann, et al. 2012].

of matter, e.g. gravitational waves), and thus the GSF in particular. What is more, such conservation laws are logically inconsistent from a general-relativistic point of view: a non-vanishing $\boldsymbol{T}$ implies a non-trivial gravitational field (through the Einstein equation) and thus a necessity of taking into account that field along with the matter one(s) for a proper accounting of energy-momentum transfer. A further technical problem is also that the formulation of conservation laws of this sort is typically predicated upon the existence of Killing vector fields or other types of symmetry generators in $\mathscr{M}$, which one does not have in general—and which do not exist in spacetimes pertinent for the GSF problem in particular.

We are therefore led to ask: how can we meaningfully define a total—*gravity plus matter*—stress-energy-momentum tensor in GR? It turns out that the precise answer to this question, while certainly not intractable, is unfortunately also not unique—or at least, it lacks a clear consensus among relativists, even today. See, e.g., Refs. [Jaramillo and Gourgoulhon 2011; Szabados 2004] for reviews of the variety of proposals that have been put forward towards addressing this question. Nonetheless, for reasons already touched upon and to be elaborated presently, what is clear and generally accepted is that such a tensor cannot be local in nature (as $\boldsymbol{T}$ is), and for this reason is referred to as "quasilocal."

Let τ_{ab} denote this quasilocal, total (matter plus gravity) stress-energy-momentum tensor that we eventually seek to use for our conservation laws. It has long been understood [Misner et al. 1973] that whatever the notion of "gravitational energy-momentum" (defined by $\boldsymbol{\tau}$) might mean, it is not something localizable: in other words, there is no way of meaningfully defining an "energy-momentum volume density" for gravity. This is, ultimately, due to the equivalence principle: locally, one can always find a reference frame in which all local "gravitational fields" (the connection coefficients), and thus any notion of "energy-momentum volume density" associated therewith, disappear. The remedy is to make $\boldsymbol{\tau}$ quasilocal: meaning that, rather than volume densities, it should define surface densities (of energy, momentum etc.)—a type of construction which is mathematically realizable and physically sensible in general.

The specific choice we make for how to define this total (matter plus gravity), quasilocal energy-momentum tensor $\boldsymbol{\tau}$ is the so-called Brown-York tensor, first put forward by the authors in [J. D. Brown and York 1993]; see also [J. D. Brown, Lau, et al. 2002] for a detailed review. This proposal was based originally upon a Hamilton-Jacobi analysis; here we will offer a simpler argument for its definition, sketched out initially in [Epp, P. L. McGrath, et al. 2013].

Consider the standard gravitational action S_{G} for the spacetime volume $\mathscr{V} = \mathrm{int}(\mathscr{B}) \subset \mathscr{M}$, where $\mathscr{B} \simeq \mathbb{R} \times \mathbb{S}^2$ is a worldtube boundary as in the previous subsection (possibly constituting a quasilocal frame, but not necessarily). This action is given by the sum of two terms, a bulk and a boundary term respectively:

$$S_{\mathrm{G}}[\boldsymbol{g}] = S_{\mathrm{EH}}[\boldsymbol{g}] + S_{\mathrm{GHY}}[\boldsymbol{\gamma}, \boldsymbol{n}] \ . \tag{5.2.7}$$

In particular, the first is the Einstein-Hilbert bulk term,

$$S_{\mathrm{EH}}[\boldsymbol{g}] = \frac{1}{2\kappa} \int_{\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{M}}\, R \,, \tag{5.2.8}$$

and the second is the Gibbons-Hawking-York boundary term [Gibbons and Hawking 1977; York 1972],

$$S_{\mathrm{GHY}}[\boldsymbol{\gamma}, \boldsymbol{n}] = -\frac{1}{\kappa} \int_{\partial\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{B}}\, \Theta \,. \tag{5.2.9}$$

Here, $\boldsymbol{\epsilon}_{\mathscr{M}} = \mathrm{d}^4 x \sqrt{-g}$ is the volume form on $\mathscr{M}$ with $g = \det(\boldsymbol{g})$, $\boldsymbol{\epsilon}_{\mathscr{B}} = \mathrm{d}^3 x \sqrt{-\gamma}$ is the volume form on $\mathscr{B}$ with $\gamma = \det(\boldsymbol{\gamma})$, and $\Theta = \mathrm{tr}(\boldsymbol{\Theta})$ is the trace of the extrinsic curvature $\Theta_{ab} = \gamma_{ac} \nabla^c n_b$ of $\mathscr{B}$ in $\mathscr{M}$. Additionally, the matter action S_{M} for any set of matter fields Ψ described by a Lagrangian L_{M} is

$$S_{\mathrm{M}}[\Psi] = \int_{\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{M}}\, L_{\mathrm{M}}[\Psi] \ . \tag{5.2.10}$$

The definition of the total (quasilocal) stress-energy-momentum tensor $\boldsymbol{\tau}$ for gravity plus matter can be obtained effectively in the same way as that of the (local) stress-energy-momentum tensor $\boldsymbol{T}$ for matter alone—from the total action in Eq. (5.2.7) rather than just, respectively, the matter action in Eq. (5.2.10). In particular, $\boldsymbol{T}$ is defined by computing the variation δ (with respect to the spacetime metric) of the matter action:

$$\delta S_{\mathrm{M}}[\Psi] = -\frac{1}{2} \int_{\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{M}}\, T_{ab} \delta g^{ab} \,. \tag{5.2.11}$$

In other words, one defines the matter stress-energy-momentum tensor as the functional derivative,

$$T_{ab} = -\frac{2}{\sqrt{-g}} \frac{\delta S_{\mathrm{M}}}{\delta g^{ab}} \,. \tag{5.2.12}$$

The definition of the Brown-York tensor follows completely analogously, except that now gravity is also included. That is, for the total action of gravity (minimally) coupled to matter,

$$S_{\mathrm{G+M}}[\boldsymbol{g}, \Psi] = S_{\mathrm{G}}[\boldsymbol{g}] + S_{\mathrm{M}}[\Psi] \ , \tag{5.2.13}$$

we have that the metric variation is:

$$\delta S_{\mathrm{G+M}}[\boldsymbol{g},\Psi] = \frac{1}{2}\left\{\int\limits_{\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{M}} \left(\frac{1}{\kappa}G_{ab} - T_{ab}\right)\delta g^{ab} - \int\limits_{\partial\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{B}} \left(-\frac{1}{\kappa}\Pi_{ab}\right)\delta\gamma^{ab}\right\} \tag{5.2.14}$$

$$= -\frac{1}{2}\int\limits_{\partial\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{B}}\, \tau_{ab}\delta\gamma^{ab}\,. \tag{5.2.15}$$

In the equality of Eq. (5.2.14), $\boldsymbol{\Pi}$ is the canonical momentum of $(\mathscr{B},\boldsymbol{\gamma},\boldsymbol{\mathcal{D}})$, given by $\boldsymbol{\Pi} = \boldsymbol{\Theta} - \Theta\boldsymbol{\gamma}$. It follows from direct computation using Eqs. (5.2.7), (5.2.10) and (5.2.11); for a review of this derivation carefully accounting for the boundary term see, e.g., Chapter 12 of [Padmanabhan 2010]. In the equality of Eq. (5.2.15), the Einstein equation $\boldsymbol{G} = \kappa\boldsymbol{T}$ has been invoked (in other words, we impose the Einstein equation to be satisfied in the bulk), thus leading to the vanishing of the bulk term; meanwhile in the boundary term, a gravity plus matter stress-energy-momentum tensor $\boldsymbol{\tau}$ (the Brown-York tensor) has been defined in direct analogy with the definition of the matter energy-momentum tensor $\boldsymbol{T}$ in Eq. (5.2.11). Hence just as Eq. (5.2.11) implies Eq. (5.2.12), Eq. (5.2.15) implies

$$\boldsymbol{\tau} = -\frac{1}{\kappa}\boldsymbol{\Pi}\,. \tag{5.2.16}$$

Henceforth, $\boldsymbol{\tau}$ refers strictly to this (Brown-York) quasilocal stress-energy-momentum tensor of Eq. (5.2.16), and not to any other definition.

It is useful to decompose $\boldsymbol{\tau}$ in a similar way as is ordinarily done with $\boldsymbol{T}$, so we define:

$$\mathcal{E} = u^a u^b \tau_{ab}\,, \tag{5.2.17}$$

$$\mathcal{P}^a = -\sigma^{ab}u^c\tau_{bc}\,, \tag{5.2.18}$$

$$\mathcal{S}^{ab} = -\sigma^{ac}\sigma^{bd}\tau_{cd}\,, \tag{5.2.19}$$

as the quasilocal energy, momentum and stress, respectively, with units of energy per unit area, momentum per unit area and force per unit length. Equivalently,

$$\tau^{ab} = u^a u^b\mathcal{E} + 2u^{(a}\mathcal{P}^{b)} - \mathcal{S}^{ab}\,. \tag{5.2.20}$$

5.2.4. Conservation laws. The construction of general conservation laws from $\boldsymbol{\tau}$ was first achieved in Refs. [Epp, P. L. McGrath, et al. 2013; P. L. McGrath, Epp, et al. 2012], and proceeds along the following lines. Let $\boldsymbol{\psi} \in T\mathscr{B}$ be an arbitrary vector field in $\mathscr{B}$. We begin by considering a projection of $\boldsymbol{\Pi}$ in the direction of $\boldsymbol{\psi}$ (in one index), i.e. $\Pi^{ab}\psi_b$, and computing its divergence in $\mathscr{B}$. By using the Leibnitz rule, we simply have

$$\mathcal{D}_a\left(\Pi^{ab}\psi_b\right) = \left(\mathcal{D}_a\Pi^{ab}\right)\psi_b + \Pi^{ab}\left(\mathcal{D}_a\psi_b\right)\,. \tag{5.2.21}$$

Next, we integrate this equation over a portion $\Delta\mathscr{B}$ of $\mathscr{B}$ bounded by initial and final constant t surfaces $\mathscr{S}_{\mathrm{i}}$ and $\mathscr{S}_{\mathrm{f}}$, as depicted in Figure 5.2. On the resulting LHS we apply

Stokes' theorem, and on the first term on the RHS we use the Gauss-Codazzi identity: $\mathcal{D}_a\Pi^{ab} = n_a\gamma^b{}_c G^{ac}$. Thus, using the notation for tensor projections in certain directions introduced in Sec. ID for ease of readability (e.g., $G_{ab}n^a\psi^b = G_{\boldsymbol{n}\boldsymbol{\psi}}$ and similarly for other contractions), we obtain:

$$\int\limits_{\mathscr{S}_\mathrm{f}-\mathscr{S}_\mathrm{i}} \boldsymbol{\epsilon}_{\mathscr{S}}\,\Pi_{\tilde{\boldsymbol{u}}\boldsymbol{\psi}} = -\int\limits_{\Delta\mathscr{B}} \boldsymbol{\epsilon}_{\mathscr{B}}\,\left(G_{\boldsymbol{n}\boldsymbol{\psi}} + \Pi^{ab}\mathcal{D}_a\psi_b\right)\,, \tag{5.2.22}$$

where $\boldsymbol{\epsilon}_{\mathscr{S}}$ denotes the volume form on the constant time closed two-surfaces $\mathscr{S}$, and we have used the notation: $\int_{\mathscr{S}_\mathrm{f}-\mathscr{S}_\mathrm{i}}(\cdot) = \int_{\mathscr{S}_\mathrm{f}}(\cdot) - \int_{\mathscr{S}_\mathrm{i}}(\cdot)$. We also remind the reader that $\tilde{\boldsymbol{u}}$ represents the unit normal to each constant time closed two-surface, which in general need not coincide with the quasilocal observers' four velocity $\boldsymbol{u}$ but is related to it by a Lorentz transformation, Eq. (5.2.4); see also Figure 5.2.

We stress that so far, Eq. (5.2.22) is a purely geometrical identity, completely general for any Lorentzian manifold $\mathscr{M}$; in other words, thus far we have said nothing about physics.

Now, to give this identity physical meaning, we invoke the definition of the Brown-York tensor in Eq. (5.2.16) (giving the boundary extrinsic geometry its meaning as stress-energy-momentum) as well as the Einstein equation [Eq. (5.2.1)], giving the spacetime curvature its meaning as the gravitational field. With these, Eq. (5.2.22) turns into:

$$\int\limits_{\mathscr{S}_\mathrm{f}-\mathscr{S}_\mathrm{i}} \boldsymbol{\epsilon}_{\mathscr{S}}\,\tilde{\gamma}\,(\tau_{\boldsymbol{u}\boldsymbol{\psi}} + \tau_{\boldsymbol{v}\boldsymbol{\psi}}) = \int\limits_{\Delta\mathscr{B}} \boldsymbol{\epsilon}_{\mathscr{B}}\,\left(T_{\boldsymbol{n}\boldsymbol{\psi}} - \tau^{ab}\mathcal{D}_{(a}\psi_{b)}\right)\,. \tag{5.2.23}$$

On the LHS we have inserted the relation $\tilde{\boldsymbol{u}} = \tilde{\gamma}(\boldsymbol{u}+\boldsymbol{v})$, with v^a representing the spatial two-velocity of fiducial observers that are at rest with respect to $\mathscr{S}$ (the hypersurface-orthogonal four-velocity of which is $\tilde{\boldsymbol{u}}$) as measured by our congruence of quasilocal observers (the four-velocity of which is $\boldsymbol{u}$), and $\tilde{\gamma} = 1/\sqrt{1-\boldsymbol{v}\cdot\boldsymbol{v}}$ is the Lorentz factor.

Observe that Eq. (5.2.23) expresses the change of some component of the quasilocal stress-energy-momentum tensor integrated over two different $t = const.$ closed two-surfaces $\mathscr{S}$ as a flux through the worldtube boundary $\Delta\mathscr{B}$ between them. The identification of the different components of $\boldsymbol{\tau}$ as the various components of the total energy-momentum of the system thus leads to the understanding of Eq. (5.2.23) as a general conservation law for the system contained inside of $\Delta\mathscr{B}$. Thus, depending on our particular choice of $\boldsymbol{\psi} \in T\mathscr{B}$, Eq. (5.2.23) will represent a conservation law for the total energy, momentum, or angular momentum of this system [Epp, P. L. McGrath, et al. 2013].

Let us now assume that $(\mathscr{B};\boldsymbol{u})$ is a rigid quasilocal frame. If we choose $\boldsymbol{\psi} = \boldsymbol{u}$, then Eq. (5.2.23) becomes the energy conservation law:

$$\int\limits_{\mathscr{S}_\mathrm{f}-\mathscr{S}_\mathrm{i}} \boldsymbol{\epsilon}_{\mathscr{S}}\,\tilde{\gamma}\,(\mathcal{E} - \mathcal{P}_{\boldsymbol{v}}) = \int\limits_{\Delta\mathscr{B}} \boldsymbol{\epsilon}_{\mathscr{B}}\,(T_{\boldsymbol{n}\boldsymbol{u}} - \boldsymbol{\alpha}\cdot\boldsymbol{\mathcal{P}})\,, \tag{5.2.24}$$

where $\alpha^a = \sigma^{ab} a_b$ is the $\mathscr{H}$ projection of the acceleration of the quasilocal observers, defined by $a^a = \nabla_{\boldsymbol{u}} u^a$.

Now suppose, on the other hand, that we instead choose $\boldsymbol{\psi} = -\boldsymbol{\phi}$ where $\boldsymbol{\phi} \in T\mathscr{H}$ is orthogonal to $\boldsymbol{u}$ (with the minus sign introduced for convenience), and represents a stationary conformal Killing vector field with respect to $\boldsymbol{\sigma}$. This means that $\boldsymbol{\phi}$ is chosen such that it satisfies the conformal Killing equation, $\mathcal{L}_{\boldsymbol{\phi}} \boldsymbol{\sigma} = (\boldsymbol{D} \cdot \boldsymbol{\phi}) \boldsymbol{\sigma}$, with $\mathcal{L}$ the Lie derivative and $\boldsymbol{D}$ the derivative on $\mathscr{H}$ (compatible with $\boldsymbol{\sigma}$). A set of six such conformal Killing vectors always exist (three for translations and three for rotations, respectively generating the action of boosts and rotations of the Lorentz group on the two-sphere) [Epp, P. L. McGrath, et al. 2013]. Then, Eq. (5.2.23) becomes the (respectively, linear and angular) momentum conservation law:

$$\int_{\mathscr{S}_{\mathrm{f}} - \mathscr{S}_{\mathrm{i}}} \boldsymbol{\epsilon}_{\mathscr{S}} \, \tilde{\gamma} \left(\mathcal{P}_{\boldsymbol{\phi}} + \mathcal{S}_{\boldsymbol{v}\boldsymbol{\phi}} \right) = - \int_{\Delta\mathscr{B}} \boldsymbol{\epsilon}_{\mathscr{B}} \left(T_{\boldsymbol{n}\boldsymbol{\phi}} + \mathcal{E} \alpha_{\boldsymbol{\phi}} + 2\nu \epsilon_{ab} \mathcal{P}^a \phi^b + \mathrm{P} \boldsymbol{D} \cdot \boldsymbol{\phi} \right) , \qquad (5.2.25)$$

where $\nu = \frac{1}{2} \epsilon^{ab}_{\mathscr{H}} \mathcal{D}_a u_b$ is the twist of the congruence (with $\epsilon^{\mathscr{H}}_{ab} = \epsilon^{\mathscr{M}}_{abcd} u^c n^d$ the induced volume form on $\mathscr{H}$), and $\mathrm{P} = \frac{1}{2} \boldsymbol{\sigma} : \boldsymbol{\mathcal{S}}$ is the quasilocal pressure (force per unit length) between the worldlines of $\mathscr{B}$. We remark that the latter can be shown to satisfy the very useful general identity (which we will expediently invoke in our later calculations):

$$\mathcal{E} - 2\mathrm{P} = \frac{2}{\kappa} a_{\boldsymbol{n}} \, . \qquad (5.2.26)$$

An analysis of the gravitational self-force problem should consider the conservation law in Eq. (5.2.25) for *linear momentum*. Thus, we will use the fact, described in greater detail in the appendix of this chapter (Section 5.6), that the conformal Killing vector $\boldsymbol{\phi} \in \mathscr{H}$ for linear momentum admits a multipole decomposition of the following form:

$$\phi^{\mathrm{i}} = \frac{1}{r} D^{\mathrm{i}} \left(\Phi^I r_I + \Phi^{IJ} r_I r_J + \cdots \right) \qquad (5.2.27)$$

$$= \frac{1}{r} \left(\Phi^I \mathfrak{B}^{\mathrm{i}}_I + 2 \Phi^{IJ} \mathfrak{B}^{\mathrm{i}}_I r_J + \cdots \right) , \qquad (5.2.28)$$

with the dots indicating higher harmonics. Here, r is the area radius of the quasilocal frame (such that $\mathscr{B}$ is a constant r hypersurface in $\mathscr{M}$), r^I denotes the the standard direction cosines of a radial unit vector in $\mathbb{R}^3$ and $\mathfrak{B}^{\mathrm{i}}_I = \partial^{\mathrm{i}} r_I$ are the boost generators on the two-sphere. See this chapter's appendix (Section 5.6) for a detailed discussion regarding conformal Killing vectors and the two-sphere. In spherical coordinates $\{\theta, \phi\}$, we have $r^I = (\sin\theta \cos\phi, \sin\theta \sin\phi, \cos\theta)$. Thus Eq. (5.2.27) gives us a decomposition of $\boldsymbol{\phi}$ in terms of multipole moments, with the $\ell = 1$ coefficients Φ^I representing vectors in $\mathbb{R}^3$ in the direction of which we are considering the conservation law.

5.3. General derivation of the gravitational self-force from quasilocal conservation laws

In this section, we will show how the GSF is a general consequence of the momentum conservation law in Eq. (5.2.25) for any system which is sufficiently localized. By that, we mean something very simple: taking the $r \to 0$ limit of a quasilocal frame around the moving object which is treated as "small", i.e. as a formal perturbation about some background. No further assumptions are for the moment needed. In particular, we do not even need to enter into the precise details of how to specify the perturbation family for this problem; that will be left to the following section, where we will carefully define and work with the family of perturbed spacetimes typically employed for applications of the GSF. For now, we proceed to show that the first-order perturbation of the momentum conservation law in Eq. (5.2.25) always contains the GSF, and that it dominates the dynamics for localized systems.

Let $\{(\mathscr{B}_{(\lambda)};\boldsymbol{u}_{(\lambda)})\}_{\lambda\geq 0}$ be an arbitrary one-parameter family of quasilocal frames (defined as in Section 5.2) each of which is embedded, respectively, in the corresponding element of the family of perturbed spacetimes $\{(\mathscr{M}_{(\lambda)},\boldsymbol{g}_{(\lambda)},\boldsymbol{\nabla}_{(\lambda)})\}_{\lambda\geq 0}$ described in the previous subsection. Consider the general geometrical identity (5.2.22) in $\mathscr{M}_{(\lambda)}$, $\forall\lambda\geq 0$:

$$\int\limits_{\mathscr{S}_{\mathrm{f}}^{(\lambda)}-\mathscr{S}_{\mathrm{i}}^{(\lambda)}} \boldsymbol{\epsilon}_{\mathscr{S}_{(\lambda)}}\,\Pi^{(\lambda)}_{\tilde{\boldsymbol{u}}_{(\lambda)}\boldsymbol{\psi}_{(\lambda)}} = -\int\limits_{\Delta\mathscr{B}_{(\lambda)}} \boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}}\,\left(G^{(\lambda)}_{\boldsymbol{n}_{(\lambda)}\boldsymbol{\psi}_{(\lambda)}} + \Pi^{ab}_{(\lambda)}\mathcal{D}^{(\lambda)}_a\psi^{(\lambda)}_b\right)\,. \tag{5.3.1}$$

For $\lambda = 0$ this gives us our conservation laws in the background, and for any $\lambda > 0$, those in the corresponding perturbed spacetime. It is the latter that we are interested in, but since we do not know how to do calculations in $\mathscr{M}_{(\lambda)}$ $\forall\lambda > 0$, we have to work with Eq. (5.3.1) transported to $\mathring{\mathscr{M}}$. This is easily achieved by using the fact that for any diffeomorphism $f:\mathscr{U}\to\mathscr{V}$ between two (oriented) smooth n-dimensional manifolds $\mathscr{U}$ and $\mathscr{V}$ and any (compactly supported) n-form $\boldsymbol{\omega}$ in $\mathscr{V}$, we have that $\int_{\mathscr{V}}\boldsymbol{\omega} = \int_{\mathscr{U}} f^*\boldsymbol{\omega}$. Applying this to the LHS and RHS of Eq. (5.3.1) respectively, we simply get

$$\int\limits_{\varphi^{-1}_{(\lambda)}(\mathscr{S}_{\mathrm{f}}^{(\lambda)})-\varphi^{-1}_{(\lambda)}(\mathscr{S}_{\mathrm{i}}^{(\lambda)})} \left(\varphi^*_{(\lambda)}\boldsymbol{\epsilon}_{\mathscr{S}_{(\lambda)}}\right)\,\varphi^*_{(\lambda)}\Pi^{(\lambda)}_{\tilde{\boldsymbol{u}}_{(\lambda)}\boldsymbol{\psi}_{(\lambda)}} = \int\limits_{\varphi^{-1}_{(\lambda)}(\Delta\mathscr{B}_{(\lambda)})} \left(\varphi^*_{(\lambda)}\boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}}\right)\,\varphi^*_{(\lambda)}\left(G^{(\lambda)}_{\boldsymbol{n}_{(\lambda)}\boldsymbol{\psi}_{(\lambda)}} + \Pi^{ab}_{(\lambda)}\mathcal{D}^{(\lambda)}_a\psi^{(\lambda)}_b\right)\,. \tag{5.3.2}$$

Denoting $\mathscr{S} = \varphi^{-1}_{(\lambda)}(\mathscr{S}_{(\lambda)}) \subset \mathring{\mathscr{M}}$ as the inverse image of a constant time two-surface and similarly $\mathscr{B} = \varphi^{-1}_{(\lambda)}(\mathscr{B}_{(\lambda)}) \subset \mathring{\mathscr{M}}$ as the inverse image of the worldtube boundary (quasilocal frame) in the background manifold, and using the fact that the tensor transport commutes with contractions, the above can simply be written in the notation we have

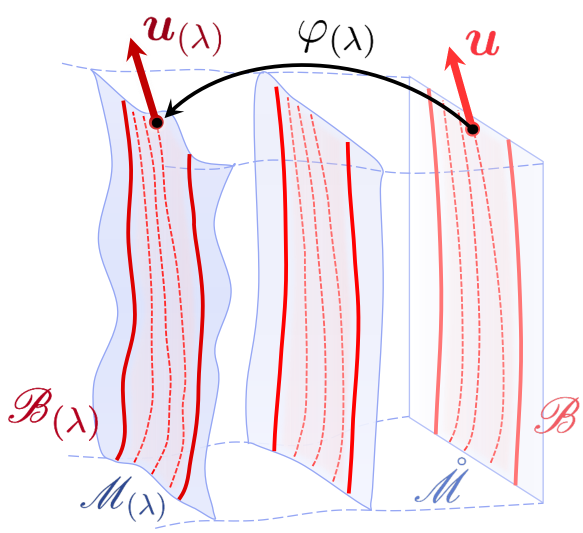

FIGURE 5.3. Representation of a one-parameter family of QFs $\{(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})\}_{\lambda \geq 0}$ embedded correspondingly in a family of spacetimes $\{\mathscr{M}_{(\lambda)}\}_{\lambda \geq 0}$.

---

established as

$$\int_{\mathscr{S}_\mathrm{f} - \mathscr{S}_\mathrm{i}} \left(\varphi^*_{(\lambda)} \boldsymbol{\epsilon}_{\mathscr{S}_{(\lambda)}}\right) \Pi_{\tilde{\boldsymbol{u}}\boldsymbol{\psi}} = \int_{\Delta\mathscr{B}} \left(\varphi^*_{(\lambda)} \boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}}\right) \left(G_{\boldsymbol{n}\boldsymbol{\psi}} + \Pi^{ab} \mathcal{D}_a \psi_b\right) . \tag{5.3.3}$$

So far we have been completely general. Now, let us restrict our attention to the momentum conservation law ($\boldsymbol{\psi} = -\boldsymbol{\phi} \in T\mathscr{H}$) given by Eq. (5.3.3), and let us assume that we do not have any matter on $\Delta\mathscr{B}$ (hence, by the Einstein equation, $G_{\boldsymbol{n}\boldsymbol{\phi}}|_{\Delta\mathscr{B}} = \kappa T_{\boldsymbol{n}\boldsymbol{\phi}}|_{\Delta\mathscr{B}} = 0$), or even simply that any matter if present there is subdominant to the linear perturbation, *i.e.* $\boldsymbol{T}|_{\Delta\mathscr{B}} = \mathcal{O}(\lambda^2)$. The LHS then expresses the change in momentum of the system (inside $\Delta\mathscr{B}_{(\lambda)}$ in the perturbed spacetime) between some initial and final time slices; for notational ease, we will simply denote this by $\Delta\mathtt{p}^{(\phi)}$. (Note that we prefer to use typewriter font for the total quasilocal momentum, so as to avoid any confusion with matter four-momentum defined in the typical way from T_{ab} and traditionally labelled by P^a, as *e.g.* in Eq. 5.1.1.) Then, inserting also the definition of the Brown-York tensor [Eq. (5.2.16)] on the RHS and replacing $\boldsymbol{\mathcal{D}}$ with $\boldsymbol{\nabla}$ since it does not affect the contractions, Eq. (5.3.3) becomes:

$$\Delta\mathtt{p}^{(\phi)} = \int_{\Delta\mathscr{B}} \left(\varphi^*_{(\lambda)} \boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}}\right) \tau^{ab} \nabla_a \phi_b \,. \tag{5.3.4}$$

We claim, and will now demonstrate, that the $\mathcal{O}(\lambda)$ part of this always contains the GSF.

Let us consider Eq. (5.3.4) term by term. First we have the transport—in this case, the pullback—under $\varphi_{(\lambda)}$ of the volume form of $\mathscr{B}_{(\lambda)}$. Now, we know that the pullback under a diffeomorphism of the volume form of a manifold is, in general, not simply the volume form of the inverse image of that manifold under the diffeomorphism. However, it is always true (see, *e.g.*, Chapter 7 of [Abraham et al. 2001]) that they are proportional, with the proportionality given by a smooth function called the Jacobian determinant and usually denoted by J. That is, in our case we have $\varphi^*_{(\lambda)}\boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}} = J\boldsymbol{\epsilon}_{\mathscr{B}}$, with $J \in C^\infty(\mathscr{B})$. In particular, this function is given by $J(p) = \det(T_p\varphi_{(\lambda)})$, $\forall p \in \mathscr{B}$, where $T_p\varphi_{(\lambda)} = (\varphi_{(\lambda)})_* : T_p\mathscr{B} \to T_{\varphi_{(\lambda)}}\mathscr{B}_{(\lambda)}$ is the pushforward, and the determinant is computed with respect to the volume forms $\boldsymbol{\epsilon}_{\mathscr{B}}(p)$ on $T_p\mathscr{B}$ and $\boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}}(\varphi_{(\lambda)}(p))$ on $T_{\varphi_{(\lambda)}(p)}\mathscr{B}_{(\lambda)}$. Now, it is clear that we have $J = 1 + \mathcal{O}(\lambda)$, as $\varphi_{(0)}$ is simply the identity map. Therefore, we have

$$\varphi^*_{(\lambda)}\boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}} = \left(1 + \mathcal{O}\left(\lambda\right)\right)\boldsymbol{\epsilon}_{\mathscr{B}}\,. \tag{5.3.5}$$

As for the other terms in the integrand of Eq. (5.3.4), we simply have

$$\tau^{ab} = \mathring{\tau}^{ab} + \lambda\delta\tau^{ab} + \mathcal{O}(\lambda^2)\,, \tag{5.3.6}$$

$$\nabla_a\phi_b = \mathring{\nabla}_a\phi_b + \lambda\delta\left(\nabla_a\phi_b\right) + \mathcal{O}(\lambda^2)\,. \tag{5.3.7}$$

Hence we can see that there will be three contributions to the $\mathcal{O}(\lambda)$ RHS of Eq. (5.3.4). Respectively, from Eqs. (5.3.5)-(5.3.7), these are the $\mathcal{O}(\lambda)$ parts of: the volume form pullback, which may not be easy to compute in practice; the Brown-York tensor $\boldsymbol{\tau}$, which may be computed from its definition [Eq. (5.2.16)]; and the derivative of the conformal Killing vector $\boldsymbol{\phi}$, which may be readily carried out and, as we will presently show, always contains the GSF. Thus we denote this contribution to the $\mathcal{O}(\lambda)$ part of $\Delta\mathrm{p}^{(\boldsymbol{\phi})}$ as $\Delta\mathrm{p}^{(\boldsymbol{\phi})}_{\text{self}}$,

$$\Delta\mathrm{p}^{(\boldsymbol{\phi})}_{\text{self}} = \lambda\int\limits_{\Delta\mathscr{B}}\boldsymbol{\epsilon}_{\mathscr{B}}\,\mathring{\tau}^{ab}\delta\left(\nabla_a\phi_b\right)\,. \tag{5.3.8}$$

Before we proceed to compute this, we remark that the rest of the $\mathcal{O}(\lambda)$ part of $\Delta\mathrm{p}^{(\boldsymbol{\phi})}$, *i.e.* the contributions due to the $\delta(\varphi^*_{(\lambda)}\boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}})$ and $\delta\tau^{ab}$ terms may simply be regarded as encoding the freedom we have at our disposal (and thus far have in no way constrained) in choosing the map $\varphi^{\mathbf{X}}_{(\lambda)}$ (the gauge) and the congruence of observers making up our worldtube boundary (the quasilocal frame). This can be seen through a simple DoF counting argument: we have four DoFs in choosing the gauge vector $\mathbf{X}|_{\mathring{\mathscr{M}}} \in T\mathring{\mathscr{M}}$ defining $\varphi^{\mathbf{X}}_{(\lambda)}$, and three DoFs in choosing the quasilocal frame $(\mathscr{B}_{(\lambda)};\boldsymbol{u}_{(\lambda)})$ in $\mathscr{M}_{(\lambda)}$ (or equivalently $(\mathscr{B};\boldsymbol{u}) = (\varphi^{-1}_{(\lambda)}(\mathscr{B}_{(\lambda)});\varphi^*_{(\lambda)}\boldsymbol{u}_{(\lambda)})$ in $\mathring{\mathscr{M}}$). So in total, we have seven DoFs available for us to fix. Thus we could, in principle, use them all up to impose $0 = \delta(\varphi^*_{(\lambda)}\boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}})$ and $0 = \delta\tau^{ab}$. The former is one equation and the latter is six equations—seven equations in total.

Now we proceed with the computation of Eq. (5.3.8). In particular, let us consider the series expansion of Eq. (5.3.8) in the areal radius r of $\mathscr{B}$. This can be defined for any time slice by $r = (\frac{1}{4\pi}\int_{\mathscr{S}} \boldsymbol{\epsilon}_{\mathscr{S}})^{1/2}$, such that a constant r slice of $\mathring{\mathscr{M}}$ defines $\mathscr{B}$ (and $\boldsymbol{n} = M\mathring{\boldsymbol{\nabla}}r$ for some positive function M on $\mathscr{B}$). It has been shown [Epp, Mann, et al. 2012] that the Brown-York tensor has, in general, the following expansion in r:

$$\mathring{\tau}^{ab} = \mathring{u}^a\mathring{u}^b\mathcal{E}_{\text{vac}} - \mathring{\sigma}^{ab}\text{P}_{\text{vac}} + \mathcal{O}(r)\,, \tag{5.3.9}$$

where

$$\mathcal{E}_{\text{vac}} = -\frac{2}{\kappa r}\,, \tag{5.3.10}$$

$$\text{P}_{\text{vac}} = -\frac{1}{\kappa r}\,, \tag{5.3.11}$$

are called the vacuum energy and vacuum pressure respectively. Some remarks regarding these are warranted before we move on. In particular, these are terms which have sometimes been argued to play the role of "subtraction terms" (to be removed from the quasilocal energy-momentum tensor); see *e.g.* [J. D. Brown, Lau, et al. 2002]. From this point of view, the definition of the Brown-York tensor [Eq. (5.2.16)] may be regarded as carrying a certain amount of freedom, inasmuch as any freedom may be assumed to exist to define a "reference" action S_0 to be subtracted from the total (gravitational plus matter) action $S_{\text{G+M}}$ in the variational principle discussed in Subsection 5.2.3. Such a subtraction of a "reference" action, while common practice in gravitational physics, has the sole function of shifting the numerical value of the action (such that, ultimately, the numerical value of the Hamiltonian constructed from the modified action $S_{\text{G+M}} - S_0$ may be interpreted as the ADM energy). However, this essentially amounts to a presumption that we are free to pick the zero of the energy—in other words, that the vacuum energy may be freely subtracted away without affecting the physics. Though we refrain from entering into much further detail here, it has been shown that these vacuum terms, Eqs. (5.3.10)-(5.3.11), are in fact crucial for our conservation laws to yield physically reasonable answers and to make mathematical sense—evidencing that the vacuum energy/pressure should be taken seriously as having physically real significance. We will now lend further credibility to this by showing that they are precisely the energy (and pressure) associated with the momentum flux that are typically interpreted as the GSF. Actually, we argue in this paper that the term implicating the vacuum energy yields the standard form of the GSF, and the vacuum pressure term is novel in our analysis.

Now that we have an expansion [Eq. (5.3.9)] of $\mathring{\boldsymbol{\tau}}$ in r, let us consider the $\delta(\boldsymbol{\nabla\phi})$ term. We see that

$$\delta\left(\nabla_a\phi_b\right) = \delta\left(\mathring{\nabla}_a\phi_b - C^d{}_{ab}\phi_d\right) = -\delta C^c{}_{ab}\phi_c\,. \tag{5.3.12}$$

Collecting all of our results so far—inserting Eqs. (5.3.9)-(5.3.12) into Eq. (5.3.8)—we thus get:

$$\Delta \mathrm{p}_{\text{self}}^{(\phi)} = \lambda \frac{2}{\kappa} \int_{\Delta\mathscr{B}} \boldsymbol{\epsilon}_{\mathscr{B}}\, \frac{1}{r} \left(\mathring{u}^a \mathring{u}^b - 2\mathring{\sigma}^{ab} \right) \delta C^c{}_{ab} \phi_c + \mathcal{O}\left(r\right) . \tag{5.3.13}$$

Let us now look at the contractions in the integrand. For the first (energy) term, inserting the connection coefficient (3.2.10), we have by direct computation:

$$\mathring{u}^a \mathring{u}^b \delta C^c{}_{ab} \phi_c = \mathring{g}^{cd} \left(\mathring{\nabla}_a h_{bd} - \frac{1}{2} \mathring{\nabla}_d h_{ab} \right) \mathring{u}^a \mathring{u}^b \phi_c \tag{5.3.14}$$

$$= - F^c[\boldsymbol{h}; \mathring{\boldsymbol{u}}] \phi_c \, , \tag{5.3.15}$$

where the functional $\boldsymbol{F}$ is precisely the GSF four-vector functional defined in the introduction [Eq. 1.5.5], and to write the final equality we have used the orthogonality property $\phi_{\mathring{\boldsymbol{u}}} = 0$. Thus we see that this is indeed the term that yields the GSF. For the second (pressure) term in Eq. (5.3.13), we similarly obtain by direct computation:

$$\mathring{\sigma}^{ab} \delta C^c{}_{ab} \phi_c = 2 \wp^c[\boldsymbol{h}; \mathring{\boldsymbol{\sigma}}] \phi_c \, , \tag{5.3.16}$$

where in expressing the RHS, it is convenient to define a general functional of two $(0,2)$-tensors similar to the GSF functional:

$$\wp^c[\boldsymbol{H}; \boldsymbol{S}] = \frac{1}{2} \mathring{g}^{cd} \left(\mathring{\nabla}_a H_{bd} - \frac{1}{2} \mathring{\nabla}_d H_{ab} \right) S^{ab} \, . \tag{5.3.17}$$

We call this novel term the *gravitational self-pressure force*.

Now we can collect all of the above and insert them into (5.3.13). Before writing down the result, it is convenient to define a total functional $\boldsymbol{\mathcal{F}}$ as the sum of $\boldsymbol{F}$ and $\wp$,

$$\mathcal{F}^a[\boldsymbol{h}; \mathring{\boldsymbol{u}}] = F^a[\boldsymbol{h}; \mathring{\boldsymbol{u}}] + \wp^a[\boldsymbol{h}; \mathring{\boldsymbol{\sigma}}] \, . \tag{5.3.18}$$

We refer to this as the *extended GSF functional*. Note that for $\boldsymbol{\mathcal{F}}$ we write only the functional dependence on $\boldsymbol{h}$ and $\mathring{\boldsymbol{u}}$ since the two-metric $\mathring{\boldsymbol{\sigma}}$ is determined uniquely by $\mathring{\boldsymbol{u}}$. With this, and setting the perturbation parameter to unity, Eq. (5.3.13) becomes:

$$\boxed{\Delta \mathrm{p}_{\text{self}}^{(\phi)} = -\frac{1}{4\pi} \int_{\Delta\mathscr{B}} \boldsymbol{\epsilon}_{\mathscr{B}}\, \frac{1}{r} \boldsymbol{\phi} \cdot \boldsymbol{\mathcal{F}}[\boldsymbol{h}; \mathring{\boldsymbol{u}}] + \mathcal{O}\left(r\right) .} \tag{5.3.19}$$

This is to be compared with Gralla's formula [Gralla 2011] discussed in the introduction, Eq. (1.5.7). While the equivalence thereto is immediately suggestive based on the general form of our result, we have to do a bit more work to show that indeed Eqn. (5.3.19), both on the LHS and the RHS, recovers—though in general will, evidently at least from our novel gravitational self-pressure force, also have extra terms added to—Eq. (1.5.7). We leave this task to the following section, the purpose of which is to consider in detail the application of our conservation law formulation to a concrete example of a perturbative family of spacetimes defined for a self-force analysis, namely the Gralla-Wald family.

Concordantly, we emphasize that the result above [Eq. (5.3.19)] holds for any family of perturbed manifolds $\{\mathscr{M}_{(\lambda)}\}_{\lambda\geq 0}$ and is completely independent of the internal description of our system, *i.e.* the worldtube interior $\mathrm{int}(\mathscr{B}_{(\lambda)}) \subset \mathscr{M}_{(\lambda)}$. In other words, what we have just demonstrated—provided only that one accepts a quasilocal notion of energy-momentum—is that the (generalized) GSF is a completely generic perturbative effect in GR for "localized" systems: it arises as a linear order contribution of any spacetime perturbation to the momentum flux of a system in the limit where its areal radius is small.

This view of the self-force may cast fresh conceptual light on the old and seemingly arcane problem of deciphering its physical origin and meaning. In particular, recall the common view that the GSF is caused by the backreaction of the "mass" of a small object upon its own motion. Yet what we have seen here is that it is actually the vacuum "mass", or vacuum energy that is responsible for the GSF. We may still regard the effect as a "backreaction," in the sense that it is the boundary metric perturbations of the system—the $\boldsymbol{h}$ on $\mathscr{B}$—which determine its momentum flux, but the point is that this flux is inexorably present and given by Eq. (5.3.19) regardless of where exactly this $\boldsymbol{h}$ is coming from. Presumably, the dominant part of $\boldsymbol{h}$ would arise from the system itself—if we further assume that the system itself is indeed what is being treated perturbatively by the family $\{\mathscr{M}_{(\lambda)}\}_{\lambda\geq 0}$, as is the case with typical self-force analyses—but in principle $\boldsymbol{h}$ can comprise absolutely any perturbations, *i.e.* its physical origin doesn't even have to be from inside the system.

In this way, we may regard the GSF as a completely geometrical, purely general-relativistic backreaction of the mass (and pressure) of the spacetime vacuum —*not* of the object—upon the motion of a localized system (*i.e.* its momentum flux). This point of view frees us from having to invoke such potentially ambiguous notions as "mass ratios" (in a two-body system for example), let alone "Coulombian m/r fields," to make basic sense of self-force effects. They simply—and always—happen from the interaction of the vacuum with any boundary perturbation, and are dominant if that boundary is not too far out.

5.4. Application to the Gralla-Wald approach to the gravitational self-force

In this section we will consider in detail the application of our ideas to a particular approach to the self-force: that is to say, a particular specification of $\{(\mathscr{M}_{(\lambda)}, \boldsymbol{g}_{(\lambda)})\}$ via a few additional assumptions aimed at encoding the notion of a "small" object being "scaled down" to zero "size" and "mass" as $\lambda \to 0$. In other words, we now identify the perturbation (which has up to this point been treated completely abstractly) defined by $\{(\mathscr{M}_{(\lambda)}, \boldsymbol{g}_{(\lambda)})\}$ as actually being that caused by the presence of the "small" object: that could mean regular matter (in particular, a compact object such as a neutron star) or a black hole.

The assumptions (on $\{\boldsymbol{g}_{(\lambda)}\}$) that we choose to work with here are those of the approach of Gralla and Wald [Gralla and Wald 2008]. Certainly, the application of our perturbed quasilocal conservation laws could just as well be carried out in the context of any other self-force analysis—such as, *e.g.*, the self-consistent approximation of Pound [Pound 2010] (the mathematical correspondence of which to the Gralla-Wald approach has, in any case, been shown in [Pound 2015a]).

Our motivation for starting with the Gralla-Wald approach in particular is two-fold. On the one hand, it furnishes a mathematically rigorous and physically clear picture (which we show in Fig. 5.4)—arguably more so than any other available GSF treatment—of what it means to "scale down" a small object to zero "size" and "mass" (or, equivalently, of perturbing any spacetime by the presence of an object with small "size" and "mass"—we will be more precise momentarily). On the other hand, it is within this approach that the formula for the GSF has been obtained (in [Gralla 2011]) as a closed two-surface (small two-sphere) integral around the object (in lieu of evaluating the GSF at a spacetime point identified as the location of the object), in the form of the Gralla "angle averaging" formula [Eq. (1.5.7)]—which our extended GSF formula (5.3.19) will recover.

In Subsection 5.4.1, we provide an overview of the assumptions and consequences of the Gralla-Wald approach to the GSF. Afterwards, in Subsection 5.4.2, we describe the general embedding of rigid quasilocal frames in the Gralla-Wald family of spacetimes, and then in Subsection 5.4.3 we describe their detailed construction in the background spacetime in this family. Having established this, we then proceed to recover the equation of motion given by Gralla's formula, up to our novel self-pressure term. In particular, we carry out the calculation with two choices of rigid quasilocal frames ("frames of reference"): first, inertially with the "point particle" approximation of the moving object in the background in Subsection 5.4.4, and second, inertially with the object itself in the perturbed spacetime in Subsection 5.4.5.

5.4.1. The Gralla-Wald approach to the GSF. The basic idea of Gralla and Wald [Gralla and Wald 2008] for defining a family $\{(\mathscr{M}_{(\lambda)}, \boldsymbol{g}_{(\lambda)})\}_{\lambda\geq 0}$ such that $\lambda > 0$ represents the inclusion of perturbations generated by a "small" object is the following one. One begins by imposing certain smoothness conditions on $\{\boldsymbol{g}_{(\lambda)}\}_{\lambda\geq 0}$ corresponding to the existence of certain limits of each $\boldsymbol{g}_{(\lambda)}$. In particular, two limits are sought corresponding intuitively to two "limiting views" of the system: first, a view from "far away" from which the "motion" of the (extended but localized) object reduces to a worldline; second, a view from "close by" the object from which the rest of the universe (and in particular, the MBH it might be orbiting as in an EMRI) looks "pushed away" to infinity. A third requirement must be added to this, namely that both of these limiting pictures nonetheless coexist in the same spacetime, *i.e.* the two limits are smoothly related (or, in other words, there is no pathological behaviour when taking these limits along different directions). While in principle this may sound rather technical, one can actually motivate each of these conditions with very sensible physical arguments as we shall momentarily elaborate further upon. From them, Gralla and Wald have shown [Gralla and Wald 2008] that it is possible to derive a number of consequences, including geodesic motion in the background at zeroth order and the MiSaTaQuWa equation [Mino et al. 1997; Quinn and Wald 1997] for the GSF at first order in λ.

Let us now be more precise. Let $\{(\mathscr{M}_{(\lambda)}, \boldsymbol{g}_{(\lambda)})\}_{\lambda\geq 0}$ be a perturbative one-parameter family of spacetimes as in the previous section. We assume that $\{\boldsymbol{g}_{(\lambda)}\}_{\lambda\geq 0}$ satisfies the following conditions, depicted visually in Fig. 5.4:

(i) Existence of an "ordinary limit": There exist coordinates $\{x^\alpha\}$ in $\mathscr{M}_{(\lambda)}$ such that $g^{(\lambda)}_{\beta\gamma}(x^\alpha)$ is jointly smooth in (λ, x^α) for $r > C\lambda$ where $C > 0$ is a constant and $r = (x_i x^i)^{1/2}$. For all $\lambda \geq 0$ and $r > C\lambda$, $\boldsymbol{g}_{(\lambda)}$ is a vacuum solution of the Einstein equation. Furthermore, $\mathring{g}_{\beta\gamma}(x^\alpha)$ is smooth in x^α including at $r = 0$, and the curve $\mathring{\mathscr{C}} = \{r = 0\} \subset \mathring{\mathscr{M}}$ is timelike.

(ii) Existence of a "scaled limit": For all t_0, define the "scaled coordinates" $\{\bar{x}^\alpha\} = \{\bar{t}, \bar{x}^i\}$ by $\bar{t} = (t - t_0)/\lambda$ and $\bar{x}^i = x^i/\lambda$. Then the "scaled metric" $\bar{g}^{(\lambda)}_{\bar{\beta}\bar{\gamma}}(t_0; \bar{x}^\alpha) = \lambda^{-2} g^{(\lambda)}_{\bar{\beta}\bar{\gamma}}(t_0; \bar{x}^\alpha)$ is jointly smooth in $(\lambda, t_0; \bar{x}^\alpha)$ for $\bar{r} = r/\lambda > C$.

(iii) Uniformity condition: Define $A = r$, $B = \lambda/r$ and $n^i = x^i/r$. Then each $g^{(\lambda)}_{\beta\gamma}(x^\alpha)$ is jointly smooth in (A, B, n^i, t).

Mathematically, the first two conditions respectively ensure the existence of an appropriate Taylor expansion (in r and λ) of the metric in a "far zone" (on length scales comparable with the mass of the MBH in an EMRI, $r \sim M$) and a "near zone" (on length scales comparable with the mass of the object, $r \sim m$) . Meanwhile, the third is simply a consistency requirement ensuring the existence of a "buffer zone" ($m \ll r \ll M$) where both expansions are valid. (This idea is in many ways similar to the method of "matched asymptotic expansions" [Mino et al. 1997]).

From a physical point of view, what is happening in the first ("ordinary") limit is that the body is shrinking down to a worldline $\mathring{\mathscr{C}}$ with its "mass" (understood as defining the perturbation) going to zero at least as fast as its radius. (As we increase the perturbative parameter λ from zero, the radius is not allowed to grow faster than linearly with λ; viewed conversely, this condition ensures that the object does not collapse to a black hole if it was not one already before reaching the point particle limit.) In the second ("scaled") limit, the object is shrinking down to zero size in an asymptotically self-similar manner (its mass is proportional to its size, and its "shape" is not changing). Finally, the uniformity condition

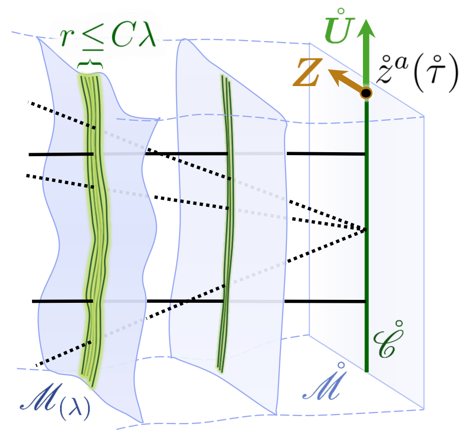

FIGURE 5.4. Representation of the Gralla-Wald family of spacetimes $\{\mathscr{M}_{(\lambda)}\}_{\lambda \geq 0}$. (This is an adaptation of Fig. 1 of [Gralla and Wald 2008].) The lined green region that "fills in" $\mathscr{M}_{(\lambda)}$ for $r \leq C\lambda$ is the "small" object which "scales down" to zero "size" and "mass" in the background $\mathring{\mathscr{M}}$. The solid black lines represent taking the "ordinary limit" (the "far away" view where the motion appears reduced to a worldline) and the dashed black lines the "scaled limit" (the "close by" view where the rest of the universe appears "pushed away" to infinity). The worldline $\mathring{\mathscr{C}}$, which can be proven to be a geodesic, is parametrized by $\mathring{z}^a(\mathring{\tau})$ and has four-velocity $\mathring{\boldsymbol{U}}$. The "deviation" vector $\boldsymbol{Z}$ on $\mathring{\mathscr{C}}$ is used for formulating the first-order correction to the motion.

ensures that there are no "bumps of curvature" in the one-parameter family. (Essentially, this guarantees that there are no inconsistencies in evaluating the limits along different directions.)

From these assumptions alone, Gralla [Gralla and Wald 2008] and Wald are able to derive the following consequences:

(a) Background motion: The worldline $\mathring{\mathscr{C}}$ is a geodesic in $\mathring{\mathscr{M}}$; writing its parametrization in terms of proper time $\mathring{\tau}$ as $\mathring{\mathscr{C}} = \{\mathring{z}^a(\mathring{\tau})\}_{\mathring{\tau}\in\mathbb{R}}$ and denoting its four-velocity by $\mathring{U}^a = \mathrm{d}\mathring{z}^a(\mathring{\tau})/\mathrm{d}\mathring{\tau}$, this means that

$$\mathring{\nabla}_{\mathring{\boldsymbol{U}}}\mathring{\boldsymbol{U}} = 0\,. \tag{5.4.1}$$

(b) Background "scaled" metric: $\mathring{\bar{\boldsymbol{g}}}$ is stationary and asymptotically flat.

(c) First-order field equation: At $\mathcal{O}(\lambda)$, the Einstein equation is sourced by the matter energy-momentum tensor of a "point particle" $\boldsymbol{T}^{\mathrm{PP}}$ supported on $\mathring{\mathscr{C}}$, *i.e.* the field equation is

$$\delta G_{ab}\left[\boldsymbol{h}\right] = \kappa T^{\mathrm{PP}}_{ab} \tag{5.4.2}$$

where

$$T^{\mathrm{PP}}_{ab} = m \int_{\mathring{\mathscr{C}}} \mathrm{d}\mathring{\tau}\, \mathring{U}_a\left(\mathring{\tau}\right) \mathring{U}_b\left(\mathring{\tau}\right) \delta_4\left(x^c - z^c\left(\mathring{\tau}\right)\right) . \tag{5.4.3}$$

(This is the same as (1.3.2) discussed in the introduction to the thesis.) Here, m is a constant along $\mathring{\mathscr{C}}$ and is interpreted as representing the "mass" of the object—or, more precisely, the mass of the point particle which approximates the object in the background. (This is a subtle point that should be kept in mind, and which will be better elucidated in our analysis further on.)

(d) First-order equation of motion: At $\mathcal{O}(\lambda)$, the correction to the motion in the Lorenz gauge—corresponding to the choice of a certain gauge vector $\mathsf{L} \in T\mathscr{N}$ defined by the condition

$$\mathring{\nabla}^b(h^{\mathsf{L}}_{ab} - \frac{1}{2}h^{\mathsf{L}}\mathring{g}_{ab}) = 0\,, \tag{5.4.4}$$

where $h = \mathrm{tr}(\boldsymbol{h})$—is given by the MiSaTaQuWa equation [Mino et al. 1997; Quinn and Wald 1997],

$$\mathring{\nabla}_{\mathring{\boldsymbol{U}}}\mathring{\nabla}_{\mathring{\boldsymbol{U}}} Z^a = -\mathring{E}_b{}^a Z^b + F^a[\boldsymbol{h}^{\mathrm{tail}}; \mathring{\boldsymbol{U}}]\,, \tag{5.4.5}$$

where $\mathring{E}_b{}^a = \mathring{R}_{cbd}{}^a \mathring{U}^c \mathring{U}^d$ is the electric part of the Weyl tensor and $\boldsymbol{h}^{\mathrm{tail}}$ is a "tail" integral of the retarded Green's functions of $\boldsymbol{h}$. The above is an equation for a four-vector $\boldsymbol{Z}$ called the "deviation" vector; the LHS is the acceleration associated therewith and the RHS is a geodesic deviation term plus the GSF. This deviation vector is defined on $\mathring{\mathscr{C}}$ and represents the first-order correction needed to move off $\mathring{\mathscr{C}}$ and onto the worldline representing the "center of mass" of the perturbed spacetime, defined as in the Hamiltonian analysis of Regge and Teitelboim [Regge and Teitelboim 1974].

Let us make a few comments on these results, specifically concerning *(a)* and *(c)*. On the one hand, it is quite remarkable that geodesic motion can be recovered as a consequence[5] of this analysis—*i.e.* without having to posit it as an assumption—just from smoothness properties (existence of appropriate limits) of our family of metrics $\{\boldsymbol{g}_{(\lambda)}\}$; and on the other, this analysis offers sensible meaning to the usual "delta function cartoon" (ubiquitous in essentially all self-force analyses) of the matter energy-momentum tensor describing the object in the background spacetime. The point is that the description of the object is completely arbitrary inside the region that is not covered by the smoothness conditions of the family $\{\boldsymbol{g}_{(\lambda)}\}$, *i.e.* for $r \leq C\lambda$ when $\lambda > 0$. (Indeed, this region

[5] See again footnote 2 and the references mentioned therein for more on this topic.

can be "filled in" even with exotic matter, *e.g.* failing to satisfy the dominant energy condition, or a naked singularity, as long as a well-posed initial value formulation exists.) Regardless of what this description is, the smoothness conditions essentially ensure that its "reduction" to $\mathring{\mathscr{M}}$ (or, more precisely, the transport of any effect thereof with respect to the family $\{\boldsymbol{g}_{(\lambda)}\}$) simply becomes that of a point particle sourcing the field equation at $\mathcal{O}(\lambda)$. In this way, the background "point particle cartoon" is justified as the simplest possible idealization of a "small" object.

What we are going to do, essentially, is to accept consequences *(a)-(c)* (in fact, we will not even explicitly need *(b)*), the proofs of which do not rely upon any further limiting conditions such as a restriction of the perturbative gauge, and to obtain, using our perturbed momentum conservation law, a more general version of the EoM, *i.e.* consequence *(d)*. For the latter, Gralla and Wald [Gralla and Wald 2008] instead rely on the typical but laborious Hadamard expansion techniques of DeWitt and Brehme [DeWitt and Brehme 1960], wherein the "mass dipole moment" of the object is set to zero. It is possible [Regge and Teitelboim 1974] to have such a notion in a well-defined Hamiltonian sense by virtue of *(b)*. While mathematically rigorous and conducive to obtaining the correct known form of the MiSaTaQuWa equation, their derivation and final result suffer not only from the limitation of having to fix the perturbative gauge, but also from the (as we shall see, potentially avoidable) technical complexity of arriving at the final answer—including the evaluation of $\boldsymbol{h}^{\text{tail}}$ (or otherwise taking recourse to a regularization procedure).

The link between this approach and our conservation law derivation of the EoM which we are about to carry out is established by the work of Gralla [Gralla 2011], who discovered that Eq. (5.4.5) can be equivalently written as:

$$\mathring{\nabla}_{\mathring{\boldsymbol{U}}}\mathring{\nabla}_{\mathring{\boldsymbol{U}}} Z^a = -\mathring{E}_b{}^a Z^b + \frac{1}{4\pi}\lim_{r\to 0}\int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2}\, F^a[\mathring{\boldsymbol{U}}, \boldsymbol{h}]\,. \tag{5.4.6}$$

Here, the GSF term $\boldsymbol{F}[\boldsymbol{h}^{\text{tail}}; \mathring{\boldsymbol{U}}]$ in the MiSaTaQuWa equation [Eq. (5.4.5)] is substituted by an integral expression—an average over the angles—of $\boldsymbol{F}$. In particular (as, strictly speaking, one cannot define integrals of vectors as such), this is evaluated by using the exponential map based on $\mathring{\mathscr{C}}$ to associate a flat metric, in terms of which the integration is performed over a two-sphere of radius r, $\mathbb{S}^2_r$, with $\boldsymbol{\epsilon}_{\mathbb{S}^2}$ denoting the volume form of $\mathbb{S}^2$.

Observe that, here, the functional dependence of $\boldsymbol{F}$ is on $\boldsymbol{h}$ itself (and not on $\boldsymbol{h}^{\text{tail}}$ or any sort of regularized $\boldsymbol{h}$) and for this reason is referred to as the "bare" GSF. Moreover, this formula is actually valid in a wider class of gauges than just the Lorenz gauge: in particular, it holds in what are referred to as "parity-regular" gauges [Gralla 2011]. We refrain from entering here into the technical details of exactly how such gauges are defined, except to say that the eponymous "parity condition" that they need to satisfy has its ultimate origin in the Hamiltonian analysis of Regge and Teitleboim [Regge and Teitelboim

1974] and is imposed so as to make certain Hamiltonian definitions—and in particular for Gralla's analysis [Gralla 2011], the Hamiltonian "center of mass"—well defined. These, however, are *not* limitations of our quasilocal formalism, where we know how to define energy-momentum notions more generally than any Hamiltonian approach. Thus, in our result, there will be *no restriction* on the perturbative gauge. This may constitute a great advantage, as the "parity-regular" gauge class—though an improvement from being limited to the Lorenz gauge in formulating the EoM—still excludes the perturbative gauges most widely employed for black hole perturbation theory, and therefore in practical EMRI calculations (*e.g.* the Regge-Wheeler gauge in Schwarzschild and the radiation gauge in Kerr).

We proceed to apply our quasilocal analysis to the Gralla-Wald family of spacetimes, beginning with a general setup in this family of rigid quasilocal frames.

5.4.2. General setup of rigid quasilocal frames in the Gralla-Wald family. Let $(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})$ be a quasilocal frame in $(\mathscr{M}_{(\lambda)}, \boldsymbol{g}_{(\lambda)}, \boldsymbol{\nabla}_{(\lambda)})$, for any $\lambda > 0$, constructed just as described in Section 5.2: with unit four-velocity $\boldsymbol{u}_{(\lambda)}$, unit normal $\boldsymbol{n}_{(\lambda)}$, induced metric $\boldsymbol{\gamma}_{(\lambda)}$ and so on. Using the fact that the tensor transport is linear and commutes with tensor products, we can compute the transport (in the five-dimensional "stacked" manifold $\mathcal{N} = \mathcal{M}_{(\lambda)} \times \mathbb{R}$ used in our perturbative setup, as in Chapter 3) of any geometrical quantity of interest to the background. For example,

$$\gamma_{ab} = \varphi^*_{(\lambda)} \gamma^{(\lambda)}_{ab} \tag{5.4.7}$$

$$= \varphi^*_{(\lambda)} (g^{(\lambda)}_{ab} - n^{(\lambda)}_a n^{(\lambda)}_b) \tag{5.4.8}$$

$$= g_{ab} - n_a n_b \tag{5.4.9}$$

$$= \mathring{\gamma}_{ab} + \lambda \delta\gamma_{ab} + \mathcal{O}\left(\lambda^2\right) , \tag{5.4.10}$$

where

$$\mathring{\gamma}_{ab} = \mathring{g}_{ab} - \mathring{n}_a \mathring{n}_b , \tag{5.4.11}$$

$$\delta\gamma_{ab} = h_{ab} - 2\mathring{n}_{(a} \delta n_{b)} . \tag{5.4.12}$$

Similarly,

$$\sigma_{ab} = \mathring{\sigma}_{ab} + \lambda \delta\sigma_{ab} + \mathcal{O}(\lambda^2) , \tag{5.4.13}$$

where

$$\mathring{\sigma}_{ab} = \mathring{\gamma}_{ab} + \mathring{u}_a \mathring{u}_b , \tag{5.4.14}$$

$$\delta\sigma_{ab} = \delta\gamma_{ab} + 2\mathring{u}_{(a} \delta u_{b)} . \tag{5.4.15}$$

Now let us assume that $(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})$ is a rigid quasilocal frame, meaning that the congruence defining it has a vanishing symmetrized strain rate tensor in $\mathscr{M}_{(\lambda)}$,

$$\theta^{(\lambda)}_{(ab)} = 0\,. \tag{5.4.16}$$

Let $\mathscr{B} = \varphi^{-1}_{(\lambda)}(\mathscr{B}_{(\lambda)})$ be the inverse image of $\mathscr{B}_{(\lambda)}$ in the background $\mathring{\mathscr{M}}$, with $\boldsymbol{u} = \varphi^{*}_{(\lambda)}\boldsymbol{u}_{(\lambda)} = \mathring{\boldsymbol{u}} + \lambda\delta\boldsymbol{u} + \mathcal{O}(\lambda^2)$ giving the transport of the quasilocal observers' four-velocity, $\boldsymbol{n} = \mathring{\boldsymbol{n}} + \lambda\delta\boldsymbol{n} + \mathcal{O}(\lambda^2)$ the unit normal and so on. In other words, $(\mathscr{B}; \boldsymbol{u})$ is the background mapping of the "perturbed" congruence $(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})$, and so will itself constitute a congruence (in the background), *i.e.* a quasilocal frame defined by a two-parameter family of worldlines with unit four-velocity $\boldsymbol{u}$ in $\mathring{\mathscr{M}}$.

However, although $(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})$ is a rigid quasilocal frame in $\mathscr{M}_{(\lambda)}$, $(\mathscr{B}; \boldsymbol{u})$ is *not* in general a rigid quasilocal frame in $\mathring{\mathscr{M}}$ (with respect to the background metric $\mathring{\boldsymbol{g}}$). One can see this easily as follows. Let $\boldsymbol{\vartheta} \in \mathscr{T}^0{}_2(\mathring{\mathscr{M}})$ be the strain rate tensor of $(\mathscr{B}; \boldsymbol{u})$, so that it is given by

$$\vartheta_{ab} = \sigma_{ca}\sigma_{bd}\mathring{\nabla}^c u^d\,. \tag{5.4.17}$$

The RHS is an series in λ, owing to the fact that $\boldsymbol{u}$ (and therefore $\boldsymbol{\sigma}$, the two-metric on the space $\mathscr{H}$ orthogonal to $\boldsymbol{u}$ in $\mathscr{B}$) are transported from a "perturbed" congruence in $\mathscr{M}_{(\lambda)}$. Upon expansion we obtain

$$\vartheta_{ab} = \mathring{\vartheta}_{ab} + \lambda\delta\vartheta_{ab} + \mathcal{O}\left(\lambda^2\right)\,, \tag{5.4.18}$$

where

$$\mathring{\vartheta}_{ab} = \mathring{\sigma}_{c(a}\mathring{\sigma}_{b)d}\mathring{\nabla}^c\mathring{u}^d \tag{5.4.19}$$

is just the strain rate tensor of the "background" congruence—*i.e.* the congruence defined by $\mathring{\boldsymbol{u}}$—and

$$\delta\vartheta_{ab} = 2\mathring{\sigma}^{(c}{}_{(a}\delta\sigma^{d)}{}_{b)}\mathring{\nabla}_c\mathring{u}_d + \mathring{\sigma}^c{}_{(a}\mathring{\sigma}_{b)d}\mathring{\nabla}_c\delta u^d \tag{5.4.20}$$

is the first-order piece in λ. Note that we are abusing our established notation slightly in writing Eq. (5.4.18), as there exists no $\boldsymbol{\vartheta}_{(\lambda)}$ in $\mathscr{M}_{(\lambda)}$ the transport (to $\mathring{\mathscr{M}}$) of which yields such a series expansion; instead $\boldsymbol{\vartheta}$ is defined directly on $\mathring{\mathscr{M}}$ (relative to the metric $\mathring{\boldsymbol{g}}$) as the strain rate tensor of a conguence with four-velocity $\boldsymbol{u}$—which itself contains the expansion in λ.

Now let us compute the transport of the rigidity condition on $(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})$ [Eq. (5.4.16)] to $\mathring{\mathscr{M}}$: we have

$$0 = \varphi^{*}_{(\lambda)}\theta^{(\lambda)}_{(ab)} \tag{5.4.21}$$

$$= \varphi^{*}_{(\lambda)}(\sigma^{(\lambda)}_{c(a}\sigma^{(\lambda)}_{b)d}\nabla^c_{(\lambda)}u^d_{(\lambda)}) \tag{5.4.22}$$

$$= \sigma_{c(a}\sigma_{b)d}\nabla^c u^d \tag{5.4.23}$$

$$= \mathring{\theta}_{(ab)} + \lambda\delta\theta_{(ab)} + \mathcal{O}(\lambda^2)\,, \tag{5.4.24}$$

where

$$\mathring{\theta}_{(ab)} = \mathring{\vartheta}_{(ab)} \,, \tag{5.4.25}$$

$$\delta\theta_{(ab)} = \delta\vartheta_{(ab)} + \mathring{\sigma}^{c}{}_{(a}\mathring{\sigma}_{b)d}\delta C^{d}{}_{ce}\mathring{u}^{e} \,. \tag{5.4.26}$$

Since $0 = \theta^{(\lambda)}_{(ab)}$ identically in $\mathscr{M}_{(\lambda)}$ (as we demand that $(\mathscr{B}_{(\lambda)};\boldsymbol{u}_{(\lambda)})$ is a rigid quasilocal frame), Eq. (5.4.24) must vanish order by order in λ. That implies, in particular, that the zeroth-order congruence (defined by $\mathring{\boldsymbol{u}}$) is a rigid quasilocal frame, and that the symmetrized strain rate tensor of the background-mapped perturbed congruence (defined by $\boldsymbol{u}$) is given by

$$\vartheta_{(ab)} = -\lambda\mathring{\sigma}^{c}{}_{(a}\mathring{\sigma}_{b)d}\delta C^{d}{}_{ce}\mathring{u}^{e} + \mathcal{O}\left(\lambda^2\right) \,. \tag{5.4.27}$$

This tells us that the deviation from rigidity of $(\mathscr{B};\boldsymbol{u})$ in $\mathring{\mathscr{M}}$ occurs only at $\mathcal{O}(\lambda)$ (and, in particular, is caused by the same perturbed connection coefficient term that is responsible for the GSF). In other words, we can treat $(\mathscr{B};\boldsymbol{u})$ as a rigid quasilocal frame at zeroth order. This zeroth order congruence actually makes up a different worldtube boundary $\mathring{\mathscr{B}}$ in $\mathring{\mathscr{M}}$, *i.e.* defined by a congruence with four-velocity $\mathring{\boldsymbol{u}}$. Clearly, for a rigid quasilocal frame with a small areal radius r constructed around a worldline $\mathscr{G}$ in $\mathring{\mathscr{M}}$ with four-velocity $\boldsymbol{U}_{\mathscr{G}}$, we would simply have $\mathring{\boldsymbol{u}} = \boldsymbol{U}_{\mathscr{G}}$ (where the RHS is understood to be transported off $\mathscr{G}$ and onto $\mathring{\mathscr{B}}$ via the exponential map), and $\mathring{\boldsymbol{\sigma}} = r^2\mathfrak{S}$, *i.e.* it is the metric of $\mathbb{S}^2_r$. This is the most trivial possible rigid quasilocal frame: at any instant of time, a two-sphere worth of quasilocal observers moving with the same four-velocity as is the point at its center (parametrizing the given worldline).

At first order, the equation $0 = \delta\theta_{(ab)}$ can be regarded as the constraint on the linear perturbations ($\delta\boldsymbol{u}$) in the motion of the quasilocal observers in terms of the metric perturbations guaranteeing that the perturbed congruence is rigid in the perturbed spacetime. (So presumably, going to n-th order in λ would yield equations for every term up to the n-th order piece of the motion of the quasilocal observers, $\delta^n\boldsymbol{u}$.)

Now recall the momentum conservation law for rigid quasilocal frames, Eq. (5.2.25). This holds for $(\mathscr{B}_{(\lambda)};\boldsymbol{u}_{(\lambda)})$ in $\mathscr{M}_{(\lambda)}$. Just as we did in the previous section with the general conservation law, we can use $\varphi_{(\lambda)}$ to turn this into an equation in $\mathring{\mathscr{M}}$:

$$\Delta \mathrm{p}^{(\boldsymbol{\phi})} = -\int_{\Delta\mathscr{B}} \varphi^{*}_{(\lambda)}\boldsymbol{\epsilon}_{\mathscr{B}_{(\lambda)}} \left(\mathcal{E}\alpha_{\boldsymbol{\phi}} + 2\nu\epsilon_{ab}\mathcal{P}^{a}\phi^{b} + \mathrm{P}\boldsymbol{D}\cdot\boldsymbol{\phi}\right) \,. \tag{5.4.28}$$

Let us now further assume that we can ignore the Jacobian determinant (discussed in the previous section) as well as the shift $\boldsymbol{v}$ of the quasilocal observers (relative to constant time surfaces). Then, dividing the above equation by Δt, where t represents the adapted "time" coordinate on $\mathscr{B}$, and taking the $\Delta t \to 0$ limit, we get the time rate of change of the momentum,

$$\dot{\mathrm{p}}^{(\phi)} = -\int\limits_{\mathscr{S}} \boldsymbol{\epsilon}_{\mathscr{S}} N \tilde{\gamma} \left(\mathcal{E} \alpha_{\boldsymbol{\phi}} + 2\nu \epsilon_{ab} \mathcal{P}^a \phi^b + \mathrm{P} \boldsymbol{D} \cdot \boldsymbol{\phi} \right) . \tag{5.4.29}$$

where $\dot{\mathrm{p}}^{(\phi)} = \mathrm{d}\mathrm{p}^{(\phi)}/\mathrm{d}t$, and we must keep in mind that the derivative is with respect to the adapted time on (the inverse image on the background of) our congruence.

5.4.3. Detailed construction of background rigid quasilocal frames. Let $\mathscr{G}$ be any timelike worldline in $\mathring{\mathscr{M}}$. Any background metric $\mathring{\boldsymbol{g}}$ on $\mathring{\mathscr{M}}$ in a neighborhood of $\mathscr{G}$ admits an expression in Fermi normal coordinates [Misner et al. 1973; Poisson et al. 2011], which we label by $\{X^a\} = \{T = X^0, X^I\}_{I=1}^3$, as a power series in the areal radius. Denoting by $A_K(T)$ and $W_K(T)$ the proper acceleration and proper rate of rotation of the spatial axes (triad) along $\mathscr{G}$ (as functions of the proper time T along $\mathscr{G}$), respectively, this is given by:

$$\mathring{g}_{00} = -\left(1 + A_K X^K\right)^2 + R^2 W_K W_L P^{KL} - \mathring{R}_{0K0L} X^K X^L + \mathcal{O}\left(R^3\right) , \tag{5.4.30}$$

$$\mathring{g}_{0J} = \epsilon_{JKL} W^K X^L - \frac{2}{3} \mathring{R}_{0KJL} X^K X^L + \mathcal{O}\left(R^3\right) , \tag{5.4.31}$$

$$\mathring{g}_{IJ} = \delta_{IJ} - \tfrac{1}{3} \mathring{R}_{IKJL} X^K X^L + \mathcal{O}\left(R^3\right) , \tag{5.4.32}$$

where $R^2 = \delta_{IJ} X^I X^J$ is the square of the radius in these coordinates (not the square of the Ricci scalar) and $P^{KL} = \delta^{KL} - X^K X^L / R^2$ projects vectors perpendicular to the radial direction X^I/R. Here we have to remember that the Riemann tensor $\mathring{\boldsymbol{R}}$ (along with $\boldsymbol{A}$ and $\boldsymbol{W}$) are understood to be evaluated on $\mathscr{G}$.

For all cases that we will be interested in, we will ignore the possibility of rotation so we set $W_I = 0$ from now on.

Let us now assume that our background rigid quasilocal frame $(\mathring{\mathscr{B}}; \mathring{\boldsymbol{u}})$ is constructed around $\mathscr{G}$: that is to say, into this coordinate system there is embedded a two-parameter family of worldlines representing a topological two-sphere worth of observers, *i.e.* a fibrated timelike worldtube $\mathring{\mathscr{B}}$ surrounding $\mathscr{G}$. This may be conveniently described, as detailed in Subsection 5.2.2, by defining a new set of coordinates $\{x^a\} = \{t, r, x^{\mathrm{i}}\}_{\mathrm{i}=1}^2$ given simply by the adapted coordinates $\{t, x^{\mathrm{i}}\}_{\mathrm{i}=1}^2$ on $\mathring{\mathscr{B}}$ supplemented with a radial coordinate r. Then denoting $\{x^{\mathrm{i}}\} = \{\theta, \phi\}$ we introduce, as done in previous calculations with rigid quasilocal frames in Fermi normal coordinates [Epp, Mann, et al. 2012], the following coordinate transformation:

$$T\left(t, r, \theta, \phi\right) = t + \mathcal{O}\left(r^2, \mathcal{R}\right) , \tag{5.4.33}$$

$$X^I\left(t, r, \theta, \phi\right) = r r^I\left(\theta, \phi\right) + \mathcal{O}\left(r^2, \mathcal{R}\right) , \tag{5.4.34}$$

where

$$r^I(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \tag{5.4.35}$$

are the standard direction cosines of a radial unit vector in spherical coordinates in $\mathbb{R}^3$, and $\mathcal{R}$ here represents the order of the perturbations of the quasilocal frame away from the round two-sphere due to the background curvature effects. In particular, for rigid quasilocal frames, we know that this is in fact simply the order of the Riemann tensor on $\mathscr{G}$, *i.e.* $\mathring{\boldsymbol{R}} = \mathcal{O}(\mathcal{R})$. Thus, one may ultimately desire to take $\mathcal{O}(\mathcal{R})$ effects into account for a full calculation, but for the moment—since, in principle, this $\mathcal{R}$ is unrelated to λ and we can assume it to be subdominant thereto—we simply omit them. Thus we can simply take $\mathring{\mathscr{S}} = \mathbb{S}^2_r$, and we can assume that there is no shift, so that $\tilde{\gamma} = 1$.

Applying the coordinate transformation in Eqs. (5.4.33)-(5.4.34) to the background metric given by Eqs. (5.4.30)-(5.4.32) with $\boldsymbol{W} = 0$, and then using all of the definitions that we have established so far, it is possible to obtain by direct computation all of the quantities appearing in the integrand of the conservation law [Eq. (5.4.29)] as series in r. We display the results only up to leading order in r, including the possibility of setting $\boldsymbol{A} = 0$:

$$\mathring{N} = 1 + rA_I r^I + \frac{1}{2} r^2 \mathring{E}_{IJ} r^I r^J + \mathcal{O}\left(r^3\right) , \tag{5.4.36}$$

$$\begin{aligned}\mathring{\mathcal{E}} &= \mathcal{E}_{\text{vac}} + \mathcal{O}\left(r\right) \\ &= -\frac{2}{\kappa r} + \mathcal{O}\left(r\right) ,\end{aligned} \tag{5.4.37}$$

$$\mathring{\alpha}_{\mathrm{i}} = rA_I \mathfrak{B}^I_{\mathrm{i}} + r^2 \left(\mathring{E}_{IJ} - A_I A_J\right) \mathfrak{B}^I_{\mathrm{i}} r^J + \mathcal{O}\left(r^3\right) , \tag{5.4.38}$$

$$\mathring{\nu} = -r\mathring{B}_{IJ} r^I r^J + \mathcal{O}\left(r^2\right) , \tag{5.4.39}$$

$$\mathring{\mathcal{P}}_{\mathrm{i}} = -\frac{1}{\kappa} r^2 \mathring{B}_{IJ} \mathfrak{R}^I_{\mathrm{i}} r^J + \mathcal{O}\left(r^3\right) , \tag{5.4.40}$$

$$\begin{aligned}\mathring{\mathrm{P}} &= \mathrm{P}_{\text{vac}} - \frac{1}{\kappa} A_I r^I + \mathcal{O}\left(r\right) \\ &= -\frac{1}{\kappa r} - \frac{1}{\kappa} A_I r^I + \mathcal{O}\left(r\right) .\end{aligned} \tag{5.4.41}$$

Here, $\mathring{E}_{IJ} = \mathring{C}_{0I0J}|_{\mathscr{G}}$ and $\mathring{B}_{IJ} = \frac{1}{2}\epsilon_I{}^{KL}\mathring{C}_{0JKL}|_{\mathscr{G}}$ are respectively the electric and magnetic parts of the Weyl tensor evaluated on on $\mathscr{G}$. Also, $\mathfrak{B}^I_{\mathrm{i}} = \partial_{\mathrm{i}} r^I$ and $\mathfrak{R}^I_{\mathrm{i}} = \epsilon^{\mathbb{S}^2}_{\mathrm{i}}{}^{\mathrm{j}}\mathfrak{B}^I_{\mathrm{j}}$ are respectively the boost and rotation generators of $\mathbb{S}^2$. See this chapter's appendix (Section 5.6) for more technical details on this. We remind the reader that $\mathcal{E}_{\text{vac}}$ and P_{vac} are respectively the vacuum energy and pressure, Eqs. (5.3.10)-(5.3.11) respectively.

The way to proceed is now clear: we expand Eq. (5.4.29) as a series in λ,

$$\dot{\mathrm{p}}^{(\phi)} = (\dot{\mathrm{p}}^{(\phi)})_{(0)} + \lambda\delta\dot{\mathrm{p}}^{(\phi)} + \mathcal{O}\left(\lambda^2\right) , \tag{5.4.42}$$

using the zeroth-order parts of the various terms written above. We need only to specify the worldline $\mathscr{G}$ in $\mathring{\mathscr{M}}$ about which we are carrying out the Fermi normal coordinate expansion (in r). We will consider two cases: $\mathscr{G} = \mathring{\mathscr{C}}$ (the geodesic, such that $\mathscr{B}$ is "inertial" with the point particle in $\mathring{\mathscr{M}}$) and $\mathscr{G} = \mathscr{C}$ (an accelerated worldline such that $\mathscr{B}_{(\lambda)}$ is "inertial" with the object in $\mathscr{M}_{(\lambda)}$, *i.e.* it is defined by a constant $r > C\lambda$ in $\mathscr{M}_{(\lambda)}$). These will give us equivalent descriptions of the dynamics of the system, from two different "points of view", or (quasilocal) frames of reference.

Before entering into the calculations, we can simplify things further by remarking that the zeroth order expansions in Eqs. (5.4.36)-(5.4.41) will always make the twist (ν) term in the conservation law [Eq. (5.4.29)] appear at $\mathcal{O}(r)$ or higher, both in $(\dot{\mathrm{p}}^{(\phi)})_{(0)}$ and $\delta\dot{\mathrm{p}}^{(\phi)}$, regardless of our choice of $\mathscr{G}$. Hence we can safely ignore it, as we are interested (at least for this work) only in the part of the conservation law which is zeroth-order in r. Thus we simply work with

$$\dot{\mathrm{p}}^{(\phi)} = -\int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2}\, r^2 N \left(\mathcal{E}\alpha_{\boldsymbol{\phi}} + \mathrm{P}\boldsymbol{D}\cdot\boldsymbol{\phi}\right) . \tag{5.4.43}$$

Into this, we furthermore have to insert the multipole expansion of the conformal Killing vector $\boldsymbol{\phi}$ given by Eq. (5.2.27). We correspondingly write

$$\dot{\mathrm{p}}^{(\phi)} = \sum_{\ell\in\mathbb{N}} \dot{\mathrm{p}}^{(\phi_\ell)} , \tag{5.4.44}$$

such that for any $\ell \in \mathbb{N}$, we have

$$\dot{\mathrm{p}}^{(\phi_\ell)} = -\Phi^{I_1\cdots I_\ell}\int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2}\, rN\left(\mathcal{E}\alpha_{\mathrm{i}} + \mathrm{P}D_{\mathrm{i}}\right) D^{\mathrm{i}}\left(\prod_{n=1}^{\ell} r_{I_n}\right) . \tag{5.4.45}$$

Explicitly, the first two terms are

$$\dot{\mathrm{p}}^{(\phi_{\ell=1})} = -\,\Phi^{I}\int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2}\, rN\left(\mathcal{E}\alpha_{\mathrm{i}} + \mathrm{P}D_{\mathrm{i}}\right)\mathfrak{B}^{\mathrm{i}}_{I} , \tag{5.4.46}$$

$$\dot{\mathrm{p}}^{(\phi_{\ell=2})} = -\,2\Phi^{IJ}\int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2}\, rN\left(\mathcal{E}\alpha_{\mathrm{i}} + \mathrm{P}D_{\mathrm{i}}\right)\left(\mathfrak{B}^{\mathrm{i}}_{I} r_J\right) . \tag{5.4.47}$$

5.4.4. Equation of motion inertial with the background "point particle". Let $\mathscr{G} = \mathring{\mathscr{C}}$. Then $\boldsymbol{A} = 0$. We will take this to be the case for the rest of this subsection—corresponding, as discussed, to a rigid quasilocal frame the inverse image in the background of which is inertial with the "point particle" approximation of the moving object in the background spacetime. This situation is displayed visually in Fig. 5.5.

Let us first compute the zeroth-order (in λ) part of $\dot{\mathrm{p}}^{(\phi)}$. Inserting (5.4.36)-(5.4.41) into the zeroth-order part of (5.4.46)-(5.4.47), and making use of the various properties in

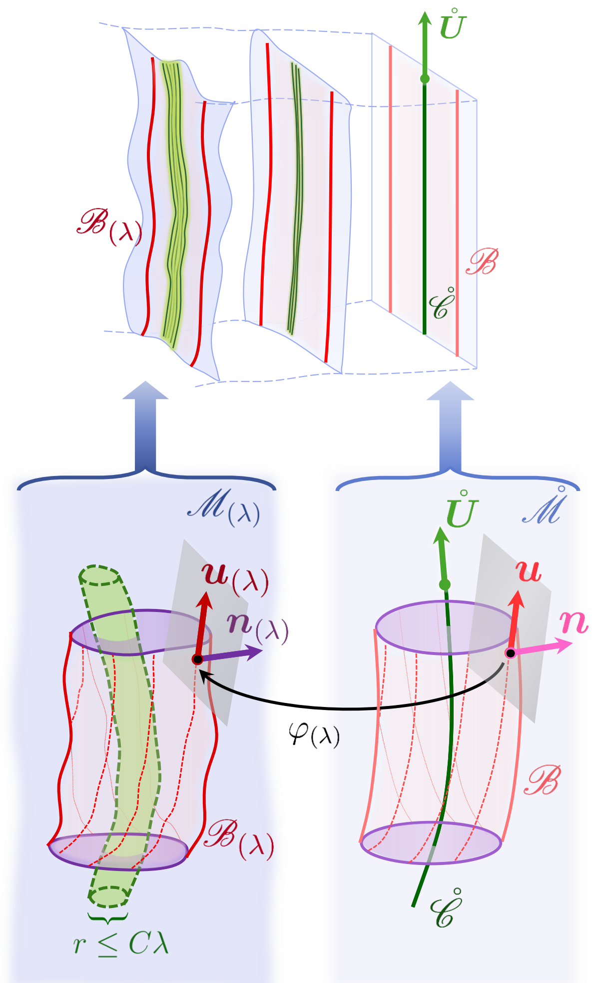

FIGURE 5.5. A family of rigid quasilocal frames $\{(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})\}$ embedded in the Gralla-Wald family of spacetimes $\{\mathscr{M}_{(\lambda)}\}$ such that the inverse image of any such perturbed quasilocal frame in the background is inertial with the "point particle" approximation of the moving object, *i.e.* is centered on the geodesic $\mathring{\mathscr{C}}$.

Section 5.6, we find by direct computation:

$$\left(\dot{\mathrm{p}}^{(\phi_{\ell=1})}\right)_{(0)} = \mathcal{O}\left(r^2\right) , \tag{5.4.48}$$

$$\left(\dot{\mathrm{p}}^{(\phi_{\ell=2})}\right)_{(0)} = \mathcal{O}\left(r^2\right) . \tag{5.4.49}$$

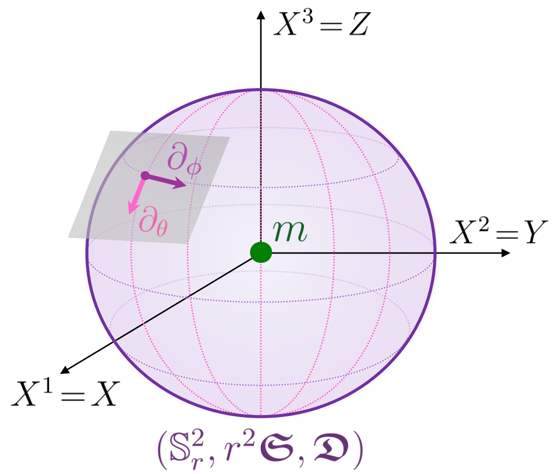

FIGURE 5.6. An instantaneous rigid quasilocal frame $(\mathbb{S}^2_r, r^2\mathfrak{S}, \mathfrak{D})$ (where $\mathfrak{S}$ and $\mathfrak{D}$ respectively are the metric and derivative compatible with the unit two-sphere) inertial with the background "point particle". This means that the latter is located at the center of our Fermi normal coordinate system.

We provide the steps of the calculation in Appendix B of [Oltean, Epp, Sopuerta, et al. 2019].

Let us now compute the $\mathcal{O}(\lambda)$, $\ell = 1$ part of $\dot{\mathrm{p}}^{(\phi)}$, *i.e.* the $\mathcal{O}(\lambda)$ part of Eq. (5.4.46) which as usual we denote by $\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}$. One can see that this will involve contributions from five $\mathcal{O}(\lambda)$ terms, respectively containing δN, $\delta\mathcal{E}$, $\delta\boldsymbol{\alpha}$, $\delta\mathrm{P}$ and $\delta\boldsymbol{D}$. For convenience, we will use the notation $(\dot{\mathrm{p}}^{(\phi_\ell)}_{(Q)})_{(n)}$ to indicate the term of $\delta^n(\dot{\mathrm{p}}^{(\phi_\ell)})$ that is linear in Q, for any ℓ, n. Thus we write

$$\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})} = \sum_{Q\in\{\delta N,\delta\mathcal{E},\delta\boldsymbol{\alpha},\delta\mathrm{P},\delta\boldsymbol{D}\}} \delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(Q)} \,. \tag{5.4.50}$$

All of the computational steps are again in Appendix B of [Oltean, Epp, Sopuerta, et al. 2019]. We find:

$$\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta N)} = -\frac{2}{\kappa}\Phi_I \int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2}\, \delta N r^I + \mathcal{O}\left(r^2\right) \,. \tag{5.4.51}$$

If δN does not vary significantly over $\mathbb{S}^2_r$, the $\mathcal{O}(r^0)$ part of the above would be negligible owing to the fact that $\int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2}\, r^I = 0$.

Next, let us consider the $\delta\mathcal{E}$ and $\delta\mathrm{P}$ terms. For this, we find it useful to depict the instantaneous quasilocal frame $(\mathbb{S}^2_r, r^2\mathfrak{S}, \mathfrak{D})$ embedded in a constant-time three-slice of $\mathring{\mathscr{M}}$ in Fig. 5.6.

The $\delta\mathcal{E}$ term can be easily determined by realizing that in our current choice of quasilocal frame, the only background matter is the "point particle" which is always at the center of our present coordinate system, *i.e.* it is always on $\mathring{\mathscr{C}}$ (on which we are here centering our Fermi normal coordinates). Interpreting the constant m as in the Gralla-Wald approach [Gralla and Wald 2008] to be the "mass" of this "point particle," this simply means that

$$\delta\mathcal{E} = \frac{m}{4\pi r^2}\,, \tag{5.4.52}$$

so that when this is integrated (as a surface energy density) over $\mathbb{S}^2_r$, we simply recover the mass: $\int_{\mathbb{S}^2_r} r^2 \boldsymbol{\epsilon}_{\mathbb{S}^2}\,\delta\mathcal{E} = m$. We remark that, by definition, it is possible to express the quasilocal energy as $\mathcal{E} = u^a u^b \tau_{ab} = -\frac{1}{\kappa}k$ with $k = \boldsymbol{\sigma} : \boldsymbol{\Theta}$ the trace of the two-dimensional boundary extrinsic curvature. Notice that the integral of this over a closed two-surface in the $r\to\infty$ limit is in fact the same as the usual ADM definition of the mass/energy; thus $\delta\mathcal{E} = -\frac{1}{\kappa}\delta k$, and so it makes sense to interpret m as the ADM mass of the object. So now, using Eq. (5.4.52), we can find that the $\delta\mathcal{E}$ contribution to $\delta\dot{\mathrm{p}}^{(\phi)}_{\ell=1}$ is also at most quadratic in r:

$$\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\mathcal{E})} = \mathcal{O}\left(r^2\right)\,. \tag{5.4.53}$$

To compute the $\delta\mathrm{P}$ term, we now employ the useful identity in Eq. (5.2.26), which tells us that

$$\delta\mathrm{P} = \frac{1}{2}\delta\mathcal{E} - \frac{1}{\kappa}\delta a_{\boldsymbol{n}}\,. \tag{5.4.54}$$

Using this, into which we insert the $\delta\mathcal{E}$ from Eq. (5.4.52), we find that the $\delta\mathrm{P}$ contribution to $\delta\dot{\mathrm{p}}^{(\phi)}_{\ell=1}$ is at most quadratic in r as well,

$$\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\mathrm{P})} = \mathcal{O}\left(r^2\right)\,. \tag{5.4.55}$$

Note that the above results may in fact be higher order in r than quadratic. We have only explicitly checked that they vanish up to linear order inclusive.

Finally we are left with the $\delta\boldsymbol{\alpha}$ and $\delta\boldsymbol{D}$ contributions to $\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}$. By direct computation, it is possible to show that their sum is in fact precisely what we have referred to as the extended GSF in our general analysis of the preceding section, *i.e.* it is the $\ell=1$ part of Eq. (5.3.19),

$$\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\boldsymbol{\alpha})} + \delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\boldsymbol{D})} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\Delta\mathrm{p}^{(\phi_{\ell=1})}_{\text{self}}\right)\,. \tag{5.4.56}$$

In particular, they respectively contribute the usual GSF (from $\delta\boldsymbol{\alpha}$) and the gravitational self-pressure force (from $\delta\boldsymbol{D}$).

Thus, we have found that the total $\mathcal{O}(\lambda)$, $\ell=1$ part of the momentum time rate of change is given at leading (zeroth) order in r by nothing more than the generalized GSF. In other words,

$$\boxed{\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})} = -\Phi_I \mathrm{F}^I + \mathcal{O}\left(r\right)}\,, \tag{5.4.57}$$

where we have defined

$$\mathrm{F}^I = -\frac{2}{\kappa} \int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2}\, \mathfrak{S}^{\mathrm{ij}} \mathfrak{B}^I_{\mathrm{i}} \mathcal{F}_{\mathrm{j}}[\boldsymbol{h}; \mathring{\boldsymbol{u}}] + \mathcal{O}\left(r\right) . \tag{5.4.58}$$

Without loss of generality, let us now pick $\Phi^I = (0, 0, 1)$ to be the unit vector in the Cartesian $X^3 = Z$ direction, and denote its corresponding conformal Killing vector as $\boldsymbol{\phi}_{\ell=1} = \boldsymbol{\phi}^Z_{\ell=1}$. (Alternately, pick the Z-axis to be oriented along Φ^I.) We know $\mathfrak{S}^{\mathrm{ij}} \mathfrak{B}^Z_{\mathrm{j}} = (-1/\sin\theta, 0)$; moreover, by the coordinate transformation $\mathcal{F}_{\mathrm{i}} = (\partial x^J/\partial x^{\mathrm{i}})\mathcal{F}_J$ we have $\mathcal{F}_\theta = \cos\theta(\cos\phi\mathcal{F}_X + \sin\phi\mathcal{F}_Y) - \sin\theta\mathcal{F}_Z$. Inserting these into Eq. (5.4.57) we get

$$\delta\dot{\mathrm{p}}^{(\phi^Z_{\ell=1})} = -\frac{2}{\kappa} \int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2}\, \mathcal{F}_Z[\boldsymbol{h}; \mathring{\boldsymbol{u}}] + \frac{2}{\kappa} \int_{\mathbb{S}^2_r} \mathrm{d}\theta \wedge \mathrm{d}\phi\, \cos\theta\left(\cos\phi\mathcal{F}_X + \sin\phi\mathcal{F}_Y\right) . \tag{5.4.59}$$

The first line is precisely in the form of the GSF term from the Gralla formula, Eq. (1.5.7) [Gralla 2011], except here in the integrand we have (the Z-component of) our extended GSF $\boldsymbol{\mathcal{F}}$ [Eq. (5.3.18)]: the usual GSF $\boldsymbol{F}$ (the only self-force term in Gralla's formula) plus our self-pressure term, $\wp$. The second line contains additional terms involving the extended GSF in the other two (Cartesian) spatial directions. Notice however that $\int_{\mathbb{S}^2_r} \mathrm{d}\theta \wedge \mathrm{d}\phi\, \cos\theta\cos\phi = 0 = \int_{\mathbb{S}^2_r} \mathrm{d}\theta \wedge \mathrm{d}\phi\, \cos\theta\sin\phi$, so if $\mathcal{F}_X$ and $\mathcal{F}_Y$ do not vary significantly over $\mathbb{S}^2_r$, their contribution will be subdominant to that of $\mathcal{F}_Z$.

Thus, we have shown that our EoM always *contains* Gralla's "angle average" of the "bare" (usual) GSF. The precise conditions under which the latter *exactly* recovers the former are still under investigation; we conjecture that a careful imposition of the parity condition on the perturbative gauge—of which we have made no explicit use so far—would achieve this, but a detailed proof is required and remains to be carried out.

5.4.5. Equation of motion inertial with the moving object in the perturbed spacetime. Now let $\mathscr{G} = \mathscr{C} \neq \mathring{\mathscr{C}}$ (so $\boldsymbol{A} \neq 0$ in general) such that the quasilocal frame $(\mathscr{B}; \boldsymbol{u})$ centered on $\mathscr{C}$ (in $\mathring{\mathscr{M}}$) is the inverse image of the rigid quasilocal frame $(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})$ defined by $r = C\lambda + \varepsilon = \mathrm{const.}$, $\forall\varepsilon > 0$, in $\mathscr{M}_{(\lambda)}$. The meaning of the r coordinate in the latter is as given in the Gralla-Wald assumptions (Subsection 5.4.1). This situation is displayed in Fig. 5.7.

We now proceed to calculate, in the same way as we did for the "point particle"-inertial case, the various terms in the expansion of the momentum conservation law, Eqs. (5.4.46)-(5.4.47). At zeroth order we obtain:

$$\left(\dot{\mathrm{p}}^{(\boldsymbol{\phi}_{\ell=1})}\right)_{(0)} = \mathcal{O}\left(r^2\right) , \tag{5.4.60}$$

$$\left(\dot{\mathrm{p}}^{(\boldsymbol{\phi}_{\ell=2})}\right)_{(0)} = \mathcal{O}\left(r^2\right) . \tag{5.4.61}$$

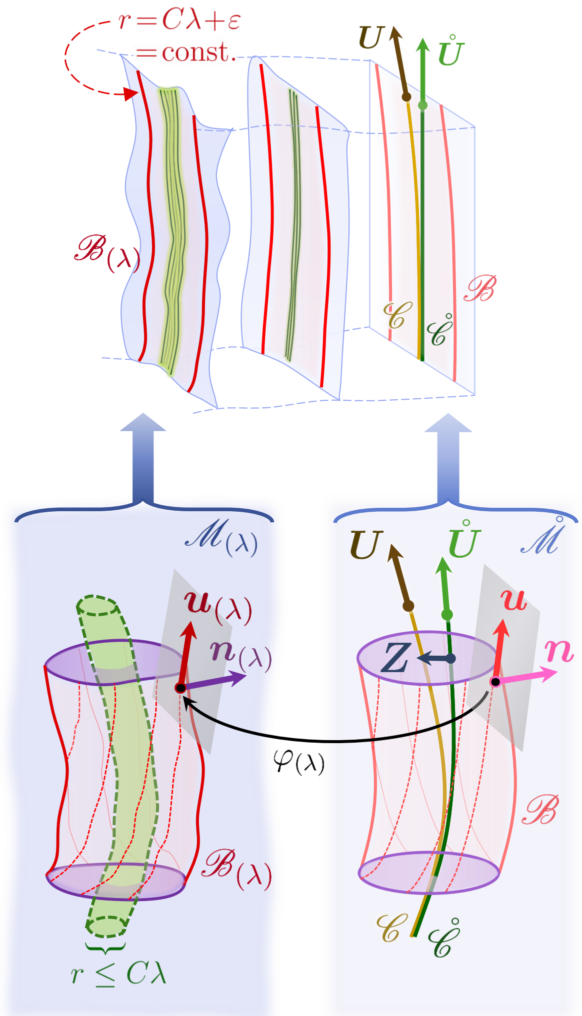

FIGURE 5.7. A family of rigid quasilocal frames $\{(\mathscr{B}_{(\lambda)}; \boldsymbol{u}_{(\lambda)})\}$ embedded in the Gralla-Wald family of spacetimes $\{\mathscr{M}_{(\lambda)}\}$ inertial with the moving object in $\mathscr{M}_{(\lambda)}$. This means that $\mathscr{B}_{(\lambda)}$ is defined by the constancy of the Gralla-Wald r coordinate in $\mathscr{M}_{(\lambda)}$, for any $r > C\lambda$. Thus, the inverse image $\mathscr{B}$ of $\mathscr{B}_{(\lambda)}$ in the background $\mathring{\mathscr{M}}$ is centered, in general, *not* on the geodesic $\mathring{\mathscr{C}}$ followed by the "point particle" background approximation of the object, but on some timelike worldline $\mathscr{C} \neq \mathring{\mathscr{C}}$, with four-velocity $\boldsymbol{U} \neq \mathring{\boldsymbol{U}}$, which may be regarded as an approximation on $\mathring{\mathscr{M}}$ of the "true motion" of the object in $\mathscr{M}_{(\lambda)}$. Between $\mathring{\mathscr{C}}$ and $\mathscr{C}$ there is a deviation vector $\boldsymbol{Z}$, which can be identified with the deviation vector ("correction to the motion") in the Gralla-Wald approach.

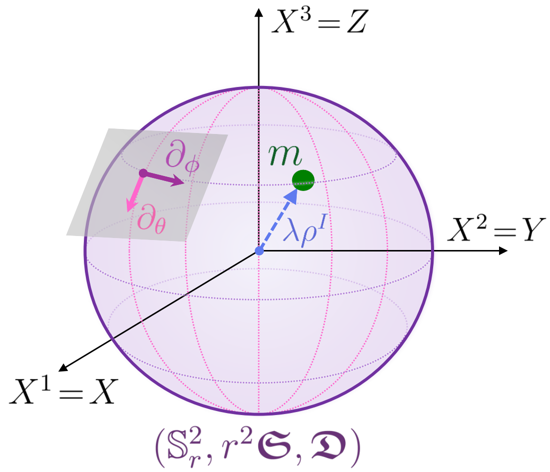

FIGURE 5.8. An instantaneous rigid quasilocal frame $(\mathbb{S}^2_r, r^2\mathfrak{S}, \mathfrak{D})$ (where $\mathfrak{S}$ and $\mathfrak{D}$ respectively are the metric and derivative compatible with the unit two-sphere) inertial with the moving object in the perturbed spacetime. This means that the "point particle" approximation of this object in the background spacetime is *not* located at the center of our Fermi normal coordinate system. Instead, it is displaced in some direction ρ^I, which must be $\mathcal{O}(\lambda)$.

The steps of all these computations are again shown in Appendix B of [Oltean, Epp, Sopuerta, et al. 2019].

Let us now compute the $\mathcal{O}(\lambda)$, $\ell = 1$ part of $\dot{\mathrm{p}}^{(\phi)}$. First, we find that $\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta N)}$ is the same as in the "point particle"-inertial case, so if δN does not vary significantly over $\mathbb{S}^2_r$, the $\mathcal{O}(r^0)$ part thereof is negligible.

Next let us look at the $\delta\mathcal{E}$ and $\delta\mathrm{P}$ parts. Again, it is useful to consider in this case the visual depiction of the instantaneous quasilocal frame, shown in Fig. 5.8.

In this case, the "particle" (delta function) will *not* be at the center of our coordinate system but instead displaced in some direction ρ^I relative thereto. Nonetheless, we know that this displacement must itself be $\mathcal{O}(\lambda)$ which means that it will only contribute $\mathcal{O}(\lambda)$ corrections to the $\delta\mathcal{E}$ having m exactly at the center, *i.e.* we have

$$\delta\mathcal{E} = \frac{m}{4\pi\left(X^I - \lambda\rho^I\right)\left(X_I - \lambda\rho_I\right)} = \frac{m}{4\pi r^2} + \mathcal{O}\left(\lambda\right), \tag{5.4.62}$$

and as before, $\delta \mathrm{P} = \frac{1}{2}\delta\mathcal{E} - \frac{1}{\kappa}\delta a_{\boldsymbol{n}}$. Using these, we find:

$$\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\mathcal{E})} = -\frac{2}{3}m\Phi_I A^I + \mathcal{O}(r) \, , \tag{5.4.63}$$

$$\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\mathrm{P})} = +\frac{1}{3}m\Phi_I A^I + \mathcal{O}(r) \, , \tag{5.4.64}$$

with the steps shown in Appendix B of [Oltean, Epp, Sopuerta, et al. 2019]. Thus,

$$\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\mathcal{E})} + \delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\mathrm{P})} = -\frac{1}{3}m\Phi_I A^I + \mathcal{O}(r) \, . \tag{5.4.65}$$

Meanwhile, we still have, exactly as in the "point particle"-inertial case,

$$\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\boldsymbol{\alpha})} + \delta\dot{\mathrm{p}}^{(\phi_{\ell=1})}_{(\delta\boldsymbol{D})} = -\Phi_I \mathrm{F}^I + \mathcal{O}(r) \, . \tag{5.4.66}$$

Now, by construction, we know that here $\delta\dot{\mathrm{p}}^{(\phi_{\ell=1})} = 0$, as we are inertial with the moving object (in the "actual" spacetime $\mathscr{M}_{(\lambda)}$). Thus summing the above and equating them to zero, we get

$$0 = \Phi_I \left(-mA^I - 3\mathrm{F}^I\right) + \mathcal{O}(r) \, . \tag{5.4.67}$$

Since Φ^I is arbitrary, we thus get the EoM

$$mA^I = -3\mathrm{F}^I \tag{5.4.68}$$

in the $r \to 0$ limit.

Finally, to cast this EoM into the same form as Gralla-Wald [Gralla and Wald 2008], *i.e.* in terms of a deviation vector $\boldsymbol{Z}$ on $\mathring{\mathscr{C}}$ rather than in terms of the proper acceleration $\boldsymbol{A}$ of $\mathscr{C}$, we use the generalized deviation equation (as the name suggests, the deviation equation between arbitrary worldlines, not necessarily geodesics), Eq. (37) of [Puetzfeld and Obukhov 2016]. In our case, this reads $\lambda\ddot{Z}^I = \lambda A^I - \lambda Z^J \mathring{E}^I{}_J + \mathcal{O}(\lambda^2)$. Combining this with Eq. (5.4.68), we finally recover the $\mathcal{O}(\lambda)$ EoM

$$\boxed{\lambda m\ddot{Z}^I = -3\lambda\mathrm{F}^I - \lambda\mathring{E}^I{}_J Z^J + \mathcal{O}\left(\lambda^2, r\right)} \, . \tag{5.4.69}$$

Note the factor of 3 multiplying the self-force term is in fact present in the general EoM in Gralla's Appendix B (including, in this case, an explicit gauge transformations out of the "parity regular" class, not needed in our EoM), Eq. (B3) of [Gralla 2011].

5.5. Discussion and conclusions

In this paper, we have used quasilocal conservation laws to develop a novel formulation of self-force effects in general relativity, one that is independent of the choice of the perturbative gauge and applicable to any perturbative scheme designed to describe the correction to the motion of a localized object. In particular, we have shown that the correction to the motion of any finite spatial region, due to any perturbation of any spacetime metric, is dominated when that region is "small" (*i.e.* at zero-th order in a series expansion in its areal radius) by an *extended* gravitational self-force: this is the standard gravitational self-force term known up to now plus a new term, not found in previous analyses and attributable to a gravitational pressure effect with no analogue in Newtonian gravity, which we have dubbed the gravitational *self-pressure force*. Mathematically, we have found that the total change in momentum $\Delta \mathrm{p}^{(\phi)} = \mathrm{p}^{(\phi)}_{\text{initial}} - \mathrm{p}^{(\phi)}_{\text{final}}$ between an initial and final time of any (gravitational plus matter) system subject to any metric perturbation $\boldsymbol{h}$ is given, in a direction determined by a conformal Killing vector $\boldsymbol{\phi}$ (see Subsection 5.2.4), by the following flux through the portion of the quasilocal frame (worldtube boundary) $(\mathscr{B}; \mathring{\boldsymbol{u}})$ delimited thereby:

$$\Delta \mathrm{p}^{(\phi)} = -\frac{c^4}{4\pi G} \int_{\Delta\mathscr{B}} \boldsymbol{\epsilon}_{\mathscr{B}} \frac{1}{r} \boldsymbol{\phi} \cdot \boldsymbol{\mathcal{F}}[\boldsymbol{h}; \mathring{\boldsymbol{u}}] + \mathcal{O}(r) \ , \tag{5.5.1}$$

where we have restored units, r is the areal radius, and $\boldsymbol{\mathcal{F}}$ is the extended self-force functional. In particular, $\boldsymbol{\mathcal{F}} = \boldsymbol{F} + \wp$ where $\boldsymbol{F}$ is the usual "bare" self-force [determined by the functional in Eq. (1.5.5)] and $\wp$ is our novel self-pressure force [determined by the functional in Eq. (5.3.17)].

The most relevant practical application of the self-force is in the context of modeling EMRIs. Ideally, one would like to compute the "correction to the motion" at the location of the moving object (SCO). Yet, once a concrete perturbative procedure is established, the latter usually ends up being described by a distribution (Dirac delta function), rendering such a computation ill-defined unless additional tactics (typically in the form of regularizations or Green's functions methods) are introduced. However, if one takes a step back from the exact point denoting the location of the "particle" (the distributional support), and instead considers a flux around it, any singularities introduced in such a model are avoided by construction.

We have, moreover, shown that our approach recovers, in the appropriate setting, the known equations of motion in the context of one particular and very common approach to the self-force, namely that of Gralla and Wald [Gralla and Wald 2008]—and specifically, contains the "angle average" term of Gralla [Gralla 2011] proposed within this approach.

We would like here to offer a concluding discussion on our results in this paper in Subsection 5.5.1, as well as outlook towards future work in Subsection 5.5.2.

5.5.1. Discussion of results. From a physical point of view, our approach offers a fresh and conceptually clear perspective on the basic mechanism responsible for the emergence of self-force effects in general relativity. In particular, we have demonstrated that the self-force may be regarded as nothing more than the manifestation of a *physical flux of gravitational momentum* passing through the boundary enclosing the "small" moving object. This gravitational momentum, and gravitational stress-energy-momentum in general, cannot be defined locally in general relativity. As we have argued at length in this paper, such notions must instead be defined quasilocally, *i.e.* as boundary rather than as a volume densities. This is why the self-force appears mathematically as a boundary integral around the moving object [Eq. (5.3.19)], dominant in the limit where the areal radius is small.

The interpretation of the physical meaning of the self-force as a consequence of conservation principles leads to many interesting implications. As we have seen, the "mass" of the moving object—*e.g.*, the mass m of the SCO in the EMRI problem—seems to have nothing to do *fundamentally* with the general existence of a self-force effect. Indeed, according to our analysis, the self-force is in fact generically present as a correction to the motion—and dominant when the moving region is "small"—whenever one has *any* perturbation $\boldsymbol{h}$ to the spacetime metric that is non-vanishing on the boundary of the system.

The usual way to understand the gravitational self-force up to now has been to regard it as a backreaction of m on the metric, *i.e.* on the gravitational field, and thus in turn upon its own motion through that field. Schematically, one thus imagines that the linear correction to the motion is "linear in m" (or more generally, that the full correction is an infinite series in m), *i.e.* that it has the form $\delta\dot{\mathsf{p}} \sim m\delta\mathsf{a}$, with a "perturbed acceleration" $\delta\mathsf{a}$ determined by $\boldsymbol{h}$ (according to some perturbative prescription) causing a correction to the momentum $\delta\dot{\mathsf{p}}$ by a (linear) coupling to the mass m.

Our analysis, instead, shows that this momentum correction $\delta\dot{\mathsf{p}}$ actually arises fundamentally in the schematic form

$$\delta\dot{\mathsf{p}} \sim \mathcal{E}_{\text{vac}}\delta\mathsf{a} + \mathrm{P}_{\text{vac}}\delta D\,, \tag{5.5.2}$$

where $\mathcal{E}_{\text{vac}}$ and P_{vac} are the *vacuum* energy and pressure [Eqs. (5.3.10)-(5.3.11) respectively], and $\delta\mathsf{a}$ and δD are perturbed acceleration and gradient terms determined by $\boldsymbol{h}$. Thus it is the vacuum energy (or "mass") and vacuum pressure, *not* the "mass" of the moving object, which are responsible for the backreaction that produces self-force corrections.

Certainly, the metric perturbation $\boldsymbol{h}$ on the system boundary determining the perturbed acceleration and gradient terms in (5.5.2) may in turn be sourced by a "small mass" present in the interior of the system. In fact, if indeed the system is "small", there may well be little physical reason for expecting that (the dominant part of) $\boldsymbol{h}$ would come for anything *other than* the presence of the "small" system itself. Concordantly, the aim of any concrete self-force analysis is to prescribe exactly *how* $\boldsymbol{h}$ is sourced thereby. Nevertheless, the correction (5.5.1) is valid *regardless* of where $\boldsymbol{h}$ comes from, and regardless of the interior description of the system, which may very well be completely empty of matter or even contain "exotic" matter (as long as a well-posed initial value formulation exists). The EMRI problem is just a special case, where $\boldsymbol{h}$ is sourced in the background, according to the approach considered here, by a rudimentary "point particle" of mass m.

This opens up many interesting conceptual questions, especially with regards to the meaning of the quasilocal vacuum energy and pressure. While traditionally these have often been regarded as unphysical, to be "subtracted away" as reference terms (for the same reason that a "reference action" is often subtracted from the total gravitational action in Lagrangian formulations of GR), our analysis in this paper reveals instead that they are absolutely indispensable to accounting for self-force effects. (Indeed, the initial work [Epp, P. L. McGrath, et al. 2013] on the formulation of the quasilocal momentum conservation laws had similarly revealed the necessity of keeping these terms for a proper accounting of gravitational energy-momentum transfer in general.) To put it simply, the vacuum energy is what seems to play the role of the "mass" in the "mass times acceleration" of the self-force; the pressure term, leading to what we have called the self-pressure force, has no Newtonian analogue.

Now let us comment on our results from a more mathematical and technical point of view. When applied to a specific self-force analysis, namely that of Gralla and Wald [Gralla and Wald 2008], we have been able to recover the "angle average" formula of Gralla [Gralla 2011]. The latter was put forward on the basis of a convenient mathematical argument in a Hamiltonian setting. As the quasilocal stress-energy-momentum definitions that we have been working with (namely, as given by the Brown-York tensor) recover the usual Hamiltonian definitions under appropriate conditions (stationary asymptotically-flat spacetimes with a parity condition), it is reasonable that our general equation of motion [Eq. (5.3.19)]—expressing the physical flux of gravitational momentum—should thereby recover that of Gralla [Eq. (5.4.6)]—expressing an "angle average" in a setting where certain surface integral definitions of general-relativistic Hamiltonian notions (in particular, a Hamiltonian "center of mass") can be well-defined. The limitation of Gralla's equation of motion (*e.g.* in terms of the perturbative gauge restriction attached to it) *vis-à-vis* our general equation of motion is therefore essentially the reflection of the general limitation of Hamiltonian notions of gravitational stress-energy-momentum (as defined for a total, asymptotically-flat spacetime with parity conditions) *vis-à-vis* general quasilocal notions of such concepts—of which the Hamiltonian ones arise simply as a special case.

For carrying out practical EMRI computations, there is a manifest advantage in formulating the self-force as a closed two-surface integral around the moving "particle" versus standard approaches. In the latter, one typically attempts to formulate the problem *at* the "particle location", *i.e.* the support of the distributional matter stress-energy-momentum tensor modeling the moving object (SCO) in the background spacetime. Of course, due to the distributional source, $\boldsymbol{h}$ actually diverges on its support, and so regularization or Green's function methods are typically employed in order to make progress. However, in principle, no such obstacles are encountered (nor the aforementioned technical solutions needed) if the self-force is evaluated on a boundary around—very close to, but at a finite distance away from—the "particle", where no formal singularity is ever encountered: $\boldsymbol{h}$ remains everywhere finite over the integration, and therefore so does the (extended) self-force functional [Eq. (5.3.18)] with it directly as its argument.

5.5.2. Outlook to future work. A numerical implementation of a concrete self-force computation using the approach developed in this paper would be arguably the most salient next step to take. To our knowledge, no numerical work has been put forth even using Gralla's "angle average" integral formula [Gralla 2011] (which would further require gauge transformations away from "parity-regular" gauges).

We stress here that our proposed equation of motion involving the gravitational self-force is entirely formulated and in principle valid in any choice of perturbative gauge. To our knowledge, this is the first such proposal bearing this feature. This may provide a great advantage for numerical work, as black hole metric perturbations $\boldsymbol{h}$ are often most easily computed (by solving the linearized Einstein equation, usually with a delta-function source motivated as in or similarly to the Gralla-Wald approach [Gralla and Wald 2008] described in Subection 5.4.1) in a gauge that is *not* in the "parity-regular" class restricting Gralla's formula [Gralla 2011]. In other words, we claim that one may solve the linearized Einstein equation [Eq. (5.4.2)] for $\boldsymbol{h}^{\mathsf{X}}$ in any desired choice of gauge X, insert this $\boldsymbol{h}^{\mathsf{X}}$ into our extended GSF functional [Eq. (5.3.18)] to obtain $\mathcal{F}^{\mathsf{X}}[\boldsymbol{h}^{\mathsf{X}}; \mathring{\boldsymbol{u}}^{\mathsf{X}}]$ (for some choice of background quasilocal frame with four-velocity $\mathring{\boldsymbol{u}}$), and then to integrate this over a "small radius" topological two-sphere surrounding the "particle" (so that $\mathring{\boldsymbol{u}}$ can be approximated by the background geodesic four-velocity of the particle, $\mathring{\boldsymbol{U}}$), to obtain the full *extended* gravitational self-force (or "correction to the motion") *directly in that gauge* X. It is easy to speculate that this may simplify some numerical issues tremendously *vis-à-vis* current approaches, where much technical machinery is needed to handle (and to do so in a sufficiently efficient way for future waveform applications) the necessary gauge transformations involving distributional source terms.

Nevertheless, further work is needed to bring the relatively abstract analysis developed in this paper into a form more readily suited for practical numerics. The most apparent technical issue to be tackled involves the fact that $\boldsymbol{h}$ is usually computed (in some kind of harmonics) in angular coordinates centered on the MBH, while the functional $\boldsymbol{\mathcal{F}}[\boldsymbol{h}; \mathring{\boldsymbol{u}}]$ is evaluated in angular coordinates (on a "small" topological two-sphere) centered on the moving "particle", *i.e.* the SCO. A detailed understanding of the transformation between the two sets of angular coordinates is thus essential to formulate this problem numerically. This issue is discussed a bit further in Gralla's paper [Gralla 2011], but a detailed implementation of such a computation remains to be attempted.

The abstraction and generality of our approach may, on the other hand, also provide useful ways to address some other technical issues surrounding the self-force problem. For example, all the calculations in this paper may be carried on to second order (in the formal expansion parameter λ)—which is conceptually straightforward given our basic perturbative setup, but of course which requires an analysis in its own right. Nonetheless, one may readily see that any higher-order correction to the motion manifestly remains here in the form of a boundary flux—only now involving nonlinear terms in the integrand.

Thus any sort of singular behaviour is avoided at the level of the equations of motion in our approach, up to any order.

As another example, if ever desired (*e.g.* for astrophysical reasons), linear or any higher-order in r (the areal radius of the SCO boundary) effects on the correction to the motion can also be computed using our approach. Moreover, any matter fluxes (described by the usual matter stress-energy-momentum tensor, $\boldsymbol{T}$) can also be accommodated thanks to our general (gravity plus matter) conservation laws [Eq. (5.2.25)].

Furthermore, while we have applied our ideas in this paper to a specific self-force approach—that of Gralla and Wald [Gralla and Wald 2008]—our general formulation (Section 5.3) can just as well be used in any other approach to the gravitational self-force, *i.e.* any other specification of a perturbative procedure (of a family of perturbed spacetimes $\{(\mathscr{M}_{(\lambda)}, \boldsymbol{g}_{(\lambda)}\})$ for this problem. In other words, our approach permits any alternative specification of what is meant by a (sufficiently) "localized source" in general relativity, as our conservation expressions always involve fluxes on their boundaries and are not conditioned in any way by the exact details of their interior modeling. Thus our equation of motion [Eq. (5.5.1)] could be used not only for a "self-consistent" computation (using, *e.g.*, an approach such as that of Refs. [Ritter, Aoudia, et al. 2016; A. D. A. M. Spallicci and Ritter 2014] for solving the field equations in this context) within the Gralla-Wald approach, but also, for example, in the context of the (mathematically equivalent) self-consistent formulation of Pound [Pound 2010].

Beyond the gravitational self-force, another avenue to explore from here—of interest at the very least for conceptual consistency—is how our approach handles the electromagnetic self-force problem. Although undoubtedly some conceptual parallels may be drawn between the gravitational and electromagnetic self-force problems (see *e.g.* [Barack and Pound 2018]), foundationally they are usually treated as separate problems. Indeed, shortly after the paper of Gralla and Wald [Gralla and Wald 2008] detailing the self-force approach used in this work, Gralla, Harte and Wald [Gralla, Harte, et al. 2009] put forth a similar analysis, with an analogous approach and level of rigour, of the electromagnetic self-force. It would be of great interest to apply our quasilocal conservation laws in this setting, as they can be used to account not just for gravitational but also (and in a consistent way) matter fluxes as well. It may thus prove insightful to study how the transfer of energy-momentum is actually accounted for (between the gravitational and the matter sector), as in our approach we are not restricted to fixing a non-dynamical metric in the spacetime. In other words, the conservation laws account completely for fluxes due to a dynamical geometry as well as matter.

5.6. Appendix: conformal killing vectors and the two-sphere

In this appendix we review some basic properties of *conformal Killing vectors* (CKVs), and in particular CKVs on the two-sphere.

A vector field $\boldsymbol{X}$ on any n-dimensional Riemannian manifold $(\mathscr{U}, \boldsymbol{g}_{\mathscr{U}}, \boldsymbol{\nabla}_{\mathscr{U}})$ is a CKV if and only if it satisfies the *conformal Killing equation*,

$$\mathcal{L}_{\boldsymbol{X}} \boldsymbol{g}_{\mathscr{U}} = \psi \boldsymbol{g}_{\mathscr{U}} \,, \tag{5.6.1}$$

where $\psi \in C^\infty(\mathscr{U})$. This function can be determined uniquely by taking the trace of this equation, yielding

$$\psi = \frac{2}{n} \boldsymbol{\nabla}_{\mathscr{U}} \cdot \boldsymbol{X} \,. \tag{5.6.2}$$

Let us now specialize to the r-radius two-sphere $(\mathbb{S}^2_r, r^2\boldsymbol{\mathfrak{S}}, \boldsymbol{\mathfrak{D}})$, where we denote our CKV by $\boldsymbol{\phi}$. Moreover for ease of notation in this appendix, the two-sphere volume form $\boldsymbol{\epsilon}_{\mathbb{S}^2}$ is equivalently denoted by $\boldsymbol{\mathfrak{E}}$, *i.e.* $\mathfrak{E}_{\mathrm{ij}} = \epsilon^{\mathbb{S}^2}_{\mathrm{ij}}$.

In this case, the conformal Killing equation (5.6.1) is

$$\mathfrak{D}^{(\mathrm{i}}\phi^{\mathrm{j})} = \frac{1}{2r^2} \mathfrak{S}^{\mathrm{ij}} \mathfrak{D}_{\mathfrak{k}} \phi^{\mathfrak{k}} \Leftrightarrow \mathfrak{D}^{\langle \mathrm{i}}\phi^{\mathrm{j}\rangle} = 0 \,, \tag{5.6.3}$$

where $\langle\cdot\cdot\rangle$ on two indices indicates taking the STF part. The solution to this equation can be usefully expressed in the form of a spherical harmonic decomposition in terms of the standard direction cosines of a radial unit vector in $\mathbb{R}^3$, which we denote by r^I. In spherical coordinates $\{x^{\mathrm{i}}\} = \{\theta, \phi\}$, it is simply given by

$$r^I(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \,. \tag{5.6.4}$$

This satisfies the following useful identity:

$$\int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2} \prod_{n=1}^{\ell} r^{I_\ell} = \begin{cases} 0 \,, & \text{for } \ell \text{ odd} \,, \\ \frac{4\pi}{(\ell+1)!!} \delta^{\{I_1 I_2} \ldots \delta^{I_{\ell-1} I_\ell\}} \,, & \text{for } \ell \text{ even} \,, \end{cases} \tag{5.6.5}$$

where $(\ell+1)!! = (\ell+1)(\ell-1)\cdots 1$ and the curly brackets on the indices denote the smallest set of permutations that make the result symmetric. In particular, the $\ell = 2$ and $\ell = 4$ cases (which suffice for the calculations presented in this paper) are:

$$\int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2} \, r^I r^J = \frac{4\pi}{3} \delta^{IJ} \,, \tag{5.6.6}$$

$$\int_{\mathbb{S}^2_r} \boldsymbol{\epsilon}_{\mathbb{S}^2} \, r^I r^J r^K r^L = \frac{4\pi}{15} \left(\delta^{IJ}\delta^{KL} + \delta^{IK}\delta^{JL} + \delta^{IL}\delta^{KJ}\right) \,. \tag{5.6.7}$$

We can construct from Eq. (5.6.4) two sets of $\ell = 1$ spherical harmonic forms on $\mathbb{S}^2_r$, namely the *boost generators*,

$$\mathfrak{B}^I_{\mathfrak{i}}(\theta,\phi) = \mathfrak{D}_{\mathfrak{i}} r^I = \begin{pmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\theta\sin\phi & \sin\theta\cos\phi & 0 \end{pmatrix}\,, \tag{5.6.8}$$

and the *rotation generators*,

$$\mathfrak{R}^I_{\mathfrak{i}}(\theta,\phi) = \mathfrak{E}_{\mathfrak{i}}{}^{\mathfrak{j}}\mathfrak{B}^I_{\mathfrak{j}} = \epsilon^I{}_{JK} r^J \mathfrak{B}^K_{\mathfrak{i}} = \begin{pmatrix} -\sin\phi & \cos\phi & 0 \\ -\sin\theta\cos\theta\cos\phi & -\sin\theta\cos\theta\sin\phi & \sin^2\theta \end{pmatrix}, \tag{5.6.9}$$

where ϵ_{IJK} is the volume form of $\mathbb{R}^3$. We can obtain from these the vector fields $\mathfrak{B}^{\mathfrak{i}}_I = \frac{1}{r^2}\delta_{IJ}\mathfrak{S}^{\mathfrak{i}\mathfrak{j}}\mathfrak{B}^J_{\mathfrak{j}} = \mathfrak{D}^{\mathfrak{j}} r_I$ and $\mathfrak{R}^{\mathfrak{i}}_I = \frac{1}{r^2}\delta_{IJ}\mathfrak{S}^{\mathfrak{i}\mathfrak{j}}\mathfrak{R}^J_{\mathfrak{j}} = \mathfrak{E}^{\mathfrak{i}}{}_{\mathfrak{j}}\mathfrak{B}^{\mathfrak{j}}_I$, which satisfy the Lorentz algebra

$$[\boldsymbol{\mathfrak{B}}_I, \boldsymbol{\mathfrak{B}}_J] = \epsilon_{IJ}{}^K \boldsymbol{\mathfrak{R}}_K\,, \tag{5.6.10}$$

$$[\boldsymbol{\mathfrak{R}}_I, \boldsymbol{\mathfrak{B}}_J] = -\,\epsilon_{IJ}{}^K \boldsymbol{\mathfrak{B}}_K\,, \tag{5.6.11}$$

$$[\boldsymbol{\mathfrak{R}}_I, \boldsymbol{\mathfrak{R}}_J] = -\,\epsilon_{IJ}{}^K \boldsymbol{\mathfrak{R}}_K\,. \tag{5.6.12}$$

From the above, it is possible to derive a number of useful properties:

$$r_I\mathfrak{B}^I_{\mathfrak{i}} = 0\,,\quad \mathfrak{D}_{\mathfrak{i}}\mathfrak{B}^I_{\mathfrak{j}} = -\mathfrak{S}_{\mathfrak{i}\mathfrak{j}} r^I \Rightarrow \mathfrak{S}^{\mathfrak{i}\mathfrak{j}}\mathfrak{D}_{\mathfrak{i}}\mathfrak{B}^I_{\mathfrak{j}} = -2r^I\,, \tag{5.6.13}$$

$$r_I\mathfrak{R}^I_{\mathfrak{i}} = 0\,,\quad \mathfrak{D}_{\mathfrak{i}}\mathfrak{R}^I_{\mathfrak{j}} = \mathfrak{E}_{\mathfrak{i}\mathfrak{j}} r^I \Rightarrow \mathfrak{S}^{\mathfrak{i}\mathfrak{j}}\mathfrak{D}_{\mathfrak{i}}\mathfrak{R}^I_{\mathfrak{j}} = 0\,. \tag{5.6.14}$$

Using these, one can show that the sets of $\ell = 1$ vector fields $\boldsymbol{\mathfrak{B}}_I$ and $\boldsymbol{\mathfrak{R}}_I$ all satisfy the conformal Killing equation, *i.e.*

$$\mathfrak{D}^{\langle\mathfrak{i}}\mathfrak{B}^{\mathfrak{j}\rangle}_I = 0 = \mathfrak{D}^{\langle\mathfrak{i}}\mathfrak{R}^{\mathfrak{j}\rangle}_I\,. \tag{5.6.15}$$

Finally, we give a list of useful relations for various contractions involving these vector fields:

$$\mathfrak{S}^{\mathfrak{i}\mathfrak{j}}\mathfrak{B}^I_{\mathfrak{i}}\mathfrak{B}^J_{\mathfrak{j}} = \mathfrak{S}^{\mathfrak{i}\mathfrak{j}}\mathfrak{R}^I_{\mathfrak{i}}\mathfrak{R}^J_{\mathfrak{j}} = -\mathfrak{E}^{\mathfrak{i}\mathfrak{j}}\mathfrak{B}^I_{\mathfrak{i}}\mathfrak{R}^J_{\mathfrak{j}} = P^{IJ}\,, \tag{5.6.16}$$

$$\mathfrak{S}^{\mathfrak{i}\mathfrak{j}}\mathfrak{B}^I_{\mathfrak{i}}\mathfrak{R}^J_{\mathfrak{j}} = \mathfrak{E}_{\mathfrak{i}\mathfrak{j}}\mathfrak{B}^{\mathfrak{i}}_I\mathfrak{B}^{\mathfrak{j}}_J = \mathfrak{E}_{\mathfrak{i}\mathfrak{j}}\mathfrak{R}^{\mathfrak{i}}_I\mathfrak{R}^{\mathfrak{j}}_J = \epsilon^{IJK} r_K\,, \tag{5.6.17}$$

$$\delta_{IJ}\mathfrak{B}^I_{\mathfrak{i}}\mathfrak{B}^J_{\mathfrak{j}} = \delta_{IJ}\mathfrak{R}^I_{\mathfrak{i}}\mathfrak{R}^J_{\mathfrak{j}} = \mathfrak{S}_{\mathfrak{i}\mathfrak{j}}\,, \tag{5.6.18}$$

$$\delta_{IJ}\mathfrak{B}^I_{\mathfrak{i}}\mathfrak{R}^J_{\mathfrak{j}} = -\,\mathfrak{E}_{\mathfrak{i}\mathfrak{j}}\,, \tag{5.6.19}$$

$$\epsilon_{IJK}\mathfrak{B}^I_{\mathfrak{i}}\mathfrak{B}^J_{\mathfrak{j}} = \epsilon_{IJK}\mathfrak{R}^I_{\mathfrak{i}}\mathfrak{R}^J_{\mathfrak{j}} = \mathfrak{E}_{\mathfrak{i}\mathfrak{j}} r_K\,, \tag{5.6.20}$$

$$\epsilon_{IJK}\mathfrak{B}^I_{\mathfrak{i}}\mathfrak{R}^J_{\mathfrak{j}} = \mathfrak{S}_{\mathfrak{i}\mathfrak{j}} r_K\,, \tag{5.6.21}$$

where $P^{IJ} = \delta^{IJ} - r^I r^J$ projects vectors perpendicular to the radial direction.

Now we have everything in hand to formulate the general solution to the conformal Killing equation (5.6.3) on $\mathbb{S}^2_r$; it can be expanded as:

$$\phi^{\mathfrak{j}} = \frac{1}{r}\left[\mathfrak{D}^{\mathfrak{j}}\left(\sum_{\ell\in\mathbb{N}}\Phi^{I_1\cdots I_\ell}\prod_{n=1}^{\ell} r_{I_n}\right) + \mathfrak{E}^{\mathfrak{j}}{}_{\mathfrak{k}}\mathfrak{D}^{\mathfrak{k}}\left(\sum_{\ell\in\mathbb{N}}\Psi^{I_1\cdots I_\ell}\prod_{n=1}^{\ell} r_{I_n}\right)\right], \tag{5.6.22}$$

$$= \frac{1}{r}\left[\left(\Phi^I\mathfrak{B}^{\mathfrak{j}}_I + \Psi^I\mathfrak{R}^{\mathfrak{j}}_I\right) + \sum_{\ell\geq 2}\ell\left(\Phi^{I_1\cdots I_\ell}\mathfrak{B}^{\mathfrak{j}}_{I_1} + \Psi^{I_1\cdots I_\ell}\mathfrak{R}^{\mathfrak{j}}_{I_1}\right)\prod_{n=2}^{\ell} r_{I_n}\right], \tag{5.6.23}$$

where to write the second equality we have used the fact that $\Phi^{I_1\cdots I_\ell}$ and $\Psi^{I_1\cdots I_\ell}$ are symmetric in their indices.

We are interested in working with the $\ell = 1$ and $\ell = 2$ parts of $\boldsymbol{\phi}$ corresponding to linear momentum only ($\boldsymbol{\Psi} = 0$):

$$\phi^{\mathfrak{i}}_{\ell=1} = \frac{1}{r}\Phi^I\mathfrak{B}^{\mathfrak{i}}_I\,, \tag{5.6.24}$$

$$\phi^{\mathfrak{i}}_{\ell=2} = \frac{2}{r}\Phi^{IJ}\mathfrak{B}^{\mathfrak{i}}_I r_J\,. \tag{5.6.25}$$

CHAPTER 6

A Frequency-Domain Implementation of the Particle-without-Particle Approach to EMRIs

Chapter summary. This chapter is based on the conference proceeding [Oltean, Sopuerta, et al. 2017] and ongoing work.

We present here a frequency-domain implementation of the Particle-without-Particle (PwP) technique previously developed for the computation of the scalar self-force, a helpful testbed for the gravitational case.

We offer a short introduction in Section 6.1, commenting briefly on the choice between time and frequency domain methods in carrying out numerics for the EMRI problem.

In Section 6.2, we formulate the problem of the scalar self-force in a non-spinning black hole spacetime in full mathematical detail. In particular, the moving particle here possesses a scalar charge due to a scalar field which does not back-react on the geometry (*i.e.* the background remains fixed). We comment on the widely-used mode-sum regularization procedure for devising a numerical implementation.

Then in Section 6.3, we discuss the Particle-without-Particle (PwP) method, a pseudospectral collocation method previously used for the computation of the scalar self-force in the time domain. The idea is to decompose quantities into linear combinations of Heaviside functions (supported, in this case, at inner and outer radii relative to the particle orbit), turning the distributionally-sourced field equations into systems of homogeneous equations (away from the particle) supplemented by "jump" (boundary) conditions connecting them (at the particle location).

In Section 6.4, we present the frequency-domain formulation of the scalar self-force problem, including the appropriate boundary and jump conditions.

Finally, in Section 6.5, we discuss in basic outline of our numerical implementation using a hyperbolic compactification and multidomain splitting of the computational grids, omitting much of the technical detail. We also present some results on circular orbits, with work on eccentric orbits in progress.

Una implementació en el domini de freqüències del mètode Partícula-sense-Partícula als *EMRIs* (chapter summary translation in Catalan). Aquest capítol es basa en la acta [Oltean, Sopuerta, et al. 2017] i treball en curs.

Aquí presentem una implementació en el domini de freqüència de la tècnica Partícula-sense-Partícula (Particle-without-Particle, PwP) desenvolupada anteriorment per a la computació de la força pròpia escalar, una prova útil per al cas gravitatori.

Oferim una breu introducció a la secció 6.1, comentant breument sobre l'elecció entre mètodes de domini de temps i domini de freqüència en la solució numèrica del problema EMRI.

A la secció 6.2, formulem el problema de la força pròpia escalar en un espai-temps de un forat negre que no gira, amb tot el detall matemàtic. En particular, la partícula en moviment aquí té una càrrega escalar a causa d'un camp escalar que no retroacciona sobre la geometria (és a dir, el fons queda fixat). Comentem el procediment de regularització de sumes de modes, molt utilitzat per idear implementacions numèriques.

A continuació, a la secció 6.3, es discuteix el mètode Partícula-sense- Partícula (PwP), un mètode de col·locació pseudospectral usat anteriorment per al càlcul de la força pròpia escalar en el domini temporal. La idea és descompondre quantitats en combinacions lineals de funcions Heaviside (suportades, en aquest cas, en els radis interns i externs respecte a l'òrbita de la partícula), convertint les equacions de camp amb fonts distributives en sistemes d'equacions homogènies (allunyades de la partícula) complementades mitjançant les condicions de "salt" (límit) que els connecten (a la ubicació de la partícula).

A la secció 6.4, es presenta la formulació de dominis de freqüència del problema de la força pròpia escalar, incloent-ne els límits i les condicions de salt adequades.

Finalment, a la secció 6.5, es discuteix en l'esquema bàsic de la nostra implementació numèrica mitjançant una compactació hiperbòlica i una divisió multidomànica de les reixes computacionals, ometent gran part del detall tècnic. També presentem alguns resultats sobre òrbites circulars, amb treballs sobre òrbites excèntriques en marxa.

Une implémentation dans le domaine fréquentiel de l'approche Particule-sans-Particule aux *EMRIs* (chapter summary translation in French). Ce chapitre est basé sur l'acte de congrès [Oltean, Sopuerta, et al. 2017] et travaux en cours.

Nous présentons ici une implémentation dans le domaine fréquentiel de la technique Particule-sans-Particule (Particle-without-Particle, PwP) développée précédemment pour le calcul de la force propre scalaire - un test utile pour le cas gravitationnel.

Nous proposons une brève introduction à la section 6.1, en commentant brièvement sur le choix entre les méthodes de domaine temporel et fréquentiel pour la réalisation des calculs pour le problème des EMRIs.

Dans la section 6.2, nous formulons le problème de la force propre scalaire dans un espace-temps de trou noir que ne tourne pas, avec tous les détails mathématiques. En particulier, la particule en mouvement possède ici une charge scalaire due à un champ scalaire que ne rétroactionne pas sur la géométrie (c'est-à-dire que le fonde reste fixe). Nous commentons la procédure de régularisation des sommes des modes largement utilisée pour concevoir une implémentation numérique.

Ensuite, dans la section 6.3, nous discutons de la méthode PwP, une méthode de collocation pseudospectrale précédemment utilisée pour le calcul de la force propre scalaire dans le domaine temporel. L'idée est de décomposer les quantités en combinaisons linéaires de fonctions de Heaviside (supportées, dans ce cas, aux rayons intérieurs et extérieurs par rapport à l'orbite de la particule), en transformant les équations du champ avec sources distributionnelles en systèmes d'équations homogènes (loin de la particule) complétées par des conditions de « saut » (aux limites) que les connecte (à l'emplacement de la particule).

Dans la section 6.4, nous présentons la formulation dans le domaine fréquentiel du problème de la force propre scalaire, y compris les conditions de limite et de saut appropriées.

Enfin, dans la section 6.5, nous discutons dans les grandes lignes de notre implémentation numérique en utilisant une compactification hyperbolique et une division en plusieurs domaines des grilles de calcul, en omettant une grande partie des détails techniques. Nous présentons également quelques résultats sur des orbites circulaires, avec des travaux sur des orbites excentriques en cours.

6.1. Introduction

The computation of the self-force and waveforms, and any other physical relevant information related to the inspiral due to radiation reaction constitute the main challenge of the EMRI problem. One possible strategy is to resort to analytic techniques by adding extra approximations to the problem, similar to those from post-Newtonian methods. However, the results may not be applicable to situations of physical relevance involving highly spinning MBHs and very eccentric orbits. To make computations without making further simplifications of the problem, numerical techniques appear to be a necessary tool.

Broadly speaking, one faces a choice in how to proceed between frequency-domain and time-domain calculations. The frequency domain approach has been used for a long time; it provides accurate results for the computation of quasinormal modes and frequencies [Chandrasekhar and Detweiler 1975; Vishveshwara 1970]. However, this approach encounters greater difficulties when one is interested in computing the waves originated from highly eccentric orbits since one has to sum over a large number of modes to obtain a good accuracy. In this sense, calculations in the time-domain can be better adapted for obtaining accurate waveforms for the physical situations of relevance. Nevertheless, overall, time-domain methods can be much slower than working in the frequency domain.

We consider in this chapter a frequency-domain implementation of a simplified EMRI model, corresponding to a charged scalar particle orbiting a non-rotating MBH. There is in this case no (gravitational) backreaction upon the background geometry (which therefore remains fixed). This offers a very useful setting to test different numerical implementations, with a view towards using those which prove most successful in the full gravitational self-force problem.

The method that we use here, and which has been developed in the past in the time domain, is called the Particle-without-Particle (PwP) method [Canizares 2011; Canizares and Sopuerta 2009b, 2014, 2011a,b; Canizares, Sopuerta, and Jaramillo 2010b; Jaramillo, Sopuerta, et al. 2011]. The basic idea is to split the computational domain into two (or more) disjoint regions whereby any non-singular quantity Q is decomposed as $Q = Q_{-}\Theta_p^{-} + Q_{+}\Theta_p^{+}$, where $\Theta_p^{\pm} = \Theta(\pm(r - r_p))$ is the Heaviside step function with the step at the particle's radial location r_p. Quantities that are not continuous will have jumps across the SCO trajectory, which we denote by $[Q]_p = \lim_{r\to r_p(t)}(Q_{+} - Q_{-})$. In this setup then, any differential equation with a singular (distributional) source is effectively replaced with homogeneous equations to the left and right of the SCO, subject to certain jump conditions across it. See also Appendix B.

6.2. The scalar self-force

In our simplified EMRI model, the SCO is represented as a scalar particle, *i.e.* a body the charge distribution of which has support only on its (timelike) worldline $\mathscr{C}$, parametrized by $z^a(\tau)$, with a charge q associated to a scalar field Φ; meanwhile, the MBH is described by a fixed Schwarzschind-Droste spacetime $(\mathring{\mathscr{M}}, \mathring{\boldsymbol{g}}, \mathring{\boldsymbol{\nabla}})$, *i.e.* a background not affected by the charged particle, with the following metric metric (for more details, see Section 3.3):

$$\mathring{g}_{ab}\mathrm{d}x^a\mathrm{d}x^b = -f\mathrm{d}t^2 + f^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2 \tag{6.2.1}$$

$$= f\left(-\mathrm{d}t^2 + \mathrm{d}r_*^2\right) + r^2\mathrm{d}\Omega^2\,, \tag{6.2.2}$$

where $f(r) = 1 - 2M/r$, $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2$ is the two-sphere line element and $r_* = r + 2M\ln(r/2M - 1)$ is the so-called radial *tortoise* coordinate.

The dynamics are in this case determined by the following action:

$$\mathcal{S}\left[\Phi, z\right] = \int \boldsymbol{\epsilon}_{\mathring{\mathscr{M}}}\left\{\boldsymbol{\nabla}\Phi\cdot\boldsymbol{\nabla}\Phi + \int_{\mathscr{C}}\mathrm{d}\tau\,\delta_4(x - z(\tau))\left[\frac{m}{2}\,\boldsymbol{u}\cdot\boldsymbol{u} + q\Phi\right]\right\}, \tag{6.2.3}$$

where $\boldsymbol{u} = \dot{\boldsymbol{z}}$ is the particle four-velocity and m is its (time-dependent) "mass". The first term is the kinetic term for the scalar field, the one proportional to m is the standard geodesic action for the particle, and the one proportional to q is a coupling between the field and the particle motion—leading to nontrivial sources in both the field equation and equation of motion.

The scalar field satisfies the following wave-like equation, obtained by extremizing the action (6.2.3) with respect to Φ (see, e.g. [Poisson et al. 2011]):

$$\mathring{\Box}\Phi(x) = -4\pi q \int_{\mathscr{C}} \mathrm{d}\tau\, \delta_4(x - z(\tau))\,, \tag{6.2.4}$$

where $\mathring{\Box} = \mathring{\nabla}^a\mathring{\nabla}_a$ is the wave operator.

The field equation (6.2.4) has to be complemented with the equation of motion for the scalar charged particle, obtained by extremizing the action (6.2.3) with respect to $\boldsymbol{z}$:

$$\nabla_{\boldsymbol{u}}(mu^a) = \mathfrak{F}^a = q\mathring{g}^{ab}\left(\mathring{\nabla}_b\Phi\right)\Big|_{\mathscr{C}}\,, \tag{6.2.5}$$

The coupled set of equations formed by the PDE for the scalar field (6.2.4) and the ODE for the particle trajectory (6.2.5) constitute our testbed model for an EMRI. The SCO generates (sources) a scalar field according to (6.2.4), which in turn affects to the SCO motion according to (6.2.5), that is, through the action of a local self-force $\mathfrak{F}^a$. This mechanism is the (scalar) analogue of the gravitational backreaction mechanism that produces the inspiral via the gravitational self-force.

Now, the retarded solution of (6.2.4) is singular at the particle location, while the force in Eq. (6.2.5) involves the gradient of the field evaluated at the particle location. Therefore, as they stand, Eqs. (6.2.4) and (6.2.5) are formal equations that require an appropriate regularization to become fully meaningful. Following [Detweiler and Whiting 2003], the retarded field can be split into two parts: a singular piece, Φ^{S}, which contains the singular structure of the field and satisfies the same wave equation as the retarded field, i.e. Eq. (6.2.4), and a regular part, Φ^{R}, that satisfies the homogeneous equation associated with Eq. (6.2.4). As it turns out, Φ^{R} is regular and differentiable at the particle position and is solely responsible for the scalar self-force [Detweiler and Whiting 2003]. We can therefore write

$$\mathfrak{F}^a_{\mathrm{R}} = q\mathring{g}^{ab}\left(\mathring{\nabla}_b\Phi^{\mathrm{R}}\right)\Big|_{\mathscr{C}}\,, \tag{6.2.6}$$

which gives a definite sense to the equations of motion of the system.

We can solve the field equation (6.2.4) by expanding the scalar field in (scalar) spherical harmonics:

$$\Phi = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} \Phi^{\ell m}(t,r)Y^{\ell m}(\theta,\varphi)\,. \tag{6.2.7}$$

The equations for each harmonic mode, $\Phi^{\ell m}(t,r)$, are decoupled from the rest and take the form of the following $(1+1)$-dimensional wave equation for $\psi^{\ell m} = r\Phi^{\ell m}$:

$$\left(-\partial_t^2 + \partial_{r_*}^2 - V_\ell(r)\right)\psi^{\ell m} = S^{\ell m}\delta(r - r_p(t))\,, \tag{6.2.8}$$

where

$$V_\ell(r) = f(r)\left[\frac{\ell(\ell+1)}{r^2} + \frac{2M}{r^3}\right] , \tag{6.2.9}$$

is the Regge-Wheeler potential for scalar (spin-zero) fields on the Schwarzschild-Droste geometry, and

$$S^{\ell m} = -\frac{4\pi q f_p}{r_p u^t}\,\bar{Y}^{\ell m}\left(\frac{\pi}{2}, \varphi_p(t)\right) , \tag{6.2.10}$$

is the coefficient of the singular source term due to the presence of the particle, $f_p = f(r_p)$, and the bar denotes complex conjugation. Here we have assumed, without loss of generality, that the particle's orbit takes place in the equatorial plane $\theta = \pi/2$. Moreover, r_p and φ_p denote the radial and azimuthal coordinates of the particle, and are functions of the coordinate time t.

The expansion in spherical harmonics is also very useful to construct the regular field, Φ^{R}. Indeed, it turns out that each harmonic mode of the retarded field, $\Phi^{\ell m}(t,r)$, is finite and continuous at the location of the particle; it is the sum over ℓ what diverges there.

Here, the *mode-sum* regularization scheme [Barack, Mino, et al. 2002; Barack and Ori 2000, 2002] comes into play: it provides analytic expressions for the singular part of each ℓ-mode of the retarded field. These expressions for the singular field are valid only near the particle location. The regularized self-force is thus obtained by computing numerically each harmonic mode of the self-force and subtracting the singular part provided by the mode-sum scheme.

The regular part of the gradient of the field, which in coordinates x^α we denote simply as $\nabla_\alpha \Phi^{\mathrm{R}} \equiv \Phi^{\mathrm{R}}_\alpha$, is given by

$$\Phi^{\mathrm{R}}_a(\boldsymbol{z}(\tau)) = \lim_{x^\mu \to z^\mu(\tau)} \sum_{\ell=0}^{\infty} \left(\Phi^\ell_\alpha(x^\mu) - \Phi^{\mathrm{S},\ell}_\alpha(x^\mu)\right) . \tag{6.2.11}$$

where

$$\Phi^\ell_\alpha(x^\mu) = \sum_{m=-\ell}^{\ell} \nabla_\alpha(\Phi^{\ell m}(t,r) Y^{\ell m}(\theta,\varphi)) , \tag{6.2.12}$$

and the structure of the singular field can can be written as:

$$\Phi^{\mathrm{S},\ell}_\alpha \;=\; q\left[\left(\ell + \frac{1}{2}\right) A_\alpha + B_\alpha + \frac{C_\alpha}{\ell+\frac{1}{2}} + \frac{D_\alpha}{(\ell-\frac{1}{2})(\ell+\frac{3}{2})} + \ldots\right] . \tag{6.2.13}$$

The expressions for the regularization parameters A_α, B_α, C_α, and D_α, can be found in the literature for generic orbits [Barack and Ori 2002; Haas and Poisson 2006; Kim 2005, 2004]. They do not depend on ℓ, but on the trajectory of the particle. The three first coefficients of (6.2.13) are responsible for the divergences, whereas the remaining terms converge to zero once they are summed over ℓ. The expressions for the regularization

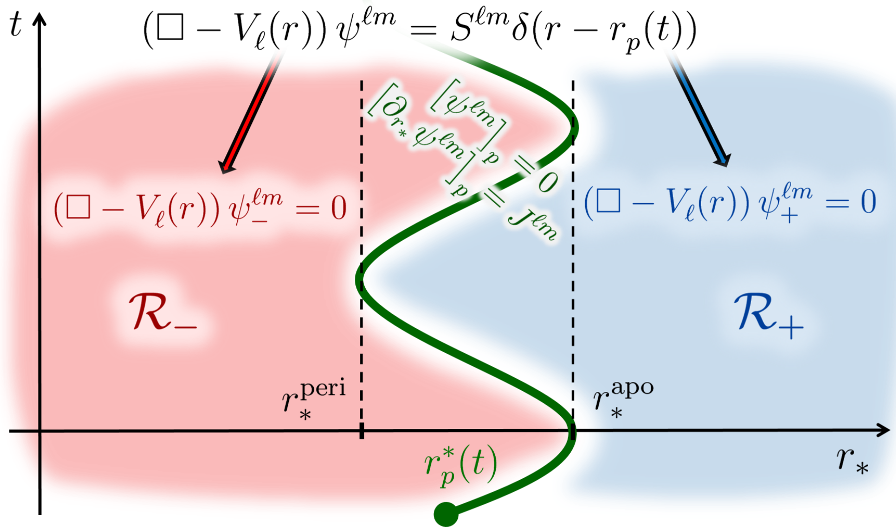

FIGURE 6.1. Schematic representation of the PwP formulation. The field equations with singular source terms become homogeneous equations at each side of the particle worldline together with a set of jump conditions to communicate their solutions.

parameters that we use in this paper are listed in the Appendix of [Canizares, Sopuerta, and Jaramillo 2010b].

6.3. The Particle-without-Particle method

The full retarded solution has to be found numerically and hence, it is very convenient to formulate the equations so that we obtain smooth solutions. However, the presence of singularities in Eqs. (6.2.8) makes the task difficult in principle. To overcome these problems, the Particle-without-Particle (PwP) method [Canizares and Sopuerta 2009a,c, 2011b; Canizares, Sopuerta, and Jaramillo 2010a] splits the computational domain (in the (t, r) space) into two disjoin regions (see Fig. 6.1): Region $\mathcal{R}_-$ to the left of the SCO trajectory ($r < r_p(t)$) and region $\mathcal{R}_+$ to the right ($r > r_p(t)$). Then, any non-singular quantity $Q(t, r)$ admits a decomposition

$$Q = Q_-\Theta^-_p + Q_+\Theta^+_p\,, \tag{6.3.1}$$

where $\Theta^-_p \equiv \Theta(r_p - r)$ and $\Theta^+_p \equiv \Theta(r - r_p)$, and Θ is the Heaviside step function. Quantities that are not continuous will have jumps across the SCO trajectory. The jump in a quantity $\mathcal{Q}$ is a time-only dependent quantity defined as: $[Q](t) = \lim_{r\to r_p(t)}(Q_+(t, r) - Q_-(t, r)) \equiv [Q]_p\,.$

Applying the PwP formulation to the scalar equation (6.2.8), *i.e.* introducing

$$\psi^{\ell m} = \psi^{\ell m}_{-}\,\Theta^{-}_{p} + \psi^{\ell m}_{+}\,\Theta^{+}_{p}\,, \tag{6.3.2}$$

it transforms into homogeneous equations (with no matter source terms) at each region $\mathcal{R}_{\pm}$:

$$\left(-\partial^2_t + \partial^2_{r_*} - V_\ell(r)\right)\psi^{\ell m}_{\pm} = 0\,, \tag{6.3.3}$$

plus a set of jump conditions on $\psi^{\ell m}_{\pm}$ and $\partial_{r_*}\psi^{\ell m}_{\pm}$, which read:

$$\left[\psi^{\ell m}\right]_p = 0\,, \tag{6.3.4}$$

$$\left[\partial_{r_*}\psi^{\ell m}_{\pm}\right]_p = \frac{S^{\ell m}}{(1-(\dot{r}^2_*)_p)f_p} \equiv J^{\ell m}\,. \tag{6.3.5}$$

In summary, at each region we have equations without the singular terms induced by the SCO. Then, since these equations are strongly hyperbolic, we obtain smooth solutions. Finally, the SCO appears in the communication between the two regions by enforcing the jump conditions. The spherical symmetry of the MBH background leads to jumps only in time and radial derivatives. For instance, for first order derivatives we find: $[\partial_t \mathcal{Q}^{\ell m}]_p = \mathrm{d}[\mathcal{Q}^{\ell m}]_p/\mathrm{d}t - \dot{r}_p[\partial_r \mathcal{Q}^{\ell m}]_p$; the same happens for derivatives of higher order. In particular, we get

$$\left[\partial_t\psi^{\ell m}\right]_p = -\frac{(\dot{r}_*)_p S^{\ell m}}{(1-(\dot{r}^2_*)_p)f_p}\,. \tag{6.3.6}$$

6.4. Frequency domain analysis

We now turn the analysis to the frequency domain. We Fourier decompose our solution:

$$\tilde{\psi}^{\ell m}_{\pm}(\omega, r) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{d}t\, e^{\mathrm{i}\omega t}\,\psi^{\ell m}_{\pm}(t,r)\,, \tag{6.4.1}$$

$$\psi^{\ell m}_{\pm}(t, r) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{d}\omega\, e^{-\mathrm{i}\omega t}\,\tilde{\psi}^{\ell m}_{\pm}(\omega,r)\,. \tag{6.4.2}$$

In this work we are interested only on bounded trajectories around the MBH, i.e. trajectories such that the SCO radial coordinate ranges over a finite interval neither crossing the event horizon nor escaping to spatial infinity [Wilkins 1972]. In that case, the motion in the radial coordinate is periodic and as a consequence, the jump of Eq. (6.3.5), can be expanded as a discrete Fourier series:

$$\left[\partial_{r_*}\psi^{\ell m}\right]_p(t) = \sum_{n=-\infty}^{n=+\infty} c_{\ell m n}e^{-\mathrm{i}\omega_{n\ell m}t}\,, \tag{6.4.3}$$

with

$$\omega_{\ell m n} = n\omega_r + m\omega_\varphi \equiv \omega_{nm}\,, \tag{6.4.4}$$

where ω_r and ω_φ are the frequencies associated with the radial motion (going from periapsis to apoapsis and back) and the azimuthal motion (going 2π around the polar axis), respectively. We only need to sum over n because the only dependence of the jump on the ω_φ frequency comes from the spherical harmonics, in the term $\exp\{-\mathrm{i}m\varphi_p(t)\}$ [see Eqs. (6.3.4)-(6.3.5) and (6.2.10)]. Since we have bounded orbits, we can expand the fields $\psi^{\ell m}_{\pm}$ in discrete Fourier series as

$$\psi^{\ell m}_{\pm}(t,r) = e^{-\mathrm{i}m\omega_\varphi t} \sum_{n=-\infty}^{+\infty} e^{-\mathrm{i}n\omega_r t}\, R^{\pm}_{\ell mn}(r)\,. \tag{6.4.5}$$

This leads to ODEs for the components of the series, that is, for the functions $R_{\ell mn}(r)$[1]. These are [from Eq. (6.3.3)]:

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}r_*^2} - V_\ell(r) + \omega_{nm}^2\right) R^{\pm}_{\ell mn} = 0\,. \tag{6.4.6}$$

To complete the problem, we need boundary conditions. In this case the boundary conditions have to be prescribed both at the horizon ($r_* \to -\infty$) and at spatial infinity ($r_* \to +\infty$). The condition at spatial infinity is that the field has to be purely outgoing:

$$\left(\partial_t + \partial_{r_*}\right) \psi^{\ell m}_{+}\Big|_{r_*\to+\infty} = 0\,, \tag{6.4.7}$$

and at the horizon it has also be outgoing (ingoing from the point of view of spatial infinity):

$$\left(\partial_t - \partial_{r_*}\right) \psi^{\ell m}_{-}\Big|_{r_*\to-\infty} = 0\,. \tag{6.4.8}$$

These boundary conditions, in terms of the radial functions $R_{\ell mn}$ become:

$$\left(-\mathrm{i}\omega_{mn} + \frac{\mathrm{d}}{\mathrm{d}r_*}\right) R^{+}_{\ell mn}\Big|_{r_*\to+\infty} = 0\,, \tag{6.4.9}$$

$$\left(\mathrm{i}\omega_{mn} + \frac{\mathrm{d}}{\mathrm{d}r_*}\right) R^{-}_{\ell mn}\Big|_{r_*\to-\infty} = 0\,. \tag{6.4.10}$$

[1] In the continuum, when the spectrum of ω is not discrete as in this case where we have bounded orbits around a Schwarzschild black hole, the radial functions are denoted by $R_{\ell m\omega}$, so the corresponding notation should be $R_{\ell m\omega_{mn}}$ as in [Barack, Ori, and Sago 2008]. However, for the sake of simplicity we will use $R_{\ell mn}$.

These equations have to be solved simultaneously with the junction conditions for the radial functions $R^{\pm}_{\ell nm}$, which are given [from Eqs. (6.3.4)-(6.3.5) and (6.4.3)] by

$$\sum_{n=-\infty}^{+\infty} e^{-\mathrm{i}n\omega_r t}\, [R_{\ell mn}]_p = 0\,, \tag{6.4.11}$$

$$e^{-\mathrm{i}m\omega_\varphi t}\sum_{n=-\infty}^{+\infty} e^{-\mathrm{i}n\omega_r t}\left[\frac{\mathrm{d}R_{\ell mn}}{\mathrm{d}r_*}\right] = J^{\ell m}\,. \tag{6.4.12}$$

Note that for circular orbits, these reduce to $[R_{\ell mn}]_p = 0$ and $[\mathrm{d}R_{\ell mn}/\mathrm{d}r_*]_p = c_{\ell mn}$ respectively, where $c_{\ell mn}$ [determined from Eq. (6.3.5)] are just the Fourier components of the jump in the gradient of the scalar field at the particle location.

6.5. Numerical implementation and results

Now, let us discuss the strategy to solve the set of equations (6.4.6), (6.4.9), (6.4.10), and (6.4.11). We offer here only an outline of our procedure and omit entering fully into the technicalities.

The first ingredient we consider is the type of grid and how many domains would be adequate to use. The tortoise coordinate has an unbounded range, $r_* \in (-\infty, +\infty)$, and hence we either truncate the physical domain or we use some different coordinate system in which we cover the horizon and spatial infinity with a coordinate with a finite range. The first option is the most widely used solution in many problems. The drawback of this choice is that if we still use the same boundary conditions, Eqs. (6.4.7) and (6.4.8), we are making an error since these boundary conditions are not exact at a finite value of r_*. Of course, the error made will depend on how far from the particle location we truncate the domain. In a time-domain setup we can always choose the truncation locations in such a way that the boundaries remain out of causal contact with the particle, avoiding the contamination of the solution around the particle from the boundaries. But in the frequency domain we are solving elliptic equations, which care about the boundaries. A possible solution is to obtain precise boundary conditions from an expansion of our equation near the horizon and spatial infinity (see [Barack, Ori, and Sago 2008]).

However, here we are going to use the second option, that is, to use a compactified coordinate so that both the horizon and spatial infinity are located at finite values of the new coordinate. We are not going to use just a compactification of the radial coordinate, as this would solve the problem of the boundary conditions but would create another problem, namely that many cycles of the radiation would accumulate near the boundaries and it would not be possible to resolve them appropriately. A solution to this is to use a hyperboloidal compactification [Zenginoglu 2011], where we also change the time coordinate (the slicing of the spacetime in t = const. hypersurfaces) so that we avoid the problem just mentioned, since in the new slicing we will only have a few number of cycles in such a way that they can be resolvable numerically with a reasonable amount of computational resources. In particular, we follow essentially the method of [Zenginoglu 2011], with a multidomain splitting of our computational grid so that the hyperboloidal compactification is applied only to suitable boundary regions (extending from a certain point to the horizon and spatial infinity respectively).

In the frequency-domain, as we have seen, the ℓ-harmonics of the SF are found by decomposing the retarded field in a Fourier series with indices (ℓ, m, ω) (where the frequency ω in the case of bounded orbits can be labeled by m and an integer n, $\omega \equiv \omega_{mn}$). Then, summing over ω and m (i.e., over m and n in the case of bounded orbits) we obtain the different ℓ-harmonics of the SF. The main advantage, as discussed, is that the computation involves only ODEs. The drawback, however, is that it was found that for the case of eccentric orbits, the sum over ω (n) has bad convergence properties as a consequence of the discontinuities at the particle location. In practice, the problem is analogous to the well-known Gibbs phenomenon that arises in standard Fourier analysis. In [Barack, Ori, and Sago 2008], a solution to this problem was proposed; the key point of the method was to use the homogeneous solutions to construct the modes of the SF instead of the inhomogeneous ones, and hence was named the method of *extended homogeneous solutions*. The method leads to spectral convergence to the value of the SF modes.

Our implementation of a frequency-domain solver for the SF modes follows very close lines to the method of extended homogeneous solutions of [Barack, Ori, and Sago 2008]. Indeed, the PwP already works with homogeneous solutions since it eliminates the explicit presence of the particle in the equations by moving it to the boundary conditions across the interface between two domains (the method is designed in such a way that the particle is always located at the interface, even for eccentric orbits), *i.e.* the jump conditions. However, the method we will develop here has some differences with the method of extended homogeneous solutions. First of all, we use a multidomain splitting with a hyperboloidal compactification. But we are going to introduce the following modification from previous implementations of the PwP: Instead of using complementary domains as has always been done until now (including in the time domain), we are going to solve the homogeneous problems on the domains $\mathcal{D}_- = \{r_* | -\infty < r_* \leq r_*^{\rm apo}\}$ and $\mathcal{D}_+ = \{r_* | r_*^{\rm peri} \leq r_* < +\infty\}$ that have an overlap. In contrast with the complementary regions $\mathcal{R}_\pm$ (shown in Figure 6.1), the regions $\mathcal{D}_\pm$ have a nonempty intersection in general. Thus $\mathcal{D}_- \cap \mathcal{D}_+ = [r_*^{\rm peri}, r_*^{\rm apo}]$. Only in the circular case the two setups, the one based on disjoint regions as in Figure 6.1 and the one based on the regions $\mathcal{D}_\pm$, coincide in the sense that there is no intersection (or just a point, the particle location).

Then, we proceed by solving for the $R^\pm_{\ell mn}$ with arbitrary Dirichlet boundary conditions at the pericenter and apocenter respectively,

$$R^\pm_{\ell mn}(r_*^{\rm peri/apo}) = \lambda^\pm \,, \tag{6.5.1}$$

for some free (non-zero) choice of $\lambda^{\pm}$. Let us call the solutions thus obtained $\hat{R}^{\pm}_{\ell mn}$ on $\mathcal{D}_{\pm}$.

Now the question is how, from these solutions (with the particular boundary conditions that we have used), we can find the solutions that we are actually interested in (taking into account the presence of the particle). Here we are going to take advantage of the linearity of our problem. Given the solutions $\hat{R}^{\pm}_{\ell mn}$ of the problem described above, i.e. for a single Fourier mode, with boundary conditions (6.5.1), the solution for our actual problem (*i.e.* including the particle) will be:

$$R_{\ell mn}(r) = \begin{cases} C^{-}_{\ell mn}\hat{R}^{-}_{\ell mn}(r) & \text{if} \quad r < r_p(t)\,, \\ C^{+}_{\ell mn}\hat{R}^{+}_{\ell mn}(r) & \text{if} \quad r > r_p(t)\,. \end{cases} \tag{6.5.2}$$

where the coefficients $C^{\pm}_{\ell mn}$ are constants to be determined. What allows us to do this is the linearity of the equations, since by multiplying the solutions $\hat{R}^{\pm}_{\ell mn}$ of the two problems defined on $\mathcal{D}_{\pm}$ by a constant, we obtain again a solution of the same equations, just with different boundary conditions than (6.5.1). The coefficients $C^{\pm}_{\ell mn}$ are then determined uniquely by enforcing the jump conditions (6.4.11) and (6.4.12) across the particle location.

While our work on eccentric orbits is still in progress, we present some results on circular orbits, where the problem simplifies a bit as discussed (the $\mathcal{D}_{\pm}$ and $\mathcal{R}_{\pm}$ regions coincide). The results are obtained from codes developed either in Matlab or in Python, and are shown in Table 1.

TABLE 1. Numerical values of the components of the gradient of the regularized field (f_r) for circular orbits. Here, N is the number of collocation grid points used and $\ell_{\max}$ the highest ℓ-harmonic used in the summation. For reference, the values for a circular orbit at the last stable circular orbit ($r = 6\,M$) obtained, using frequency-domain methods, in Diaz-Rivera et al. 2004 was: 1.6772834×10^{-4}.

N	$\ell_{\max}$	f_r
50	20	$1.674125346413219 \times 10^{-4}$
50	30	$1.680135078016693 \times 10^{-4}$
80	20	$1.673179411442940 \times 10^{-4}$
80	30	$1.675719825341073 \times 10^{-4}$
80	40	$1.676358608421948 \times 10^{-4}$
80	50	$1.676586279346283 \times 10^{-4}$
80	60	$1.675397087197872 \times 10^{-4}$
100	20	$1.673697449614004 \times 10^{-4}$
100	30	$1.676237863251235 \times 10^{-4}$
100	40	$1.676876512268239 \times 10^{-4}$
100	50	$1.677108498387443 \times 10^{-4}$
150	80	$1.677232315338774 \times 10^{-4}$
200	20	$1.673570555862998 \times 10^{-4}$
200	40	$1.676749618219680 \times 10^{-4}$
200	50	$1.676981603683444 \times 10^{-4}$
200	60	$1.677085514921244 \times 10^{-4}$

CHAPTER 7

Conclusions

Conclusion summary. The theory of general relativity has now withstood its first century of existence, one at the end of which it has decisively opened the door to a new era in astronomy. The electromagnetic waves that once told us nearly everything we knew of the Universe beyond our Earth are now only one—albeit a still largely dominant—voice in the story. Gravitational waves have begun to tell their own story, the first pages of which are being written as we speak.

In this thesis, we have investigated two-body gravitational systems in the strong-field—that is, fully general relativistic—and extreme-mass-ratio regime, known as extreme-mass-ratio inspirals. These are expected to be among the main and most interesting sources of the future space-based gravitational-wave detector LISA. Prospective observations of such systems will furnish us with a wealth of opportunities to probe strong gravity, as the complicated orbits of the inspiraling object (stellar-mass black hole or neutron star) will effectively "map out" the gravitational field around the more massive one (the massive black hole at a galactic center). The problem of modeling such systems to sufficient accuracy—that is, for producing the theoretical waveform templates needed by LISA in its envisioned search for them—has witnessed significant progress over the last few decades, yet remains today an open one.

The understanding of this problem is intimately connected with concepts such as gravitational energy-momentum and mathematical techniques such as spacetime decompositions—for example, via canonical or quasilocal approaches—as well as perturbation theory. In the first half of this thesis, we have developed in detail the basic methods needed for dealing with these. In the second half, we have presented our novel contributions in these areas, notably on the issues of entropy, motion and the self-force in general relativity.

In what follows, we summarize briefly the results obtained in this work. This then leads us into offering some closing reflections on the broad conceptual issues that have historically been at the basis of the interpretation of general relativity. In view of the intrinsic dichotomy of the theory, as Einstein himself saw it, between "measuring rods and clocks [and] all other things", it is perhaps unsurprising that more subtle notions such as entropy, energy-momentum and the self-force continue to elude a clear consensus among relativists to this day. Our contributions in this thesis have sought to offer some fresh perspectives on these basic issues.

Conclusions (conclusion summary translation in Catalan). La teoria de la relativitat general ha viscut ara el seu primer segle d'existència, un al final del qual ha obert decisivament la porta a una nova era en l'astronomia. Les ones electromagnètiques que abans ens van dir gairebé tot el que sabíem de l'Univers fora de la nostra Terra són ara només una veu de la història. Les ones gravitacionals han començat a transmetre la seva pròpia història, les primeres pàgines de la qual estan sent escrites en aquests mateixos moments.

En aquesta tesi, hem investigat sistemes gravitacionals de dos cossos en el règim de camps forts - és a dir, en la teoria completa de la relativitat general - i raons de masses extremes (conegudes com a caigudes en espiral amb raó de masses extrema, *EMRIs*), que s'esperaven figurar entre les principals fonts del futur detector d'ones gravitacionals LISA, situada en l'espai. Les observacions possibles d'aquests sistemes ens proporcionaran una gran varietat d'oportunitats per provar la gravetat en el règim fort, ja que les òrbites complicades de l'objecte caient en espiral (un forat negre de massa estel·lar o una estrella de neutrons) realitzaran un "mapa" del camp gravitatori al voltant del masiu (un forat negre massiu d'un centre galàctic). El problema de modelar aquests sistemes amb una precisió suficient (és a dir, per produir les plantilles teòriques de formas d'onas necessàries per LISA en la seva cerca prevista) ha vist progressos significatius durant les últimes dècades, encara que avui en dia queda obert.

La comprensió d'aquest problema està íntimament relacionada amb conceptes com ara l'energia i la quantitat de moviment gravitatòria, i tècniques matemàtiques com les descomposicions de l'espai-temps - per exemple, mitjançant enfocaments canònics o quasilocals - i també amb la teoria de pertorbacions. En la primera meitat d'aquesta tesi, hem desenvolupat en detall els mètodes bàsics necessaris per tractar-los. En la segona meitat, hem presentat les nostres contribucions en aquestes àrees, en particular sobre els temes de l'entropia, el moviment i la força pròpia en la relativitat general.

A continuació, resumim breument els resultats obtinguts en aquest treball. Això ens porta a oferir algunes reflexions tancades sobre els grans temes conceptuals que històricament han estat a la base de la interpretació de la relativitat general. A la vista de la dicotomia intrínseca de la teoria, tal com ho va veure el mateix Einstein, entre "varetes i rellotges de mesurament [i] totes les altres coses", potser no és sorprenent que nocions més subtils com l'entropia, l'energia i la quantitat de moviment gravitatòria i la força pròpia actualment continuen eludint un consens clar entre els relativistes. Les nostres contribucions en aquesta tesi han buscat oferir algunes perspectives noves sobre aquests temes bàsics.

Conclusions (conclusion summary translation in French). La théorie de la relativité générale a maintenant traversé son premier siècle d'existence, à l'issue duquelle elle a ouvert de manière décisive la porte d'une nouvelle ère dans l'astronomie. Les ondes électromagnétiques qui nous disaient à peu près tout ce que nous savions de l'univers en dehors de notre Terre ne sont plus qu'une voix dans l'histoire, même si les différents fréquences de la lumière restent le messager principal aujourd'hui. Les ondes gravitationnelles ont commencé à transmettre leur propre histoire, dont les premières pages sont tout de suite en cours d'écriture.

Dans cette thèse, nous avons étudié les systèmes gravitationnels à deux corps dans le régime des champs forts - c'est-à-dire, dans la théorie complète de la relativité générale - et les quotients extrêmes des masses (appelés inspirals avec quotients extrêmes des masses, *EMRIs*), qui devraient être parmi les principales sources du futur détecteur spatial d'ondes gravitationnelles LISA. Les observations prospectives de tels systèmes nous fourniront une grande variété de posibilités pour tester la gravité forte, car les orbites compliquées de l'objet spirallant (un trou noir à masse stellaire ou une étoile à neutrons) « cartographieront » effectivement le champ gravitationnel autour du plus massif (un trou noir massif au centre galactique). Le problème de la modélisation de tels systèmes avec une précision suffisante - c'est-à-dire pour la production des modèles de formes des ondes théoriques requis par LISA dans sa recherche envisagée - a connu des progrès significatifs au cours des dernières décennies, mais reste aujourd'hui ouvert.

La compréhension de ce problème est intimement liée aux concepts tels que l'énergie et la quantité de mouvement gravitationnelles et les techniques mathématiques telles que les décompositions de l'espace-temps - par exemple, en usant des approches canoniques ou quasi-locales - ainsi que la théorie des perturbations. Dans la première partie de cette thèse, nous avons développé en détail les méthodes de base nécessaires pour y faire face. Dans la deuxième partie, nous avons présenté nos nouvelles contributions dans ces domaines, en particulier sur les problèmes de l'entropie, du mouvement et de la force propre dans la relativité générale.

Dans ce qui suit, nous résumons brièvement les résultats obtenus dans ce travail. Cela nous amène ensuite à proposer des réflexions finales sur les grandes questions conceptuelles qui ont toujours été à la base de l'interprétation de la relativité générale. Compte tenu de la dichotomie intrinsèque de la théorie, telle que l'a vue Einstein lui-même, entre "bâtonnets de mesure et horloges [et] tout le reste", il n'est peut-être pas surprenant que des notions plus subtiles telles que l'entropie, l'énergie et la quantité de mouvement gravitationnelles et la force propre continuent à éluder un consensus clair parmi les relativistes à ce jour. Nos contributions dans cette thèse ont cherché à offrir de nouvelles perspectives sur ces questions fondamentales.

*

We now offer a brief concluding summary of the novel contributions of this thesis.

In Chapter 4, we have studied entropy theorems in classical mechanics and general relativity, with a focus on the gravitational two-body problem. In particular, we have proved that canonical theories of classical particles for certain classes of Hamiltonians, as well as of some typical matter (in particular, scalar and electromagnetic) fields in curved spacetime, do not admit any monotonically increasing function of phase space (along trajectories of the Hamiltonian flow). Thus, such theories preclude the existence of entropy in what we have referred to as a "mechanical" sense, i.e. as a phase space functional. We have then looked at why these proofs do not carry over to general relativity, which we do know to manifest the existence of entropy in such a sense. We have furthermore discussed another method of proof based on a topological argument, in particular, phase space compactness, and have investigated the meaning of these results for the gravitational two-body problem, in particular, by proving the non-compactness of the phase space of perturbed Schwarzschild-Droste spacetimes. In the absence today of a general formula for gravitational entropy, an understanding of why general relativity differs from, for example, classical mechanics or Maxwellian electromagnetism in this sense can give helpful indications for future progress.

In Chapter 5, we have presented a novel derivation, based on conservation laws, of the basic equations of motion for the EMRI problem. They are formulated with the use of a quasilocal (rather than matter) stress-energy-momentum tensor—in particular, the Brown-York tensor—so as to capture gravitational effects in the momentum flux of the object, including the gravitational self-force. Our formulation and resulting equations of motion are independent of the choice of the perturbative gauge. We have shown that, in addition to the usual gravitational self-force term, they also lead to an additional "self-pressure" force not found in previous analyses, and also that our results correctly recover known formulas under appropriate conditions. Our approach thus offers a fresh geometrical picture from which to understand the self-force fundamentally, and potentially useful new avenues for computing it practically.

In Chapter 6, we have presented some numerical work based on the Particle-without-Particle (PwP) approach, a pseudospectral collocation method previously developed for the computation of the scalar self-force—a helpful testbed for the gravitational case. The basic idea of this method is to discretize the computational domain into two (or more) disjoint grids such that the "particle"—the distributional source in the field equations of the self-force problem—is always at the interface between them; thus, one only needs to solve homogeneous equations in each domain, with the source effectively replaced by jump (boundary) conditions thereon. Here we have presented some results on the numerical computation of the scalar self-force, using this method, for circular orbits in the frequency domain. Moreover, in Appendix B, we present a generalization of this method to general partial differential equations with distributional sources, including also applications to other areas of applied mathematics. We generically obtain improved convergence rates relative to other implementations in these areas, typically relying on delta function approximations on the computational grid.

*

As we have seen, the EMRI problem is intimately connected with conceptual as well as technical questions regarding entropy, energy-momentum and motion in general relativity. It is remarkable that, despite the multiplicity of fruitful insights which have so far been achieved towards their understanding, relativists today continue to lack a clear, general consensus on the conceptual interpretation and, strictly speaking, even the formal mathematical expression of such notions.

These considerations naturally invite us to reflect back upon our discussion in the introduction, specifically regarding the interpretation of general relativity and more generally the evolution of our ideas about gravitation in physics.

It may be argued that, with regard to the basic content of his theory, Einstein's key physical insight was to realize what sort of object it should be that the gravitational field equations describe (in particular, the metric tensor of spacetime, or something like it), much more so, in a certain sense, than eventually obtaining the exact final form of these equations—an effort which relied essentially on mathematical reasoning and consistency with the Newtonian theory once the spacetime geometry was understood to be the basic object of study.

There is a simple *gedankenexperiment* that Einstein frequently used to illustrate how the local effects of special relativity plus the requirement that physical laws be formulated in any coordinate frame of reference together logically imply that our spacetime must be, in general, globally curved. It is worthwhile to recount it here, from his 1921 lecture series [Einstein 1922] (taken from [Einstein 2002]):

> [Let K be an inertial coordinate system, with spatial Cartesian coordinates x, y, z.] Imagine a circle drawn about the origin in the [Cartesian] $x'y'$ plane of [another coordinate system] K' [the z' axis of which coincides with the z axis of K], and a diameter of this circle. Imagine, further, that we have given a large number of rigid rods, all equal to each other. We suppose these laid in series along the periphery and the diameter of the circle, at rest relatively to K'. If U is the number of these rods along the periphery, D the number along the diameter, then, if K' does not rotate relatively to K we shall have
>
> $$\frac{U}{D} = \pi \, .$$
>
> But if K' rotates we get a different result. Suppose that at a definite time t, of K we determine the ends of all the rods. With respect to K all the rods upon the periphery experience the Lorentz contraction, but the rods upon the diameter do not experience this contraction (along their lengths!).* It therefore follows

> that
> $$\frac{U}{D} > \pi \, .$$
> [...] Space and time, therefore, cannot be defined with respect to K' as they were in the special theory of relativity with respect to inertial systems. But, according to the principle of equivalence, K' may also be considered as a system at rest, with respect to which there is a gravitational field (field of centrifugal force, and force of Coriolis). We therefore arrive at the result: the gravitational field influences and even determines the metrical laws of the space-time continuum. If the laws of configuration of ideal rigid bodies are to be expressed geometrically, then in the presence of a gravitational field the geometry is not Euclidean.
>
> ---
>
> * These considerations assume that the behavior of rods and clocks depends only upon velocities, and not upon accelerations, or, at least, that the influence of acceleration does not counteract that of velocity.

Notwithstanding the elegant simplicity of the above argument in capturing the essence of the motivation for general relativity (that is, for devising a theory of gravitation in terms of spacetime geometry), Einstein was certainly aware that there is greater subtlety here than first meets the eye. Referring in particular to special relativity, he writes in his *Autobiographical Notes* [Einstein 1949]:

> One is struck [by the fact] that the theory (except for the four-dimensional space) introduces two kinds of physical things, i.e., (1) measuring rods and clocks, (2) all other things, e.g., the electro-magnetic field, the material point, etc. This, in a certain sense, is inconsistent; strictly speaking measuring rods and clocks would have to be represented as solutions of the basic equations (objects consisting of moving atomic configurations), not, as it were, as theoretically self-sufficient entities. However, the procedure justifies itself because it was clear from the very beginning that the postulates of the theory are not strong enough to deduce from them sufficiently complete equations for physical events sufficiently free from arbitrariness, in order to base upon such a foundation a theory of measuring rods and clocks.

It would be fair to say that this basic dichotomy in the physical foundations of the theory—between "measuring rods and clocks" and "all other things"—has never been fully resolved, at least not at the level that Einstein would have regarded as "sufficiently complete" within this discussion. Nevertheless, this sort of "inherent contradiction" of general relativity, if one is inclined to regard it as such, is one which he certainly saw, at the very least, as a reasonable exchange for the Newtonian ones it has come to replace. With what Arthur Koestler might have called "sleepwalker's assurance", Einstein writes a few years after the discovery of general relativity [Einstein 1921] (English translation taken from [Goenner et al. 1999]):

> The concept of the measuring-rod and the concept of the clock coordinated with it in the theory of relativity do not find an exactly corresponding object in the real world [there are no perfectly rigid rods]. It is also clear that the solid body and clock do not play the role of irreducible elements in the conceptual edifice of physics, but that of composite structures, which may not play any independent role in theoretical physics. But it is my conviction that in the present stage of development of theoretical physics these concepts must still be employed as independent concepts; for we are still far from possessing such certain knowledge of the theoretical foundations as to be able to give theoretical constructions of such structures.

Early in our introduction to this thesis, we briefly traced Johannes Kepler's struggle with the idea of the "force" governing planetary motion in his emerging vision of a clockwork universe [Koestler 1959]. At that time, he could do no better than to visualize it as a sort of vortex, "a raging current which tears all the planets, and perhaps all the celestial ether, from West to East" [Kepler 1609]; it took the arrival of Newton for this concept to encounter its first clear formulation. Today, our technically-advanced struggle with the "self-force" and related concepts should not obscure the fact that we are still, to a large extent, following in the inertia of sleepwalking—not least evinced by the manifestly neo-Newtonian nomenclature to which we still stubbornly cling. We may take with a good dose of welcome encouragement Arthur Koestler's remark [Koestler 1959] that "[t]he contemporary physicist grappling with the paradoxa of relativity and quantum mechanics will find [in Kepler's own struggle] an echo of his perplexities".

APPENDIX A

Topics in Differential Geometry: Maps on Manifolds, Lie Derivatives, Forms and Integration

Appendix summary. In this appendix, $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ is any n-dimensional (oriented, smooth, topological [Lee 2002]) manifold of any signature, with metric $\boldsymbol{g}$ and compatible derivative $\boldsymbol{\nabla}$.

We define and develop here four broad geometrical notions used amply throughout this thesis: maps on manifolds in Section A.1, Lie derivatives in Section A.2, differential forms in Section A.3, and finally integration on manifolds in Section A.4. At the end of each of these sections we offer a brief example from physics. The exposition is mainly based on Appendices B and C of [Wald 1984] and [Lee 2002].

Temes en geometria diferencial (appendix summary translation in Catalan). En aquest apèndix, $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ és qualsevol varietat n-dimensional (orientada, suau, topològica [Lee 2002]) de qualsevol signatura, amb $\boldsymbol{g}$ el tensor mètric i derivada compatible $\boldsymbol{\nabla}$.

Definim i desenvolupem aquí quatre nocions geomètriques àmplies que s'utilitzen de forma extensiva al llarg de aquesta tesi: fonctions sobre varietats a la secció A.1, derivats de Lie a la secció A.2, formes diferencials de la secció A.3 i, finalment, integració sobre varietats de la secció A.4. Al final de cadascuna d'aquestes seccions oferim un breu exemple de física. L'exposició es basa principalment en els apèndixs B i C de [Wald 1984] i [Lee 2002].

Sujets dans la géométrie différentielle (appendix summary translation in French). Dans cette annexe, $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ nous décrivons n'importe quelle variété n-dimensionnelle (orientée, lisse, topologique [Lee 2002]) de n'importe quelle signature, avec $\boldsymbol{g}$ le tenseur métrique et dérivé compatible $\boldsymbol{\nabla}$.

Nous définissons et développons ici quatre grandes notions géométriques largement utilisées tout au long de cette thèse : applications sur le variétés dans la section A.1, dérivées de Lie dans la section A.2, formes différentielles dans la section A.3 et enfin intégration sur variétés dans la section A.4. À la fin de chacune de ces sections, nous proposons un bref exemple tiré de la physique. L'exposition est principalement basée sur les annexes B et C de [Wald 1984] et [Lee 2002].

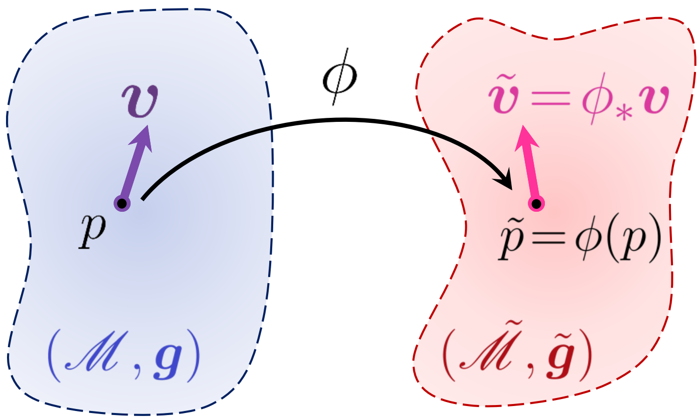

FIGURE A.1. An illustration of two manifolds $(\mathscr{M}, \boldsymbol{g})$ and $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ with a map $\phi : \mathscr{M} \to \tilde{\mathscr{M}}$ between them. This identifies any point $p \in \mathscr{M}$ with $\tilde{p} = \phi(p) \in \tilde{\mathscr{M}}$, and can be used, for example, to push-forward the vector $\boldsymbol{v} \in T_p\mathscr{M}$ to $\phi_*\boldsymbol{v} \in T_{\tilde{p}}\tilde{\mathscr{M}}$. If ϕ is a diffeomorphism, a general transport of tensors from one manifold to the other can be defined. Note that in this notation, the metric $\tilde{\boldsymbol{g}}$ of $\tilde{\mathscr{M}}$ is not necessarily the same as the metric transported from $\mathscr{M}$, i.e. $\phi_*\boldsymbol{g}$. If indeed $\phi_*\boldsymbol{g} = \tilde{\boldsymbol{g}}$, then ϕ is called an *isometry*—a symmetry of the metric.

A.1. Maps on manifolds

Let $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ be an n-dimensional manifold and $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{\nabla}})$ an $\tilde{n}$-dimensional manifold. They could be of the same dimension, and could even be the same manifold, but not necessarily.

An important question, one that often arises in physics and especially in GR, is how to establish an identification of points between manifolds (or between points on the same manifold), i.e. how to relate a point $p \in \mathscr{M}$ with a point $\tilde{p} \in \tilde{\mathscr{M}}$, and more generally, an arbitrary tensor $\boldsymbol{A} \in \mathscr{T}^k{}_l(\mathscr{M})$ at $p \in \mathscr{M}$ with another tensor $\tilde{\boldsymbol{A}} \in \mathscr{T}^k{}_l(\tilde{\mathscr{M}})$ at $\tilde{p} \in \tilde{\mathscr{M}}$. In this section, overset tildes will generally be used to indicate objects living on $\tilde{\mathscr{M}}$.

First, in order to identify the points themselves, we suppose that there exists a smooth map between these manifolds,

$$\phi : \mathscr{M} \to \tilde{\mathscr{M}} \tag{A.1.1}$$

$$p \mapsto \phi(p) = \tilde{p}, \tag{A.1.2}$$

such that any point $p \in \mathscr{M}$ is identified with its image under this map, $\tilde{p} = \phi(p)$. See Fig. A.1.

Consider now any function $\tilde{f} : \tilde{\mathscr{M}} \to \mathbb{R}$. The map $\phi : \mathscr{M} \to \tilde{\mathscr{M}}$ can be used to define a new function f on $\mathscr{M}$, referred to as the *pull-back* of $\tilde{f}$, via simple composition. We denote this as $f = \phi^* \tilde{f} : \mathscr{M} \to \mathbb{R}$, and it is simply given by:

$$\phi^* \tilde{f} = \tilde{f} \circ \phi \,. \tag{A.1.3}$$

This idea can be extended from functions to tensors of higher rank. First, suppose $v^a \in T\mathscr{M}$ is a vector field in the tangent bundle of $\mathscr{M}$. The map ϕ can be used to "carry along" this vector field to another vector field $\tilde{v}^a = \phi_* v^a \in T\tilde{\mathscr{M}}$ in the tangent bundle of $\tilde{\mathscr{M}}$, from point to point, via a map called the *push-forward*, $\phi_* : T_p\mathscr{M} \to T_{\tilde{p}}\tilde{\mathscr{M}}$. Its action is defined by

$$(\phi_* \boldsymbol{v})\,(\tilde{f}) = \boldsymbol{v}(\phi^* \tilde{f})\,, \tag{A.1.4}$$

for any function $\tilde{f} : \tilde{\mathscr{M}} \to \mathbb{R}$, with its pull-back $\phi^* \tilde{f} = \tilde{f} \circ \phi$ as given by (A.1.3). See again Fig. A.1. One can push-forward vector fields from $T\tilde{\mathscr{M}}$ to $T\mathscr{M}$ in the same way using instead the inverse map $\phi^{-1} : \tilde{\mathscr{M}} \to \mathscr{M}$ if it exists.

One can continue in this way to define the pull-back of co-vectors ($(0,1)$-tensors) from the co-tangent bundle of $\tilde{\mathscr{M}}$ to that of $\mathscr{M}$. Let $\tilde{w}_a \in T^*\tilde{\mathscr{M}}$. Then the pull-back $\phi^* : T^*_{\tilde{p}}\tilde{\mathscr{M}} \to T^*_p\mathscr{M}$ is defined by

$$\left(\phi^* \tilde{\boldsymbol{w}}\right)_a v^a = \tilde{w}_a \left(\phi_* \boldsymbol{v}\right)^a \,. \tag{A.1.5}$$

Suppose now that $\phi : \mathscr{M} \to \tilde{\mathscr{M}}$ is a *diffeomorphism*, meaning that it is bijective (one-to-one and onto) and has a smooth inverse $\phi^{-1} : \tilde{\mathscr{M}} \to \mathscr{M}$. Then a generalization of the push-forward and pull-back maps can be defined to relate an arbitrary tensor $\boldsymbol{A} \in \mathscr{T}^k{}_l(\mathscr{M})$ at $p \in \mathscr{M}$ with another tensor $\tilde{\boldsymbol{A}} \in \mathscr{T}^k{}_l(\tilde{\mathscr{M}})$ at $\tilde{p} \in \tilde{\mathscr{M}}$. In particular, we write $\tilde{\boldsymbol{A}} = \phi_* \boldsymbol{A}$ and refer to this generally as the *transport* of $\boldsymbol{A}$ (under ϕ) from $\mathscr{M}$ to $\tilde{\mathscr{M}}$. (Note that in this notation, the metric $\tilde{\boldsymbol{g}}$ of $\tilde{\mathscr{M}}$ is not necessarily $\phi_* \boldsymbol{g}$. If it is, then ϕ is called an *isometry*—a symmetry of the metric.) For any set of k co-vectors $\{\tilde{w}^{(j)}_a\}^k_{j=1}$ in $T^*\tilde{\mathscr{M}}$ and any set of l vectors $\{\tilde{v}^a_{(j)}\}^l_{j=1}$ in $T\tilde{\mathscr{M}}$, the transport map

$$\phi_* : \left(\mathscr{T}^k{}_l\,(\mathscr{M})\right)_p \to \left(\mathscr{T}^k{}_l(\tilde{\mathscr{M}})\right)_{\tilde{p}} \tag{A.1.6}$$

is defined by

$$(\phi_* \boldsymbol{A})^{a_1 \cdots a_k}{}_{b_1 \cdots b_l} \tilde{w}^{(1)}_{a_1} \cdots \tilde{w}^{(k)}_{a_k} \tilde{v}^{b_1}_{(1)} \cdots \tilde{v}^{b_l}_{(l)} = A^{a_1 \cdots a_k}{}_{b_1 \cdots b_l} w^{(1)}_{a_1} \cdots v^{b_l}_{(l)}\,, \tag{A.1.7}$$

where $\boldsymbol{w}^{(j)} = \phi^* \tilde{\boldsymbol{w}}^{(j)}$ is the pull-back (under ϕ) of each co-vector and $\boldsymbol{v}_{(j)} = (\phi^{-1})_* \tilde{\boldsymbol{v}}_{(j)}$ the push-forward (under ϕ^{-1}) of each vector, from $\tilde{\mathscr{M}}$ to $\mathscr{M}$.

The transport of any tensor from $\tilde{\mathscr{M}}$ to $\mathscr{M}$, i.e. the transport under ϕ^{-1}, is denoted by super-scripting the star,

$$\left(\phi^{-1}\right)_* = \phi^* : \left(\mathscr{T}^k{}_l(\tilde{\mathscr{M}})\right)_{\tilde{p}} \to \left(\mathscr{T}^k{}_l\,(\mathscr{M})\right)_p \,. \tag{A.1.8}$$

We now enumerate some useful properties of the tensor transport (A.1.6). See Theorem 10.6 of [Felsager 2012]. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathscr{T}^k{}_l(\mathscr{M})$ and $\boldsymbol{B}' \in \mathscr{T}^{k'}{}_{l'}(\mathscr{M})$ be any tensors in $\mathscr{M}$ and $c \in \mathbb{R}$ a constant. Then we have the following:

(1) ϕ_* *is linear*, i.e.

$$\phi_* (\boldsymbol{A} + \boldsymbol{B}) = \phi_* \boldsymbol{A} + \phi_* \boldsymbol{B} , \quad \phi_* (c\boldsymbol{A}) = c\phi_* \boldsymbol{A} . \tag{A.1.9}$$

(2) ϕ_* *commutes with the tensor product*, i.e.

$$\phi_* \left(\boldsymbol{A} \otimes \boldsymbol{B}'\right) = (\phi_* \boldsymbol{A}) \otimes \left(\phi_* \boldsymbol{B}'\right) . \tag{A.1.10}$$

(3) ϕ_* *commutes with contractions*, i.e.

$$\phi_* \left(A^{a_1 \cdots c \cdots a_{k-1}}{}_{b_1 \cdots c \cdots b_{l-1}}\right) = (\phi_* \boldsymbol{A})^{a_1 \cdots c \cdots a_{k-1}}{}_{b_1 \cdots c \cdots b_{l-1}} .$$

Example: gauge freedom in GR. If the two manifolds $(\mathscr{M}, \boldsymbol{g})$ and $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ are (four-dimensional, Lorentzian) spacetimes, the existence of a diffeomorphism $\phi : \mathscr{M} \to \tilde{\mathscr{M}}$ is interpreted as signifying that the spacetimes describe *the same physical situation*. In other words, any solution of the field equations of a theory for some collection of fields ψ is considered to be physically indistinguishable from the solution $\phi_*\psi$. (Thus, we may speak of an equivalence class of solutions with the equivalence relation $\psi \sim \phi_*\psi$.) Conversely, if there exists no diffeomorphism $\phi : \mathscr{M} \to \tilde{\mathscr{M}}$, then the two spacetimes (and the correspondent solutions to the field equations thereon) are seen to represent physically different situations.

The existence in GR of the freedom to transform the spacetime metric $\boldsymbol{g}$ by a diffeomorphism, (such that $\boldsymbol{g} \sim \tilde{\boldsymbol{g}}$), is often referred to as the "active view" of gauge freedom. Equivalently, one may take the "passive view", where gauge freedom can be seen to manifest itself as coordinate transformations. Concretely, suppose $\{x^\alpha\}$ is a coordinate system covering a neighborhood $\mathscr{U}$ of a point $p \in \mathscr{M}$, and $\{y^\alpha\}$ one covering a neighborhood $\mathscr{V}$ of $\tilde{p} = \phi(p) \in \tilde{\mathscr{M}}$. One can then define a new coordinate system $\{x'^\alpha\}$ in a neighborhood $\phi^{-1}(\mathscr{V})$ of $p \in \mathscr{M}$ by setting $x'^\alpha(q) = y^\alpha(\phi(q))$, for all $q \in \phi^{-1}(\mathscr{V})$. From this point of view, one may thus regard the effect of ϕ as leaving p and all tensors at p unchanged, but instead inducing a local coordinate transformation $x^\alpha \mapsto x'^\alpha$. In other words, the components of any $\phi_*\boldsymbol{A}$ at $\tilde{p} = \phi(p)$ in the coordinates $\{y^\alpha\}$ (in the "active" viewpoint) are the same as those of $\boldsymbol{A}$ at p in the coordinates $\{x'^\alpha\}$ (in the "passive" viewpoint).

A.2. Lie derivatives

Let $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ be any manifold, and let $v^a \in T\mathscr{M}$ be any vector field. An important question to address is: how do tensors change "in the direction" of $\boldsymbol{v}$? Or, more precisely, how do they change (from point to point) along the curves in $\mathscr{M}$ to which $\boldsymbol{v}$ is tangent? Firstly, it is necessary to formalize the meaning of the latter concept: the set of curves in $\mathscr{M}$ to which $\boldsymbol{v}$ is tangent are referred to as the *integral curves* of $\boldsymbol{v}$. These are defined by a *one-parameter group* of diffeomorphisms in $\mathscr{M}$, referred to as the *flow* of $\boldsymbol{v}$,

$$\phi_t^{(\boldsymbol{v})} : \mathscr{M} \times \mathbb{R} \to \mathscr{M} \,, \tag{A.2.1}$$

which are solutions to the ODE

$$\frac{\mathrm{d}\phi_t^{(\boldsymbol{v})}}{\mathrm{d}t} = \boldsymbol{v} \circ \phi_t^{(\boldsymbol{v})} \,. \tag{A.2.2}$$

In this case, $\boldsymbol{v}$ is referred to as the *generator* of the flow.

Thanks to our discussion in the previous section, we have a precise way of "comparing tensors" at different points on a manifold. In particular, we can compare the values of tensors at different points along the integral curves of $\boldsymbol{v}$ simply by transporting them under the flow $\phi_t^{(\boldsymbol{v})}$.

To be more precise, let $\boldsymbol{A}$ be any (k, l)-tensor. We may ask, for example, how its value at a point $p_0 \in \mathscr{M}$ on the manifold corresponding to $t = 0$ in the flow parametrization changes relative that at a point $\phi_t^{(\boldsymbol{v})}(p_0) = p_t \in \mathscr{M}$ at some parameter value $t > 0$. In this case, one needs to compare $\boldsymbol{A}$ at p_0 with the transport of $\boldsymbol{A}$ from p_t to p_0, i.e. with $((\phi_t^{(\boldsymbol{v})})^{-1})_* \boldsymbol{A} = (\phi_{-t}^{(\boldsymbol{v})})_* \boldsymbol{A}$. See Fig. A.2. The limit in which t is small gives rise precisely to the notion of the *Lie derivative*,

$$\mathcal{L}_{\boldsymbol{v}} \boldsymbol{A} = \lim_{t \to 0} \frac{(\phi_{-t}^{(\boldsymbol{v})})_* \boldsymbol{A} - \boldsymbol{A}}{t} \,. \tag{A.2.3}$$

To make this more concrete, notice firstly that when applied to a function $\boldsymbol{A} = f : \mathscr{M} \to \mathbb{R}$, (A.2.3) immediately recovers the usual notion of the "directional derivative",

$$\mathcal{L}_{\boldsymbol{v}} f = \boldsymbol{v}\,(f) \,. \tag{A.2.4}$$

Moreover, for any $\boldsymbol{A}$, it is to instructive to consider the result of the formula (A.2.3) in a choice of coordinates $\{x^\alpha\}$ adapted to $\boldsymbol{v}$. This means that the action of ϕ_t on a point corresponds to a transformation in one coordinate $x^1 \to x^1 + t$ with the rest $x^2, \ldots, x^n$ being held fixed. In such a coordinate system, it is possible to show (see Appendix C of [Wald 1984]) that (A.2.3) yields:

$$\mathcal{L}_{\boldsymbol{v}} A^{\alpha_1 \cdots \alpha_k}{}_{\beta_1 \cdots \beta_l} = \frac{\partial A^{\alpha_1 \cdots \alpha_k}{}_{\beta_1 \cdots \beta_l}}{\partial x^1} \,, \tag{A.2.5}$$

again recovering the notion of a general "directional derivative", expressed here in the coordinates adapted to the direction of change. From (A.2.5), it is then possible (see Appendix C of [Wald 1984]) to obtain a completely general, *coordinate-independent* abstract

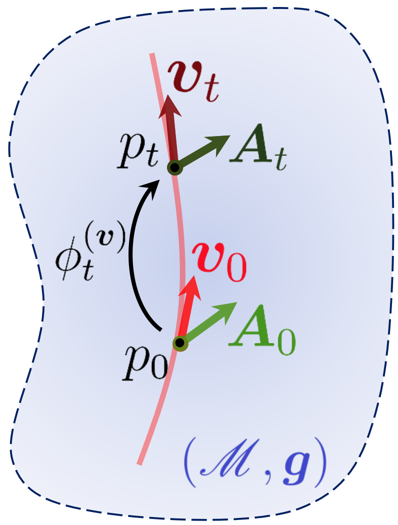

FIGURE A.2. An illustration of the meaning of the Lie derivative along a vector field $\boldsymbol{v}$ in $(\mathscr{M}, \boldsymbol{g})$ of a tensor $\boldsymbol{A}$, depicted here for ease of visualization in the case where $\boldsymbol{A}$ is a vector. In particular, one compares $\boldsymbol{A}_0 = (\boldsymbol{A})_{p_0}$ at the point p_0 corresponding to $t = 0$ in the flow $\phi_t^{(\boldsymbol{v})}$ with its value $\boldsymbol{A}_t = (\boldsymbol{A})_{p_t}$ at the point $p_t = \phi_t^{(\boldsymbol{v})}(p_0)$ for some $t > 0$ by transporting the latter back to p_0, i.e. one compares $\boldsymbol{A}_0$ with $(\phi_{-t}^{(\boldsymbol{v})})_* \boldsymbol{A}$ at p_0. Their difference divided by t, in the small t limit, is the Lie derivative.

index formula for (A.2.3),

$$\mathcal{L}_{\boldsymbol{v}} A^{a_1 \cdots a_k}{}_{b_1 \cdots b_l} = \nabla_{\boldsymbol{v}} A^{a_1 \cdots a_k}{}_{b_1 \cdots b_l} - \sum_{i=1}^{k} A^{a_1 \cdots c \cdots a_k}{}_{b_1 \cdots b_l} \nabla_c v^{a_i} + \sum_{j=1}^{l} A^{a_1 \cdots a_k}{}_{b_1 \cdots c \cdots b_l} \nabla_{b_j} v^c \,. \tag{A.2.6}$$

Example: perturbative gauge freedom and Killing vectors in GR. As developed in Chapter 3, a perturbative gauge transformation generated by a vector field $\boldsymbol{\xi} \in T\mathscr{M}$ changes the linear perturbation $\boldsymbol{h}$ according to $\boldsymbol{h} \mapsto \boldsymbol{h} + \mathcal{L}_{\boldsymbol{\xi}} \mathring{\boldsymbol{g}}$, where $\mathring{\boldsymbol{g}}$ is the background metric. Using (A.2.6), we see that

$$\mathcal{L}_{\boldsymbol{\xi}} \mathring{g}_{ab} = \xi^c \mathring{\nabla}_c \mathring{g}_{ab} + \mathring{g}_{cb} \mathring{\nabla}_a \xi^c + \mathring{g}_{ac} \mathring{\nabla}_b \xi^c \tag{A.2.7}$$

$$= 2 \mathring{\nabla}_{(a} \xi_{b)} \,. \tag{A.2.8}$$

The Lie derivative is also used to define *Killing vector fields* in any manifold $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ as those vector fields $\boldsymbol{\Xi} \in T\mathscr{M}$ satisfying the *Killing equation*, $\mathcal{L}_{\boldsymbol{\Xi}}\boldsymbol{g} = 0$. This is equivalent to the statement that the flow generated by such vector fields, $\phi_t^{(\boldsymbol{\Xi})}$, are *isometries* of the metric, i.e. $(\phi_t^{(\boldsymbol{\Xi})})_*\boldsymbol{g} = \boldsymbol{g}$.

A.3. Differential forms

Let $(\mathscr{M}, \boldsymbol{g}, \boldsymbol{\nabla})$ be any n-dimensional manifold (of any signature). If a $(0,k)$-tensor $\boldsymbol{\alpha} \in \mathscr{T}^0{}_k(\mathscr{M})$ is totally antisymmetric, i.e. if

$$\alpha_{a_1 \cdots a_k} = \alpha_{[a_1 \cdots a_k]}\,, \tag{A.3.1}$$

then $\boldsymbol{\alpha}$ is referred to as a (*differential*) *k-form*. The notion of a form is a crucial one in geometry, and was first developed in the work of [Cartan 1899]. As we shall see in the following section, forms serve as the basis for defining integration over (regions of) manifolds. For the remainder of the current section, we summarize some useful definitions and properties.

The set of k-forms on $\mathscr{M}$ is typically denoted $\Lambda^k(\mathscr{M})$. As the simplest examples, $\Lambda^0(\mathscr{M}) = \mathscr{F}(\mathscr{M})$ is the set of smooth functions on $\mathscr{M}$, and $\Lambda^1(\mathscr{M}) = T^*\mathscr{M}$ is just the cotangent bundle. Any k-form for $k > n$ vanishes identically due to the antisymmetry.

A useful operation between forms is the *wedge product*, $\wedge$. It can be applied between any k-form $\boldsymbol{\alpha}$ and any l-form $\boldsymbol{\beta}$ to produce a $(k+l)$-form $\boldsymbol{\alpha}\wedge\boldsymbol{\beta}$, given by their completely antisymmetrized outer product. That is,

$$\wedge : \Lambda^k(\mathscr{M}) \times \Lambda^l(\mathscr{M}) \to \Lambda^{k+l}(\mathscr{M}) \tag{A.3.2}$$

$$(\alpha_{a_1\cdots a_k}, \beta_{b_1\cdots b_l}) \mapsto (\boldsymbol{\alpha}\wedge\boldsymbol{\beta})_{a_1\cdots a_k b_1\cdots b_l} = \frac{(k+l)!}{k!l!}\alpha_{[a_1\cdots a_k}\beta_{b_1\cdots b_l]}\,. \tag{A.3.3}$$

This definition implies $\boldsymbol{\alpha}\wedge\boldsymbol{\beta} = (-1)^{kl}\boldsymbol{\beta}\wedge\boldsymbol{\alpha}$. (So in particular, $\boldsymbol{\alpha}\wedge\boldsymbol{\alpha} = 0$ for k odd.)

Another useful operation is the *exterior derivative* d, which takes k-forms to $(k+1)$-forms, defined as their completely antisymmetrized derivative. That is,

$$\mathrm{d} : \Lambda^k(\mathscr{M}) \to \Lambda^{k+1}(\mathscr{M}) \tag{A.3.4}$$

$$\alpha_{a_1\cdots a_k} \mapsto (\mathrm{d}\boldsymbol{\alpha})_{a_0 a_1\cdots a_k} = (k+1)\partial_{[a_0}\alpha_{a_1\cdots a_k]}\,. \tag{A.3.5}$$

In fact, due to the antisymmetry of forms and the symmetry of the connection coefficient (between any two derivative operators on $\mathscr{M}$), one can show that the above definition is independent of the choice of the derivative operator. (So we simply write it in terms of the partial derivative $\boldsymbol{\partial}$.)

Thus, given a basis $\{\mathrm{d}x^\alpha\}$ of the cotangent space $T_p^*\mathscr{M}$ at any point $p \in \mathscr{M}$, any k-form $\boldsymbol{\alpha}$ and its exterior derivative $\mathrm{d}\boldsymbol{\alpha}$ can locally always be written, respectively, as

$$\boldsymbol{\alpha} = \alpha_{\alpha_1\cdots\alpha_k}\mathrm{d}x^{\alpha_1} \wedge \cdots \wedge \mathrm{d}x^{\alpha_k} \,, \tag{A.3.6}$$

$$\mathrm{d}\boldsymbol{\alpha} = \frac{\partial\alpha_{\alpha_1\cdots\alpha_k}}{\partial x^{\alpha_0}}\mathrm{d}x^{\alpha_0} \wedge \mathrm{d}x^{\alpha_1} \wedge \cdots \wedge \mathrm{d}x^{\alpha_k} \,. \tag{A.3.7}$$

A k-form $\boldsymbol{\alpha}$ is called *closed* if $\mathrm{d}\boldsymbol{\alpha} = 0$, and *exact* if there exists a $(k-1)$-form $\boldsymbol{\beta}$ such that $\boldsymbol{\alpha} = \mathrm{d}\boldsymbol{\beta}$. All exact forms are closed, as the definition of the exterior derivative implies $\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} = 0$, a result known as the *Poincaré lemma*. The converse is true, however, only *locally*: all closed forms are locally exact, but globally they may not be in general[1].

In the study of integration in the next section, we will work with forms of the same rank as the dimension of the manifold, n. These are sometimes referred to as *top forms*. (As we shall see, integration in any n-dimensional space is defined for n-forms.) Let $\boldsymbol{\alpha} \in \Lambda^n(\mathscr{M})$ be any n-form. Its antisymmetry in all n indices implies that, locally, its expansion (A.3.6) is simply

$$\boldsymbol{\alpha} = \alpha\,(x)\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n \,, \tag{A.3.8}$$

where $x = (x^1, \ldots, x^n)$ and $\alpha(x)$ is a function on $\mathscr{M}$.

Recall that zero-forms are also functions. Indeed, it is generally true, for any $0 \leq k \leq n$, that $\Lambda^k(\mathscr{M})$ is isomorphic to $\Lambda^{n-k}(\mathscr{M})$. Moreover, this isomorphism is provided by another famous and very useful operator, the *Hodge star* $\star : \Lambda^k(\mathscr{M}) \to \Lambda^{n-k}(\mathscr{M})$ defined uniquely, for any two k-forms $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, by the relation $\boldsymbol{\alpha} \wedge \star\boldsymbol{\beta} = \alpha_{\alpha_1\cdots\alpha_k}\beta^{\beta_1\cdots\beta_k}\sqrt{\det(\boldsymbol{g})}\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$. For more on this, see Chapter 7 of [Nakahara 2003].

Example: electromagnetism. Consider the electromagnetic four-vector potential A^a (traditionally written as (ϕ, A^i), with ϕ the scalar potential and A^i the vector potential). One can (using the spacetime metric to lower the index) think of this instead as a one-form $\boldsymbol{A} = A_a\mathrm{d}x^a$. Then, the Faraday tensor $F_{ab} = 2\partial_{[a}A_{b]}$ is a two-form (as $F_{ab} = F_{[ab]}$), given simply by the exterior derivative of $\boldsymbol{A}$, i.e. $\boldsymbol{F} = \mathrm{d}\boldsymbol{A}$. Automatically, the Poincaré lemma implies

$$\mathrm{d}\boldsymbol{F} = \mathrm{d}^2\boldsymbol{A} = 0 \,. \tag{A.3.9}$$

In traditional notation, with $\vec{E}$ denoting the electric field and $\vec{B}$ the magnetic field (from $\boldsymbol{A}$), this is equivalent to two Maxwell equations: the Gauss law for magnetism $\vec{\nabla} \cdot \vec{B} = 0$ and the Faraday induction law $\vec{\nabla} \times \vec{E} = -\partial_t\vec{B}$. These are in fact the constraints of the theory, manifested here as conditions of geometrical consistency. (Recall from Chapter 2 that the constraints of GR are also conditions of geometrical consistency.)

[1] The study of the implications of this leads to a very useful set of geometrical invariants called *de Rham cohomology groups*. See Chapter 11 of [Lee 2002].

The other two Maxwell equations are the dynamical evolution equations: the Gauss law for electrostatics $\vec{\nabla} \cdot \vec{E} = \rho$ and the Ampère-Maxwell law $\vec{\nabla} \times \vec{B} = \vec{j} + \partial_t \vec{E}$, where ρ and $\vec{j}$ are the charge and current densities. These are obtained from the Maxwell Lagrangian, given by $\mathcal{L} = -\frac{1}{4}\boldsymbol{F} : \boldsymbol{F} + \boldsymbol{A} \cdot \boldsymbol{J}$, where here we denote $\boldsymbol{J} = J_a \mathrm{d}x^a$ with $J_a = (\rho, j_i)$ the four-current written as a one-form. (See Chapter 4.) One obtains from the stationary action principle the dynamical Maxwell equations,

$$\star\, \mathrm{d} \star \boldsymbol{F} = \boldsymbol{J}\,. \tag{A.3.10}$$

A.4. Integration on manifolds

In this section we review integration over (regions of) $\mathscr{M}$.

First, suppose $\{(\mathscr{U}_i, \varphi_i)\}$ is an atlas for $\mathscr{M}$, where $\mathscr{U}_i \subset \mathscr{M}$ are the open subsets covering $\mathscr{M}$ (such that $\bigcup_i \mathscr{U}_i = \mathscr{M}$), and the homeomorphisms $\varphi_i : \mathscr{U}_i \to \mathbb{R}^n$ are the associated coordinate maps (or charts).

We begin by defining the integral of an n-form $\boldsymbol{\alpha}$ over any $\mathscr{U}_i$. In particular, this is defined to be the same as the integral of the transported (pushed-forward) form over the image of $\mathscr{U}_i$ in Euclidean space under φ_i, i.e.

$$\int_{\mathscr{U}_i} \boldsymbol{\alpha} = \int_{\varphi_i(\mathscr{U}_i)} (\varphi_i)_* \boldsymbol{\alpha} = \int_{\varphi_i(\mathscr{U}_i)} \alpha\,(x)\, \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n\,, \tag{A.4.1}$$

where to write the last equality we have used the expansion (A.3.8). This can now be made sense of as a usual integral in $\mathbb{R}^n$ by identifying the wedge product of basis forms $\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$ with the usual Riemann (or Lebesgue) measure $\mathrm{d}\mu_{\mathbb{R}^n} = \mathrm{d}x^1 \cdots \mathrm{d}x^n$ on $\mathbb{R}^n$, defined in the usual way. This finally gives us:

$$\int_{\mathscr{U}_i} \boldsymbol{\alpha} = \int_{\varphi_i(\mathscr{U}_i)} \mathrm{d}\mu_{\mathbb{R}^n} \alpha\,(x)\ . \tag{A.4.2}$$

The extension of this definition to integration over the entire manifold $\mathscr{M}$ is not quite so straightforward. In particular, it requires a result which is rather technical, and the proof of which we omit here (see, e.g., Chapter 2 of [Lee 2002]): namely the existence in any manifold of *partitions of unity*. These are a set of functions $\{\psi_i : \mathscr{M} \to [0,1] \subset \mathbb{R}\}$, said to be *subordinate* to $\{\mathscr{U}_i\}$, satisfying the following conditions: *(i)* they are supported entirely within each $\mathscr{U}_i$ (i.e. $\mathrm{supp}(\psi_i) \subset \mathscr{U}_i$); *(ii)* at any point $p \in \mathscr{M}$, they are nonzero for a finite number of i and add up to unity, i.e.

$$\sum_i \psi_i(p) = 1\,. \tag{A.4.3}$$

With this in hand, integration over the entire manifold $\mathscr{M}$ of a k-form $\boldsymbol{\alpha}$ can be defined. In particular, one "inserts the identity" (A.4.3) into the integrand of $\int_{\mathscr{M}} \boldsymbol{\alpha}$ to transform it into a sum of integrals over each $\mathscr{U}_i$ (given that each ψ_i is only supported therein), which are themselves in turn given by (A.4.2). That is, we define:

$$\int_{\mathscr{M}} \boldsymbol{\alpha} = \sum_i \int_{\mathscr{U}_i} \psi_i \boldsymbol{\alpha} \,. \tag{A.4.4}$$

An important and often used result is *Stokes' theorem*[2]. The modern geometrical version of this theorem was first formulated by [Cartan 1945]. It states that for any $(n-1)$-form $\boldsymbol{\alpha}$,

$$\int_{\mathscr{M}} \mathrm{d}\boldsymbol{\alpha} = \int_{\partial\mathscr{M}} \boldsymbol{\alpha} \,,$$

where $\partial\mathscr{M}$ is the boundary[3] of $\mathscr{M}$.

Now that we know how to make sense of integration of forms, we would also like to give meaning to the integration of a function $f : \mathscr{M} \to \mathbb{R}$ over (regions of) $\mathscr{M}$. In order to do this, one first requires the definition of a *volume form.* This is a nowhere-vanishing n-form, denoted as $\boldsymbol{\epsilon}_{\mathscr{M}} \in \Lambda^n(\mathscr{M})$, such that the total volume or *total measure* $\mu(\mathscr{M})$ of the manifold is given by its integral thereover,

$$\mu\left(\mathscr{M}\right) = \int_{\mathscr{M}} \boldsymbol{\epsilon}_{\mathscr{M}} \,. \tag{A.4.5}$$

The volume form is usually defined by the condition

$$\frac{1}{n!} \epsilon^{\mathscr{M}}_{a_1 \cdots a_n} \epsilon_{\mathscr{M}}^{a_1 \cdots a_n} = (-1)^s \;, \tag{A.4.6}$$

where s denotes the number of minus signs in the signature of $\boldsymbol{g}$ (so $s = 0$ if it is Riemannian and $s = 1$ if it is Lorentzian). It can be shown (see e.g. Appendix B of [Wald 1984]) that (A.4.6) implies that $\boldsymbol{\epsilon}_{\mathscr{M}}$ has a local expansion (A.3.6) given by:

$$\boldsymbol{\epsilon}_{\mathscr{M}} = \sqrt{|g|} \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n \,, \tag{A.4.7}$$

where $g = \det(\boldsymbol{g})$.

With this, we can now define the integral of a function $f : \mathscr{M} \to \mathbb{R}$ over $\mathscr{M}$ as

$$\int_{\mathscr{M}} \boldsymbol{\epsilon}_{\mathscr{M}} f \,. \tag{A.4.8}$$

One final useful result that we state here concerns the transport under a diffeomorphism $\phi : \mathscr{M} \to \tilde{\mathscr{M}}$ of the volume form of $\mathscr{M}$ to another manifold $\tilde{\mathscr{M}}$. Indeed, this does not simply yield the volume from of $\tilde{\mathscr{M}}$ itself; however, they are proportional, with the

[2] This has evolved through different versions throughout history. See [Katz 1979] for a detailed account. The first record of its appearance is in an 1850 letter by Lord Kelvin to Stokes, who then used it for several years as a problem in the Smith's Prize exam at Cambridge. (It is unknown if any of the students managed to prove it.) The first published proof is in [Hankel 1861].

[3] Formally, this can be defined if one of the coordinates in the charts φ_i of $\mathscr{M}$, say x^n, is always non-negative. Then one takes $\partial\mathscr{M} = \{p \in \mathscr{M} | \varphi_i(p) = (x^1, \ldots, x^{n-1}, 0)\}$.

proportionality factor given by a smooth function on $\tilde{\mathscr{M}}$ called the *Jacobian determinant*, $J \in \mathscr{F}(\tilde{\mathscr{M}})$. To be more precise, let $\boldsymbol{\epsilon}_{\tilde{\mathscr{M}}}$ denote the volume form of $\tilde{\mathscr{M}}$. Then we have (see Chapter 7 of [Abraham et al. 2001]):

$$\phi_* \boldsymbol{\epsilon}_{\mathscr{M}} = J \boldsymbol{\epsilon}_{\tilde{\mathscr{M}}} \,, \tag{A.4.9}$$

where $J = \det(\phi^*)$, with $\phi* : T\tilde{\mathscr{M}} \to T\mathscr{M}$ here indicating the push-forward.

Example: diffeomorphism-invariant action functionals. A useful result (see Proposition 10.20 of [Lee 2002]) is that for any diffeomorphism $\phi : \mathscr{M} \to \tilde{\mathscr{M}}$ and any compactly supported n-form $\boldsymbol{\alpha}$ on $\mathscr{M}$, we have

$$\int_{\mathscr{M}} \boldsymbol{\alpha} = \int_{\tilde{\mathscr{M}}} \phi_* \boldsymbol{\alpha} \,. \tag{A.4.10}$$

Suppose $\mathcal{S}[\psi] = \int_{\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{M}} f[\psi]$ is an action functional of a field theory, as defined in Chapter 2. The above result can then be regarded as a statement of diffeomorphism invariance: we have

$$\mathcal{S}[\psi] = \int_{\mathscr{V}} \boldsymbol{\epsilon}_{\mathscr{M}} f \,, \tag{A.4.11}$$

$$= \int_{\phi(\mathscr{V})} \phi_* \left(\boldsymbol{\epsilon}_{\mathscr{M}} f \right) \tag{A.4.12}$$

$$= \int_{\phi(\mathscr{V})} \left(\phi_* \boldsymbol{\epsilon}_{\mathscr{M}} \right) \left(f \circ \phi^{-1} \right) \tag{A.4.13}$$

$$= \int_{\phi(\mathscr{V})} \boldsymbol{\epsilon}_{\tilde{\mathscr{M}}} J \cdot f \circ \phi^{-1} \,. \tag{A.4.14}$$

In the second line we have used (A.4.10), in the third line the fact that the transport commutes with tensor products as well as (A.1.3), and finally in the last line (A.4.9). The latter may readily be recognized simply as a general version of the "change of coordinates" formula from standard multi-variable calculus in Euclidean space.

APPENDIX B

Particle-without-Particle: A Practical Pseudospectral Collocation Method for Linear Partial Differential Equations with Distributional Sources

Appendix summary. This appendix is based on the publication [Oltean, Sopuerta, et al. 2019].

Partial differential equations with distributional sources—involving (derivatives of) delta distributions—have become increasingly ubiquitous in numerous areas of physics and applied mathematics. It is often of considerable interest to obtain numerical solutions for such equations, but any singular ("particle"-like) source modeling invariably introduces nontrivial computational obstacles. A common method in the literature used to circumvent these is through some form of delta function approximation procedure on the computational grid; however, this often carries significant limitations on the efficiency of the numerical convergence rates, or sometimes even the resolvability of the problem at all.

In this appendix, we present an alternative technique for tackling such equations which avoids the singular behavior entirely: the Particle-without-Particle method. Previously introduced in the context of the self-force problem in gravitational physics, the idea is to discretize the computational domain into two (or more) disjoint pseudospectral (Chebyshev-Lobatto) grids such that the "particle" is always at the interface between them; thus, one only needs to solve homogeneous equations in each domain, with the source effectively replaced by jump (boundary) conditions thereon. We prove here that this method yields solutions to any linear PDE the source of which is any linear combination of delta distributions and derivatives thereof supported on a one-dimensional subspace of the problem domain. We then implement it to numerically solve a variety of relevant PDEs with applications in neuroscience, finance and acoustics. We generically obtain improved convergence rates relative to typical past implementations relying on delta function approximations.

Following an introduction in Section B.1 and some mathematical preliminaries in Section B.2, we prove in Section B.3 how the Particle-without-Particle method can be formulated and applied to problems with the most general possible "point" source, that is, one containing an arbitrary number of (linearly combined) one-dimensional delta functions and derivatives supported at an arbitrary number of points. Thus, one can use it on any type of (linear) PDE involving such sources.

Then in Sections B.4-B.7, we illustrate the application of this method, respectively, to first-order hyperbolic problems (with applications in neuroscience), parabolic problems (with applications in finance), second-order hyperbolic problems (with applications in acoustics), and finally elliptic problems.

Section B.8 offers some concluding remarks.

Partícula-sense-Partícula (appendix summary translation in Catalan). Aquest apèndix es basa en la publicació [Oltean, Sopuerta, et al. 2019].

Les equacions diferencials parcials amb fonts distributives - en particular, que impliquen (derivats de) distribucions delta - s'han tornat cada vegada més omnipresents en nombroses àrees de la física i les matemàtiques aplicades. Sovint és d'interès considerable obtenir solucions numèriques per a aquestes equacions, però qualsevol model de font singular (de tipus "partícula") introdueix invariablement obstacles computacionals no privats. Un mètode comú per evitar-les és mitjançant una forma d'aproximació de la funció delta a la graella computacional; no obstant això, sovint comporta limitacions importants en l'eficiència de les taxes de convergència numèrica, o, fins i tot, en la possibilitat de resoldre el problema en si mateix.

En aquest apèndix, presentem una tècnica alternativa per abordar aquestes equacions que evita completament el comportament singular: el mètode Partícula-sense-Partícula (Particle-without-Particle). Anteriorment introduïda en el context del problema de la força pròpia en la física gravitatòria, la idea és discretitzar el domini computacional en dues (o més) reixes disjuntes pseudospectrals (Chebyshev-Lobatto) de manera que la "partícula" sempre estigui a la interfície entre elles. Per tant, només cal resoldre equacions homogènies en cada domini, efectivament substituint la font per condicions de salt (de frontera). Aquí demostrem que aquest mètode produeix solucions a qualsevol equació diferencial parcial lineal la font de la qual és qualsevol combinació lineal de distribucions delta i derivats de les mateixes suportades en un subespai unidimensional del domini de la problema. A continuació, l'implementem per resoldre numèricament diverses equacions diferencials parcials rellevants amb aplicacions en neurociència, finances i acústica. Obtenim genèricament taxes de convergència millorades respecte a les implementacions anteriors típiques basades en aproximacions de la funció delta.

Després d'una introducció a la secció B.1 i d'alguns preliminaris matemàtics a la secció B.2, demostrem a la secció B.3 com es pot formular i aplicar el mètode Partícula-sense-Partícula a problemes amb la font de punt més general possible, és a dir, que conté un nombre arbitrari de funcions delta unidimensionals (linealment combinades) i derivades suportades en un nombre arbitrari de punts. Per tant, es pot utilitzar en qualsevol tipus d'equació diferencial parcial (lineal) que impliqui aquestes fonts.

A continuació, a les Seccions B.4-B.7, il·lustrem l'aplicació d'aquest mètode, respectivament, a problemes hiperbòlics de primer ordre (amb aplicacions en neurociència), problemes parabòlics (amb aplicacions en finances), problemes hiperbòlics de segon ordre (amb aplicacions en acústica) i, finalment, problemes el·líptics.

La secció B.8 ofereix algunes observacions finals.

Particule-sans-Particule (appendix summary translation in French). Cette annexe est basée sur la publication [Oltean, Sopuerta, et al. 2019].

Les équations aux dérivées partielles (EDP) avec sources distributionnelles - en particulier, impliquant (dérivées de) distributions delta - sont devenues de plus en plus omniprésentes dans de nombreux domaines de la physique et des mathématiques appliquées. Il est souvent d'un intérêt considérable d'obtenir des solutions numériques pour de telles équations, mais toute modélisation de source singulière (semblable à une « particule ») introduit invariablement des obstacles de calcul non triviaux. Une méthode possible pour les contourner consiste à utiliser une procédure d'approximation de la fonction delta sur la grille de calcul ; cependant, cela limite souvent considérablement l'efficacité des taux de convergence numérique, voire parfois même la posibilité de resoudre le problème.

Dans cette annexe, nous présentons une technique alternative pour traiter de telles équations, qui évite totalement le comportement singulier : la méthode Particule-sans-Particule (Particle-without-Particle, PwP). Auparavant introduite dans le contexte du problème de la force propre dans la physique gravitationnelle, l'idée est de discrétiser le domaine de calcul en deux (ou plus) grilles disjointes pseudospectraux (Chebyshev-Lobatto) de telle sorte que la « particule » soit toujours à l'interface entre eux ; il suffit donc de résoudre des équations homogènes dans chaque domaine, la source étant effectivement remplacée par des conditions de saut (aux limites). Nous montrons ici que cette méthode fournit des solutions à toute EDP linéaire dont la source est quelque combinaison linéaire de distributions delta et de leurs dérivées supportées sur un sous-espace unidimensionnel du domaine du problème. Nous l'implémentons ensuite pour résoudre numériquement divers types des EDP pertinentes dans les domaines des neurosciences, de la finance et de l'acoustique. Nous obtenons de manière générique des taux de convergence meilleurs par rapport aux implémentations passées typiques reposant sur des approximations de fonctions delta.

Après une introduction dans la section B.1 et quelques préliminaires mathématiques dans la section B.2, nous montrons à la section B.3 comment la méthode Particule-sans-Particule peut être formulée et appliquée aux problèmes avec la source « ponctuelle » la plus générale possible, c'est-à-dire contenant un nombre arbitraire de fonctions delta unidimensionnelles (combinées linéairement) et des dérivées avec support à un nombre arbitraire de points. Ainsi, on peut l'utiliser sur n'importe quel type d'EDP (linéaire) impliquant de telles sources.

Ensuite, dans les sections B.4 à B.7, nous illustrons l'application de cette méthode, respectivement, aux problèmes hyperboliques du premier ordre (avec applications dans la neuroscience), aux problèmes paraboliques (avec des applications dans la finance), aux problèmes hyperboliques du second ordre (avec applications dans l'acoustique) et enfin des problèmes elliptiques.

La section B.8 propose quelques remarques de conclusion.

B.1. Introduction

Mathematical models often have to resort—be it out of expediency or mere ignorance—to deliberately idealized descriptions of their contents. A common idealization across different fields of applied mathematics is the use of the Dirac delta distribution, often simply referred to as the *delta "function"*, for the purpose of describing highly localized phenomena: that is to say, phenomena the length scale of which is significantly smaller, in some suitable sense, than that of the problem into which they figure, and the (possibly complicated) internal structure of which can thus be safely (or safely enough) ignored in favour of a simple "point-like" cartoon. Canonical examples of this from physics are notions such as "point masses" in gravitation or "point charges" in electromagnetism.

Yet, despite their potentially powerful conceptual simplifications, introducing distributions into any mathematical model is something that must be handled with great technical care. In particular, let us suppose that our problem of interest has the very general form

$$\mathcal{L}u = S \quad \text{in } \mathscr{U} \subseteq \mathbb{R}^n \,, \tag{B.1.1}$$

where $\mathcal{L}$ is an n-dimensional (partial, if $n > 1$) differential operator (of arbitrary order m), u is a quantity to be solved for (a function, a tensor etc.) and we assume that S—the "source"—is distributional in nature, *i.e.* we have $S : \mathcal{D}(\mathscr{U}) \to \mathbb{R}$, where we use the common notation $\mathcal{D}(\mathscr{U})$ to refer to the set of smooth compactly-supported functions, *i.e.* "test functions", on $\mathscr{U}$. It follows, therefore, that u—if it exists—must also be distributional in nature. So strictly speaking, from the point of view of the classic theory of distributions [Schwartz 1957], the problem (B.1.1) is only well-defined—and hence may admit distributional solutions u—provided that $\mathcal{L}$ is linear[1].

The problem with a nonlinear $\mathcal{L}$ is essentially that, classically, products of distributions do not make sense [Schwartz 1954]. While there has certainly been work by mathematicians aiming to generalize the theory of distributions so as to accommodate this possibility [Bottazzi 2017; Colombeau 2013; Li 2007], in the standard setting we are only really allowed to talk of *linear* problems of the form (B.1.1). Opportunely, very many of the typical

[1] Here the terms "linear"/"nonlinear" have their standard meaning from the theory of partial differential equations.

problems in physics and applied mathematics involving distributions take precisely this form.

The inspiration for considering (B.1.1) in general in this paper actually comes from a setting where one does, in fact, encounter non-linearities a priori: namely, gravitational physics. (For a general discussion regarding the treatment of distributions therein, see Ref. [Geroch and Traschen 1987].) In particular, equations such as (B.1.1) arise when attempting to describe the backreaction of a body with a "small" mass upon the spacetime through which it moves—known as its *self-force* [Blanchet et al. 2011; Detweiler and Whiting 2003; Gralla and Wald 2011, 2008; Mino et al. 1997; Poisson et al. 2011; Pound 2015b; Quinn and Wald 1997; A. D. A. M. Spallicci, Ritter, and Aoudia 2014; Wardell 2015]. (A similar version of this problem exists in electromagnetism, where a "small" charge backreacts upon the electromagnetic field that determines its motion [Barut 1980; DeWitt and Brehme 1960; Dirac 1938; Poisson et al. 2011].) In the full Einstein equations of general relativity, which can be regarded as having the schematic form (B.1.1) with u describing the gravitational field (that is, the spacetime geometry, in the form of the metric) and S denoting the matter source (the stress-energy-momentum tensor), $\mathcal{L}$ is a nonlinear operator. Nevertheless, for a distributional S (representing the "small" mass as a "point particle" source) one *can* legitimately seek solutions to a *linearized* version of (B.1.1) in the context of perturbation theory, *i.e.* at first order in an expansion of $\mathcal{L}$ in the mass. The detailed problem, in this case, turns out to be highly complex, and in practice, u must be computed numerically. The motivation for this, we may add, is not just out of purely theoretical or foundational concern—the calculation of the self-force is also of significant applicational value for gravitational wave astronomy. To wit, it will in fact be indispensable for generating accurate enough waveform templates for future space-based gravitational wave detectors such as LISA [Amaro-Seoane et al. 2017, 2013] *vis-à-vis* extreme-mass-ratio binary systems, which are expected to be among the most fruitful sources thereof. For these reasons, having at our disposal a practical and efficient numerical method for handling equations of the form (B.1.1) is of consequential interest.

What is more, these sorts of partial differential equations (PDEs) arise frequently in other fields as well; indeed, (B.1.1) can adequately characterize quite a wide variety of (linear) mathematical phenomena assumed to be driven by "localized sources". A few examples, which we will consider one by one in different sections of this paper, are the following:

(i) *First-order hyperbolic PDEs*: in neuroscience, advection-type PDEs with a delta function source can be used in the modeling of neural populations [Cáceres, Carrillo, et al. 2011; Cáceres and Schneider 2016; Casti et al. 2002; Haskell et al. 2001];

(ii) *Parabolic PDEs*: in finance, heat-type PDEs with delta function sources are sometimes used to model price formation [Achdou et al. 2014; Burger et al. 2013; Caffarelli et al. 2011; Lasry and Lions 2007; Markowich et al. 2009; Pietschmann 2012];

(iii) *Second-order hyperbolic PDEs*: in acoustics, wave-type PDEs with delta function (or delta derivative) sources are used to model monopoles (or, respectively, multipoles) [Kaltenbacher 2017; Petersson and Sjogreen 2010]; more complicated equations of this form also appear, for example, in seismology models [Aki and Richards 2009; Madariaga 2007; Petersson and Sjogreen 2010; Romanowicz and Dziewonski 2007; Shearer 2009], which we will briefly comment upon.

(iv) *Elliptic PDEs*: Finally, we will look at a simple Poisson equation with a singular source [Tornberg and Engquist 2004]; such equations can describe, for example, the potential produced by a very localized charge in electrostatics.

B.1.1. Scope of this work. The purpose of this work is to explicate and generalize a practical method for numerically solving equations like (B.1.1), as well as to illustrate its broad applicability to the various problems listed in (i)-(iv) above. Previously implemented with success only in the specific context of the self-force problem [Canizares 2011; Canizares and Sopuerta 2009b, 2014, 2011a,b; Canizares, Sopuerta, and Jaramillo 2010b; Jaramillo, Sopuerta, et al. 2011; Oltean, Sopuerta, et al. 2017], we dub it the "Particle-without-Particle" (PwP) method. (Other methods for the computation of the self-force have also been developed based on matching the properties of the solutions on the sides of the delta distributions—see, *e.g.*, the indirect (source-free) integration method of Refs. [Aoudia and A. D. A. M. Spallicci 2011; Ritter, Aoudia, et al. 2015a,b; Ritter, A. D. A. M. Spallicci, et al. 2011; A. D. A. M. Spallicci and Ritter 2014; A. D. Spallicci et al. 2012].) The basic idea of the PwP approach is the following: One begins by writing u as a sum of distributions each of which has support outside (plus, if necessary, at the location of) the points where S is supported; one then solves the equations for each of these pieces of u and finally matches them in such a way that their sum satisfies the original problem (B.1.1). In fact, as we shall soon elaborate upon, this approach will not work in general for all possible problems of the form (B.1.1). However, we will prove that it will *always* work if, rather than the source being a distribution defined on all of $\mathscr{U}$, we have instead $S : \mathcal{D}(\mathscr{I}) \to \mathbb{R}$ with $\mathscr{I} \subseteq \mathbb{R}$ representing a *one-dimensional subspace* of $\mathscr{U}$.

To make things more concrete, let us briefly describe this procedure using the simplest possible example: let $f : \mathscr{U} \to \mathbb{R}$ be an arbitrary given function and suppose $S = f\delta$ where $\delta : \mathcal{D}(\mathscr{I}) \to \mathbb{R}$ is the delta function supported at some point $x_p \in \mathscr{I}$. Then, to solve (B.1.1), one would assume the decomposition (or "ansatz") $u = u^- \Theta^- + u^+ \Theta^+$ with $\Theta^\pm : \mathcal{D}(\mathscr{I}) \to \mathbb{R}$ denoting appropriately defined Heaviside distributions (supported to the right/left of x_p, respectively), and $u^{\pm} : \mathscr{U} \to \mathbb{R}$ being simple functions (not distributions) to be solved for. Inserting such a decomposition for u into (B.1.1), one obtains *homogeneous* equations $\mathcal{L}u^{\pm} = 0$ on the appropriate domains, supplemented by the necessary boundary conditions (BCs) for these equations at $x_p \in \mathscr{I}$, explicitly determined by f. Generically, the latter arise in the form of relations between the limits of u^- and u^+ (and/or the derivatives thereof) at x_p, and for this reason are called "jump conditions" (JCs). Effectively, the latter completely replace the "point" source S in the original problem, now simply reduced to solving sourceless equations—hence the nomenclature of the method.

While in principle one can certainly contemplate the adaptation of these ideas into a variety of established approaches for the numerical solution of PDEs, we will focus specifically on their implementation through pseudospectral collocation (PSC) methods on Chebyshev-Lobatto (CL) grids. The principal advantages thereof lie in their typically very efficient (exponential) rates of numerical convergence as well as the ease of incorporating and modifying BCs (JCs) throughout the evolution. Indeed, PSC methods have enjoyed very good success in past work [Canizares 2011; Canizares and Sopuerta 2009b, 2014, 2011a,b; Canizares, Sopuerta, and Jaramillo 2010b; Jaramillo, Sopuerta, et al. 2011; Oltean, Sopuerta, et al. 2017] on the PwP approach for self-force calculations (and in gravitational physics more generally [Grandclément and Novak 2009], including arbitrary precision implementations [Santos-Oliván and Sopuerta 2018]), and so we shall not deviate very much from this recipe in the models considered in this paper. Essentially the main difference will be that here, instead of the method of lines which featured in most of the past PwP self-force work, we will for the most part carry out the time evolution using the simplest first-order forward finite difference scheme; we do this, on the one hand, so that we may illustrate the principle of the method explicitly in a very elementary way without too many technical complications, and on the other, to show how well it can work even with such basic tactics. Depending on the level of accuracy and computational efficiency required for any realistic application, these procedures can naturally be complexified (to higher order, more domains, more complicated domain compactifications etc.) for properly dealing with the sophistication of the problem at hand.

To summarize, past work using the PwP method only solved a specific form of Eq. (B.1.1) pertinent to the self-force problem: that is, with a particular choice of $\mathcal{L}$ and S (upon which we will comment more later). It did *not* consider the question of the extent to which the idea of the method could be useful in general for solving distributionally-sourced PDEs. These appear, as enumerated above, in many other fields of study—and we submit that a method such as this would be of valuable benefit to researchers working therein. The novelty of the present paper will thus be to formulate a *completely general* PwP method for *any* distributionally-sourced (linear) problem of the form (B.1.1) with the single limiting condition that $\mathrm{supp}(S) \subset \mathscr{I} \subseteq \mathscr{U}$ where $\dim(\mathscr{I}) = 1$. We will prove rigorously why and how the method works for such problems, and then we will implement it to obtain numerical solutions to the variety of different applications mentioned earlier in order to illustrate its broad practicability. We will see that, in general, this method either matches or improves upon the results of other methods existent in the literature for tackling distributionally-sourced PDEs—and we turn to a more detailed discussion of this topic in the next subsection.

B.1.2. Comparison with other methods in the literature. Across all areas of application, the most commonly encountered—and, perhaps, most naively suggestible—strategy for numerically solving equations of the form (B.1.1) is to rely upon some sort of delta function approximation procedure on the computational grid [Jung 2009; Jung and Don 2009; Petersson, O'Reilly, et al. 2016; Tornberg and Engquist 2004]. For instance, the simplest imaginable choice in this vein is just a narrow hat function (centered at the point where the delta function is supported, and having total measure 1) which, for better accuracy, one can upgrade to higher-order polynomials, or even trigonometric functions. Another readily evocable possibility is to use a narrow Gaussian—and indeed, this is one option that has in fact been tried in self-force computations as well (see Ref. [López-Alemán et al. 2003], for example). However, this unavoidably introduces into the problem an additional, artificial length scale: that is, the width of the Gaussian, which a priori need not have anything to do with the actual ("physical") length scale of the source. Moreover, there is the evident drawback that no matter how small this artificial length scale is chosen, the solutions will never be well-resolved close to the distributional source location: there will always be some sort of Gibbs-type phenomenon[2] there.

Methods for solving (B.1.1) which are closer in spirit to our PwP method have been explored in Refs. [Field et al. 2009 and [Shin and Jung 2011]. In particular, both of these works have used the idea of placing the distributional source at the interface of computational grids—however, they tackle the numerical implementation differently than we do.

In the case of Ref. [Field et al. 2009]—which, incidentally, is also concerned with the self-force problem—the difference is that the authors use a discontinuous Galerkin method (rather than spectral methods, as in our PwP approach), and the effect of the distributional source is accounted for via a modification of the numerical flux at the "particle" location. This relies essentially upon a weak formulation of the problem, wherein a choice has to be made about how to assign measures to the distributional terms over the relevant computational domains. In contrast, we directly solve only for smooth solutions supported away

[2] The Gibbs phenomenon, originally discovered by Henry Wilbraham [Wilbraham 1848] and rediscovered by J. Willard Gibbs [Gibbs 1899], refers generally to an overshoot in the approximation of a piecewise continuously differentiable function near a jump discontinuity.

from the "particle" location, and account for the distributional source simply by imposing adequate boundary—*i.e.* jump— conditions there.

Ref. [Shin and Jung 2011] is closer to our approach in this sense, as the authors there also use spectral methods and also account for the distributional source via jump conditions. However, the difference with our method is that Ref. [Shin and Jung 2011] treats these jump conditions as additional constraints (rather than built-in boundary conditions) for the smooth solutions away from the distributional source, thus over-determining the problem. That being the case, the authors are led to the need to define a functional (expressing how well the differential equations plus the jump conditions are satisfied) to be minimized—constituting what they refer to as a "least squares spectral collocation method". There is however no unique way to choose this functional. Moreover, the complication of introducing it is not at all necessary: our approach, in contrast, simply replaces the discretization of (the homogeneous version of) the differential equations at the "particle" location with the corresponding jump conditions (*i.e.* it imposes the jump conditions *as* boundary conditions, *by construction*—something which PSC methods are precisely designed to be able to handle), leading to completely determined systems in all cases which are solved directly, without further complications.

Finally, neither Ref. [Field et al. 2009] nor [Shin and Jung 2011] analyzed to any significant extent the conditions under which their methods might be applicable to more general distributionally-sourced PDEs. As mentioned, in the present paper we will devote a careful proof entirely to this issue.

This paper is structured as follows. Following some mathematical preliminaries in Section B.2, we prove in Section B.3 how the PwP method can be formulated and applied to problems with the most general possible "point" source $S : \mathcal{D}(\mathscr{I}) \to \mathbb{R}$, that is, one containing an arbitrary number of (linearly combined) delta derivatives and supported at an arbitrary number of points in $\mathscr{I}$. Thus, one can use it on any type of (linear) PDE involving such sources, which we illustrate with the applications listed in (i)-(iv) above in Sections B.4-B.7 respectively. Finally, we give concluding remarks in Section B.8.

B.2. Setup

We wish to begin by establishing some basic notation and then reviewing some pertinent properties of distributions that we will need to make use of later on. While we will certainly strive to maintain a fair level of mathematical rigour here and throughout this paper (at least, insofar as a certain amount of formal precaution is inevitably necessary when dealing with distributions), our principal aim remains that of presenting practical methodologies; hence the word "distribution" may at times be liberally interchanged for "function" (*e.g.* we may say "delta function" instead of "delta distribution") and some notation possibly slightly abused, when the context is clear enough to not pose dangers for confusion.

B.2.1. Distributionally-sourced linear PDEs. Consider the problem (B.1.1) with $S : \mathcal{D}(\mathscr{I}) \to \mathbb{R}$, where $\mathscr{I} \subseteq \mathbb{R}$ is a one-dimensional subspace of $\mathscr{U} \subseteq \mathbb{R}^n$, as discussed in the introduction. Then we can view $\mathscr{U}$ as a product space, $\mathscr{U} = \mathscr{I} \times \mathscr{V}$ with $\mathscr{V} = \mathscr{U}/\mathscr{I} \subseteq \mathbb{R}^{n-1}$, and write coordinates on $\mathscr{U}$ as $\mathbf{x} = (x, \mathbf{y})$ with $x \in \mathscr{I}$ and $\mathbf{y} = (y_1, \ldots, y_{n-1}) \in \mathscr{V}$, such that

$$\begin{aligned} f : \mathscr{U} = \mathscr{I} \times \mathscr{V} \subseteq \mathbb{R} \times \mathbb{R}^{n-1} = \mathbb{R}^n \to \mathbb{R} \\ \mathbf{x} = (x, \mathbf{y}) = (x, y_1, \ldots, y_{n-1}) \mapsto f(\mathbf{x}) \end{aligned} \tag{B.2.1}$$

denotes any arbitrary function on $\mathscr{U}$.

It is certainly possible, in the setup we are about to describe, to have $\mathscr{V} = \emptyset$, *i.e.* problems involving just *ODEs* (on $\mathscr{U} = \mathscr{I}$) of the form (B.1.1)—and, in fact, our first elementary example illustrating the PwP method in the following section will be of such a kind. For the more involved numerical examples we will study in later sections, we will most often be dealing with functions of two variables, $x \in \mathscr{I}$ for "space" (or some other pertinent parameter) and $t \in \mathscr{V} \subseteq \mathbb{R}$ for time.

For any function (B.2.1) involved in these problems, we will sometimes use the notation $f' = \partial_x f$ for the "spatial" derivative; also, we may employ $\dot{f} = \partial_t f$ for the partial derivative with respect to time t when $\{t\}$ is (a subspace of) $\mathscr{V}$.

Now, as in the introduction, let $\mathcal{L}$ be any general m-th order *linear* differential operator. The sorts of PDEs (B.1.1) that we will be concerned with have the basic form

$$\mathcal{L}u = S = f\delta_{(p)} + g\delta'_{(p)} + \cdots, \tag{B.2.2}$$

where $f(\mathbf{x})$, $g(\mathbf{x})$ etc. are "source" functions prescribed by the problem at hand, and we employ the convenient notation

$$\delta_{(p)}(x) = \delta(x - x_p(\mathbf{y})) \tag{B.2.3}$$

to indicate the Dirac delta distribution on $\mathscr{I}$ centered at the "particle location" $x_p(\mathbf{y})$—the functional form of which can be either specified a priori, or determined via some given prescription as the solution u itself is evolved. When there is no risk of confusion, we may sometimes omit the $\mathbf{y}$ dependence in our notation and simply write x_p.

In fact, our PwP method can even deal with multiple, say M, "particles". PwP computations of the self-force have actually only required $M = 1$ (there being only one "particle" involved in the problem), so the general $M \geq 1$ case has not been considered up to now. Concordantly, to express our problem of interest (B.2.2) in the most general possible form, let us employ the typical PDE notation for "multi-indices" [Evans 1998],

$\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ with each $\alpha_I \in \mathbb{Z}^{\geq}$ being a non-negative integer (indexed from $I = 0$ to $I = n-1$ so as to make sense *vis-à-vis* our coordinate notation on $\mathscr{U}$, instead of the more usual practice to label them from 1 to n), and $|\alpha| = \sum_{I=0}^{n-1} \alpha_I$. Furthermore, we define $\alpha! = \prod_{I=0}^{n-1} \alpha_I!$. Thus, the most general m-th order linear partial differential operator can be written as $\mathcal{L} = \sum_{|\alpha| \leq m} \xi^\alpha(\mathbf{x}) D^\alpha$ where $\xi^\alpha : \mathscr{U} \to \mathbb{R}$ are arbitrary functions and $D^\alpha = \partial^{|\alpha|} / \partial x^{\alpha_0} \partial y_1^{\alpha_1} \cdots \partial y_{n-1}^{\alpha_{n-1}}$. Hence, we are dealing with any problem which can be placed into the form

$$\sum_{|\alpha| \leq m} \xi^\alpha(\mathbf{x}) D^\alpha u(\mathbf{x}) = \sum_{i=1}^{M} \sum_{j=0}^{K} f^{ij}(\mathbf{x}) \delta^{(j)}(x - x_{p_i}(\mathbf{y})) \,, \tag{B.2.4}$$

with $f^{ij} : \mathscr{U} \to \mathbb{R}$ denoting the "source" functions (for the j-th delta derivative of the i-th particle) and $K \in \mathbb{Z}^{\geq}$ the highest order of the delta function derivatives in S, appropriately supplemented by initial/boundary conditions (ICs/BCs).

Let us give a few basic examples to render this setup more palpable. One very simple example—that which will serve as our first illustration of the PwP method in the next section—is the simple harmonic oscillator with a constant delta function forcing (source) term—that is, the ODE (with $\mathscr{V} = \emptyset$):

$$u'' + u = a\delta_{(p)} \,, \tag{B.2.5}$$

where $\delta_{(p)}(x) = \delta(x - x_p)$ for some fixed $x_p \in \mathscr{I}$, and $a \in \mathbb{R}$. Another example is the wave equation with a moving singular source,

$$\left(\partial_t^2 - \partial_x^2\right) u(x,t) = f(x,t)\, \delta(x - x_p(t)) \,, \tag{B.2.6}$$

with $x_p(t)$ specified as a function of time.

B.2.2. Properties of distributions. We now wish to remind the reader of a few basic properties of distributions before proceeding to describe the PwP procedure; for a good detailed exposition, see *e.g.* Ref. [Stakgold and M. J. Holst 2011].

Let $f : \mathscr{U} \to \mathbb{R}$ be, as before, any function involved in the problem (B.2.4). We denote by

$$\begin{aligned} f_p : \mathscr{V} &\to \mathbb{R} \\ \mathbf{y} &\mapsto f_p(\mathbf{y}) = f(x_p(\mathbf{y}), \mathbf{y}) \end{aligned} \tag{B.2.7}$$

the function evaluated at the "particle" position.

Furthermore, let $\phi \in \mathcal{D}(\mathscr{I})$ be any test function on $\mathscr{I}$. Then we define the action of the distribution associated with f as:

$$\langle f, \phi \rangle = \int_{\mathscr{I}} \mathrm{d}x\, f(x, \mathbf{y})\, \phi(x) \,. \tag{B.2.8}$$

We say that two functions f and g are *equivalent* in the sense of distributions if

$$\langle f,\phi\rangle = \langle g,\phi\rangle \Leftrightarrow f \equiv g\,. \tag{B.2.9}$$

An identity which will be important for us in discussing the PwP method is the following [Cortizo 1995; Li 2007]:

$$f\left(x,\mathbf{y}\right)\delta_{(p)}^{(n)}\left(x\right) \equiv \left(-1\right)^{n}\sum_{j=0}^{n}\left(-1\right)^{j}\binom{n}{j}f_{p}^{(n-j)}\left(\mathbf{y}\right)\delta_{(p)}^{(j)}\left(x\right)\,, \tag{B.2.10}$$

where $\delta_{(p)}^{(n)} = \partial_x^n\delta_{(p)}$. For concreteness, let us write down the first three cases explicitly here:

$$f\left(x,\mathbf{y}\right)\delta_{(p)}\left(x\right) \equiv f_{p}\left(\mathbf{y}\right)\delta_{(p)}\left(x\right)\,, \tag{B.2.11}$$

$$f\left(x,\mathbf{y}\right)\delta'_{(p)}\left(x\right) \equiv -f'_{p}\left(\mathbf{y}\right)\delta_{(p)}\left(x\right) + f_{p}\left(\mathbf{y}\right)\delta'_{(p)}\left(x\right)\,, \tag{B.2.12}$$

$$f\left(x,\mathbf{y}\right)\delta''_{(p)}\left(x\right) \equiv f''_{p}\left(\mathbf{y}\right)\delta_{(p)}\left(x\right) - 2f'_{p}\left(\mathbf{y}\right)\delta'_{(p)}\left(x\right) + f_{p}\left(\mathbf{y}\right)\delta''_{(p)}\left(x\right)\,. \tag{B.2.13}$$

For the interested reader, we offer in Appendix A of [Oltean, Sopuerta, et al. 2019] a proof by induction of the formula (B.2.10), which is instructive for appreciating the subtleties generally involved in manipulating distributions.

Let

$$\Theta_{(p)}^{\pm}\left(x\right) = \Theta\left(\pm\left(x - x_{p}\left(\mathbf{y}\right)\right)\right) \tag{B.2.14}$$

be the Heaviside function which is supported to the right/left (respectively) of x_p. Then, we have:

$$\partial_x\Theta_{(p)}^{\pm} = \pm\,\delta_{(p)}\,, \tag{B.2.15}$$

$$\partial_{y_j}\Theta_{(p)}^{\pm} = \mp\left(\partial_{y_j}x_p\right)\delta_{(p)}\,, \tag{B.2.16}$$

and so on for higher order partials.

For notational expediency, we may sometimes omit the (p) subscript on the Heaviside functions (and derivatives thereof) when the context is sufficiently clear.

B.3. The Particle-without-Particle method

As discussed heuristically in the introduction, the basic idea of our method for solving (B.2.4) is to effectively eliminate the "point"-like source or "particle" from the problem by decomposing the solution u into a series of distributions: specifically, Heaviside functions $\Theta^i : \mathcal{D}(\mathscr{I}) \to \mathbb{R}$ supported in each of the $M+1$ disjoint regions of $\mathscr{I}\backslash\mathrm{supp}(S)$ (*i.e.* $\mathrm{supp}(\Theta^i)\cap\mathrm{supp}(S) = \emptyset, \forall i$ and $\mathrm{supp}(\Theta^i)\cap\mathrm{supp}(\Theta^j) = \emptyset, \forall i \neq j$) and, if necessary, delta functions (plus delta derivatives) at $\mathrm{supp}(S)$:

$$u = \sum_{i=0}^{M}u^{i}\Theta^{i} + \sum_{i=1}^{M}\sum_{j=0}^{K-m}h^{ij}\delta_{(p_i)}^{(j)}\,, \tag{B.3.1}$$

where $u^i : \mathscr{U} \to \mathbb{R}$ and we need to include the second sum with $h^{ij} : \mathscr{V} \to \mathbb{R}$ only if $K \geq m$.

We will prove in this section that one can always obtain solutions of the form (B.3.1) to the problem (B.2.4). In particular, inserting (B.3.1) into (B.2.4) will always yield homogeneous equations

$$\mathcal{L}u^i = 0 \quad \text{in } \left(\mathscr{I} \backslash \text{supp}(S)\right) \times \mathscr{V} \,, \tag{B.3.2}$$

along with JCs on (the derivatives of) u—and possibly (derivatives of) h^{ij} if applicable. In general, we define the "jump" $[\cdot]_p : \mathscr{V} \to \mathbb{R}$ in the value of any function $f : \mathscr{U} \to \mathbb{R}$ at $x_p(\mathbf{y})$ as

$$[f]_p\left(\mathbf{y}\right) = \lim_{x \to x_p(\mathbf{y})^+} f\left(x, \mathbf{y}\right) - \lim_{x \to x_p(\mathbf{y})^-} f\left(x, \mathbf{y}\right) \,. \tag{B.3.3}$$

Henceforth, for convenience, we will generally omit the $\mathbf{y}$-dependence and simply write $[f]_p$.

First we will work through a simple example in order to offer a more concrete sense of the method, and afterwards we will show in general how (B.3.1) solves (B.2.4).

B.3.1. Simple example. We illustrate here the application of our PwP method to a very simple ODE (and single-particle) example. We will consider the problem

$$\mathcal{L}u = u'' + u = a\delta + b\delta^{'} \,, \;\; x \in \mathscr{I} = [-L, L] \,, \;\; u\left(\pm L\right) = 0 \,, \tag{B.3.4}$$

where δ is simply the delta function centered at $x_p = 0$.

We begin by decomposing u as

$$u = u^- \Theta^- + u^+ \Theta^+ \,, \tag{B.3.5}$$

where $\Theta^\pm(x) = \Theta(\pm x)$, and we insert this into (B.3.4). Using (B.2.15), the LHS becomes simply

$$\begin{aligned} \mathcal{L}u\left(x\right) &= u''\left(x\right) + u\left(x\right) && \text{(B.3.6)} \\ &= \left\{\mathcal{L}u^-\left(x\right)\right\}\Theta^-\left(x\right) + \left\{\mathcal{L}u^+\left(x\right)\right\}\Theta^+\left(x\right) \\ &\quad + \left\{-2\left(u^-\left(x\right)\right)' + 2\left(u^+\left(x\right)\right)'\right\}\delta\left(x\right) \\ &\quad + \left\{-u^-\left(x\right) + u^+\left(x\right)\right\}\delta'\left(x\right) \,. && \text{(B.3.7)} \end{aligned}$$

Now before we can equate this to the distributional terms in the source (RHS), we must apply the identity (B.2.10). In particular, we use $f(x)\delta(x) \equiv f_p\delta(x)$ and $f(x)\delta'(x) \equiv$

$-f_p'\delta(x) + f_p\delta'(x)$. Thus, the above becomes

$$\begin{aligned}\mathcal{L}u\,(x) \equiv &\left\{\mathcal{L}u^-\,(x)\right\}\Theta^-\,(x) + \left\{\mathcal{L}u^+\,(x)\right\}\Theta^+\,(x)\\ &+\left\{-2\left(u^-\right)_p' + 2\left(u^+\right)_p'\right\}\delta\,(x)\\ &+\left\{\left(u^-\right)_p' - \left(u^+\right)_p'\right\}\delta\,(x) + \left\{-u_p^- + u_p^+\right\}\delta'\,(x) \qquad \text{(B.3.8)}\\ =&\left\{\mathcal{L}u^-\,(x)\right\}\Theta^-\,(x) + \left\{\mathcal{L}u^+\,(x)\right\}\Theta^+\,(x) + \left[u'\right]_p\delta\,(x) + [u]_p\,\delta'\,(x)\;. \qquad \text{(B.3.9)}\end{aligned}$$

Plugging this into the DE (B.3.4), we have

$$\left\{\mathcal{L}u^-\right\}\Theta^- + \left\{\mathcal{L}u^+\right\}\Theta^+ + \left[u'\right]_p\delta + [u]_p\,\delta' \equiv a\delta + b\delta^{'}\;. \qquad \text{(B.3.10)}$$

Therefore the original problem is equivalent to the system of equations:

$$\begin{cases}\mathcal{L}u^- = 0, & x\in\mathscr{D}^- = [-L,0]\;,\;\; u^-\,(-L) = 0\,,\\ \mathcal{L}u^+ = 0, & x\in\mathscr{D}^+ = [0,L]\;,\;\; u^+\,(L) = 0\,,\\ [u]_p = b, & [u']_p = a\,.\end{cases} \qquad \text{(B.3.11)}$$

Let us solve (B.3.11), for simplicity, taking $L = \pi/4$. The left homogeneous equation in (B.3.11) has the general solution $u^- = A^-\cos(x) + B^-\sin(x)$, and the BC tells us that $0 = u^-(-\pi/4) = \frac{1}{\sqrt{2}}(A^- - B^-)$, *i.e.*

$$A^- - B^- = 0\,. \qquad \text{(B.3.12)}$$

The right homogeneous equation in (B.3.11) similarly has general solution $u^+ = A^+\cos(x) + B^+\sin(x)$, with the BC stating $0 = u^+(\pi/4) = \frac{1}{\sqrt{2}}(A^+ + B^+)$, *i.e.*

$$A^+ + B^+ = 0\,. \qquad \text{(B.3.13)}$$

So far we have two equations (B.3.12)-(B.3.13) for four unknowns (the integration constants in the general solutions). It is the JCs in (B.3.11) that provide us with the remaining necessary equations to fix the solution. We have $u^-(0) = A^-$, $(u^-)'(0) = B^-$, $u^+(0) = A^+$ and $(u^+)'(0) = B^+$ (understood in the appropriate limit approaching $x_p = 0$). Hence the JCs tell us:

$$b = [u]_p = u^+(0) - u^-(0) = A^+ - A^-\,, \qquad \text{(B.3.14)}$$

$$a = \left[u'\right]_p = (u^+)'(0) - (u^-)'(0) = B^+ - B^-\,. \qquad \text{(B.3.15)}$$

(We can think of the JCs as a mixing of the degrees of freedom in the homogeneous solutions in such a way that they "link together" to produce the solution generated by the original distributional source.) Solving (B.3.12)-(B.3.15), we get $A^- = -\frac{a+b}{2} = B^-$, $A^+ = -\frac{a-b}{2} = -B^+$. We now have the full solution to our original problem (B.3.4):

$$u\,(x) = -\frac{a+b}{2}\left(\cos\,(x) + \sin\,(x)\right)\Theta\,(-x) - \frac{a-b}{2}\left(\cos\,(x) - \sin\,(x)\right)\Theta\,(x)\;. \qquad \text{(B.3.16)}$$

B.3.2. General proof. Suppose we have M “particles” located at $\mathrm{supp}(S) = \{x_{p_i}\}_{i=1}^{M} \subset \mathscr{I}$, as in the problem (B.2.4), with $x_{p_1} < x_{p_2} < \cdots < x_{p_M}$. (NB: For $M \geq 2$, if there exists any subset of $\mathscr{V}$ where it should happen that $x_{p_i}(\mathbf{y}) > x_{p_{i+1}}(\mathbf{y})$ as a consequence of the $\mathbf{y}$-evolution, we can, without loss of generality, simply swap indices within that subset so as to always have $x_{p_i} < x_{p_{i+1}}, \forall i$.) Furthermore let us assume for the moment that the maximum order of delta function derivatives in the source is one less than the order of the PDE (or smaller), *i.e.* $K = m - 1$. In this case, we do not need to consider the second term on the RHS of (B.3.1), *i.e.* u is just split up into pieces which are supported only in between all the particle locations: $u^0(\mathbf{x})$ to the left of x_{p_1}, $u^1(\mathbf{x})$ between x_{p_1} and x_{p_2}, ..., $u^i(\mathbf{x})$ between x_{p_i} and $x_{p_{i+1}}$, ..., and finally u^M to the right of x_{p_M}. Thus, we take

$$u = \sum_{i=0}^{M} u^i \Theta^i \,, \tag{B.3.17}$$

where we define

$$\Theta^i = \begin{cases} \Theta^-_{(p_1)} \,, & i = 0 \,, \\ \Theta^+_{(p_i)} - \Theta^+_{(p_{i+1})} \,, & 1 \leq i \leq M-1 \,, \\ \Theta^+_{(p_M)} \,, & i = M \,, \end{cases} \tag{B.3.18}$$

denoting, as before, $\Theta^{\pm}_{(p_i)}(x) = \Theta(\pm(x - x_{p_i}(\mathbf{y})))$. Another way of stating this is that we assume for u a piecewise decomposition

$$u = \begin{cases} u^0 \,, & x \in \mathscr{D}^0 \,, \\ & \vdots \\ u^M \,, & x \in \mathscr{D}^M \,, \end{cases} \tag{B.3.19}$$

where the $\mathscr{D}^i$'s are disjoint subsets of $\mathscr{I}$ between each “particle location”, *i.e.*

$$\mathscr{I} = \mathrm{supp}\,(S) \cup \left(\bigcup_{i=0}^{M} \mathscr{D}^i \right) \,, \tag{B.3.20}$$

where

$$\mathscr{D}^i = \begin{cases} \{x \in \mathscr{I} | x < x_{p_1}\} \,, & i = 1 \,, \\ \left\{x \in \mathscr{I} | x_{p_i} < x < x_{p_{i+1}}\right\} \,, & 1 \leq i \leq M-1 \,, \\ \{x \in \mathscr{I} | x_{p_M} < x\} \,, & i = M \,. \end{cases} \tag{B.3.21}$$

The general strategy, then, is to insert (B.3.17) into (B.2.4), and to obtain a set of equations by matching (regular function) terms multiplying the same derivative order of the Heaviside distributions. Explicitly, using the Leibniz rule, we get

$$\mathcal{L}u = \sum_{i=0}^{M} \sum_{|\alpha| \leq m} \sum_{|\beta| \leq |\alpha|} \binom{\alpha}{\beta} \xi^{\alpha} \left(D^{\alpha-\beta} u^i\right) \left(D^{\beta} \Theta^i\right) = \sum_{i=1}^{M} \sum_{j=0}^{m-1} f^{ij} \delta^{(j)}_{(p_i)} \,. \tag{B.3.22}$$

At zeroth order in derivatives of the Heaviside functions, *i.e.* the sum of all $|\beta| = 0$ terms in the LHS above, we will always simply obtain—in the absence of any Heaviside functions on the RHS—a set of $M+1$ homogeneous equations, which constitute simply the original equation on each disjoint subset of $\mathscr{I}$ but *with no source*:

$$\sum_{i=0}^{M}\left(\sum_{|\alpha|\leq m}\xi^{\alpha}D^{\alpha}u^{i}\right)\Theta^{i}=0\Leftrightarrow\mathcal{L}u^{i}=0\ \text{ in }\ \mathscr{D}^{i}\times\mathscr{V},\,\forall i\,. \tag{B.3.23}$$

At first order and higher in the Heaviside derivatives (thus, zeroth order and higher in delta function derivatives), *i.e.* the sum of all $|\beta| \neq 0$ terms in the LHS of (B.3.22), we have terms of the form

$$D^{\beta}\Theta^{i}=\partial_{x}^{\beta_0}\partial_{y_1}^{\beta_1}\cdots\partial_{y_{n-1}}^{\beta_{n-1}}\begin{cases}\Theta^{-}_{(p_1)}\,, & i=0\,,\\ \Theta^{+}_{(p_i)}-\Theta^{+}_{(p_{i+1})}\,, & 1\leq i\leq M-1\,,\\ \Theta^{+}_{(p_M)}\,, & i=M\,,\end{cases} \tag{B.3.24}$$

$$=\sum_{j=0}^{|\beta|-1}\begin{cases}F^{0j}\delta^{(j)}_{(p_1)}\,, & i=0\,,\\ F^{ij}\delta^{(j)}_{(p_i)}+G^{ij}\delta^{(j)}_{(p_{i+1})}\,, & 1\leq i\leq M-1\,,\\ F^{Mj}\delta^{(j)}_{(p_M)}\,, & i=M\,,\end{cases} \tag{B.3.25}$$

for some $\mathbf{y}$-dependent functions $F^{ij}:\mathscr{V}\to\mathbb{R}$ and $G^{ij}:\mathscr{V}\to\mathbb{R}$ which arise from the implicit differentiation (*e.g.*, Eqns. (B.2.15)-(B.2.16)), and the precise form of which does not concern us for the present purposes. Plugging (B.3.25) into (B.3.22) and manipulating the sums, we get

$$\sum_{i=1}^{M}\sum_{|\alpha|\leq m}\sum_{0<|\beta|\leq|\alpha|}\sum_{j=0}^{|\beta|-1}\Phi^{\alpha,\beta,ij}\delta^{(j)}_{(p_i)}=\sum_{i=1}^{M}\sum_{j=0}^{m-1}f^{ij}\delta^{(j)}_{(p_i)}\,. \tag{B.3.26}$$

where for convenience we have defined

$$\Phi^{\alpha,\beta,ij}\left(\mathbf{x}\right)=\binom{\alpha}{\beta}\xi^{\alpha}\left(\mathbf{x}\right)\Big(F^{ij}\left(\mathbf{y}\right)D^{\alpha-\beta}u^{i}\left(\mathbf{x}\right)+H^{ij}\left(\mathbf{y}\right)D^{\alpha-\beta}u^{i-1}\left(\mathbf{x}\right)\Big)\,, \tag{B.3.27}$$

for some $\mathbf{y}$-dependent functions $H^{ij}:\mathscr{V}\to\mathbb{R}$ (related to F^{ij} and G^{ij}, and the precise form of which is also unimportant). At this point, one must be careful: *before* drawing conclusions regarding the equality of terms (the coefficients of the delta function derivatives) in (B.3.26), one should apply the identity (B.2.10). Doing this, one obtains:

$$\sum_{\alpha,\beta,i,j}\sum_{k=0}^{j}(-1)^{j+k}\binom{j}{k}\left(\partial_x^{j-k}\Phi^{\alpha,\beta,ij}\right)_{p_i}\delta^{(k)}_{(p_i)}=\sum_{i,j}\sum_{k=0}^{j}(-1)^{j+k}\binom{j}{k}\left(\partial_x^{j-k}f^{ij}\right)_{p_i}\delta^{(k)}_{(p_i)}\,, \tag{B.3.28}$$

with the omitted summation limits as before. Thus, we see that on the LHS, we have terms involving

$$\partial_x^{j-k}\Phi^{\alpha,\beta,ij} = \partial_x^{j-k}\left\{\binom{\alpha}{\beta}\xi^\alpha\left(F^{ij}D^{\alpha-\beta}u^i + H^{ij}D^{\alpha-\beta}u^{i-1}\right)\right\} \tag{B.3.29}$$

$$= \binom{\alpha}{\beta}\left\{F^{ij}\partial_x^{j-k}\left(\xi^\alpha D^{\alpha-\beta}u^i\right) + H^{ij}\partial_x^{j-k}\left(\xi^\alpha D^{\alpha-\beta}u^{i-1}\right)\right\} \tag{B.3.30}$$

$$= \binom{\alpha}{\beta}\sum_{l=0}^{j-k}\binom{j-k}{l}\left(\partial_x^{j-k-l}\xi^\alpha\right)\left[F^{ij}\left(\partial_x^l D^{\alpha-\beta}u^i\right) + H^{ij}\left(\partial_x^l D^{\alpha-\beta}u^{i-1}\right)\right]. \tag{B.3.31}$$

Thus, defining the $\mathbf{y}$-dependent functions

$$\Psi^{\alpha,\beta,ijkl} = (-1)^{j+k}\binom{j}{k}\binom{j-k}{l}\binom{\alpha}{\beta}\left(\partial_x^{j-k-l}\xi^\alpha\right)_{p_i}, \tag{B.3.32}$$

$$\psi^{ijk} = (-1)^{j+k}\binom{j}{k}\left(\partial_x^{j-k}f^{ij}\right)_{p_i}, \tag{B.3.33}$$

we can use (B.3.31) to write (B.3.28) in the form:

$$\sum_{\alpha,\beta,i,j,k,l}\Psi^{\alpha,\beta,ijkl}\left[F^{ij}\left(\partial_x^l D^{\alpha-\beta}u^i\right)_{p_i} + H^{ij}\left(\partial_x^l D^{\alpha-\beta}u^{i-1}\right)_{p_i}\right]\delta_{(p_i)}^{(k)} = \sum_{i,j,k}\psi^{ijk}\delta_{(p_i)}^{(k)}, \tag{B.3.34}$$

where the terms involving u^i partials "at the particle" should be understood as the limit evaluated from the appropriate direction, *i.e.*

$$\left(D^\gamma u^i(\mathbf{x})\right)_{p_i} = \lim_{x\to x_{p_i}^+} D^\gamma u^i(x,\mathbf{y}), \tag{B.3.35}$$

$$\left(D^\gamma u^{i-1}(\mathbf{x})\right)_{p_i} = \lim_{x\to x_{p_i}^-} D^\gamma u^i(x,\mathbf{y}). \tag{B.3.36}$$

Having obtained (B.3.34), we can finally match the coefficients of each $\delta_{(p_i)}^{(k)}$ to obtain the JCs with which the homogeneous equations (B.3.23) must be supplemented.

Let us now extend this method to problems where the maximum order of delta function derivatives in the source equals or exceeds the order of the PDE, *i.e.* $K \geq m$, a case not previously required—and hence not yet considered—in any of the past PwP work on the self-force. To do this, we just add to our ansatz the second term on the RHS of (B.3.1), which for convenience we denote u^δ; that is:

$$u = \sum_{i=0}^{M}u^i\Theta^i + u^\delta, \quad u^\delta = \sum_{i=1}^{M}\sum_{j=0}^{K-m}h^{ij}\delta_{(p_i)}^{(j)}, \tag{B.3.37}$$

with $h^{ij}(\mathbf{y})$ to be solved for. Inserting (B.3.37) into (B.2.4) we get, on the LHS of the PDE, the homogeneous problems (at zeroth order) as before, then the LHS of (B.3.34) due again to the sum of Heaviside functions term in (B.3.37), plus the following due to the sum of delta function derivatives:

$$\mathcal{L}u^\delta = \sum_{|\alpha|\leq m} \xi^\alpha D^\alpha \sum_{i=1}^{M} \sum_{j=0}^{K-m} h^{ij} \delta^{(j)}_{(p_i)} \tag{B.3.38}$$

$$= \sum_{\alpha,i,j} \sum_{|\beta|\leq|\alpha|} \binom{\alpha}{\beta} \xi^\alpha \left(D^{\alpha-\beta} h^{ij}\right) \left(D^\beta \delta^{(j)}_{(p_i)}\right) , \tag{B.3.39}$$

using the Leibniz rule. Next, we employ the Faà di Bruno formula [Constantine and Savits 1996] to carry out the implicit differentiation of the delta function derivatives; writing $(n-1)$ dimensional multi-indices on $\mathscr{V}$ (pertaining only to the $\mathbf{y}$ variables) with tildes, *e.g.* $\tilde{\beta} = (\beta_1, \ldots, \beta_{n-1})$, we have the following:

$$D^\beta \delta^{(j)}_{(p_i)} = D^{\tilde{\beta}} \delta^{(j+\beta_0)}_{(p_i)} = \tilde{\beta}! \sum_{l=1}^{|\tilde{\beta}|} \delta^{(j+\beta_0+l)}_{(p_i)} \sum_{s=1}^{|\tilde{\beta}|} \sum_{\mathscr{P}_s(\tilde{\beta},l)} \prod_{k=1}^{s} \frac{\left(-D^{\tilde{\lambda}_k} x_{p_i}\right)^{q_k}}{q_k! \left(\tilde{\lambda}_k!\right)^{q_k}} , \tag{B.3.40}$$

where $\mathscr{P}_s(\tilde{\beta}, l) = \{(q_1, \ldots, q_s; \tilde{\lambda}_1, \ldots, \tilde{\lambda}_s) : q_k > 0, 0 \prec \tilde{\lambda}_1 \prec \cdots \prec \tilde{\lambda}_s, \sum_{k=1}^s q_k = l$ and $\sum_{k=1}^s q_k \tilde{\lambda}_k = \tilde{\beta}\}$. Therefore, with all the summation limits the same as above, we get

$$\mathcal{L}u^\delta = \sum_{\alpha,\beta,i,j,l,s} \sum_{\mathscr{P}_s(\tilde{\beta},l)} \binom{\alpha}{\beta} \tilde{\beta}! \xi^\alpha \left(D^{\alpha-\beta} h^{ij}\right) \delta^{(j+\beta_0+l)}_{(p_i)} \prod_{k=1}^{s} \frac{\left(-D^{\tilde{\lambda}_k} x_{p_i}\right)^{q_k}}{q_k! \left(\tilde{\lambda}_k!\right)^{q_k}} . \tag{B.3.41}$$

Finally, we use the distributional identity (B.2.10) to obtain

$$\mathcal{L}u^\delta \equiv \sum_{\alpha,\beta,i,j,l,s} \sum_{\mathscr{P}_s(\tilde{\beta},l)} \binom{\alpha}{\beta} \tilde{\beta}! \left(\left(D^{\alpha-\beta} h^{ij}\right) \prod_{k=1}^{s} \frac{\left(-D^{\tilde{\lambda}_k} x_{p_i}\right)^{q_k}}{q_k! \left(\tilde{\lambda}_k!\right)^{q_k}} \right)_{p_i}$$

$$\times (-1)^{j+\beta_0+l} \sum_{r=0}^{j+\beta_0+l} (-1)^r \binom{j+\beta_0+l}{r} \left(\partial_x^{j+\beta_0+l-r} \xi^\alpha\right)_{p_i} \delta^{(r)}_{(p_i)} , \tag{B.3.42}$$

with which the higher order delta function derivatives on the RHS of (B.2.4) can be matched.

B.3.3. Limitations of the method. Let us now discuss more amply the potential issues one is liable to encounter in any attempt to extend the PwP method further beyond the setup we have described so far.

Firstly, we stress once more that the method is applicable only to *linear* PDEs. As pointed out in the introduction, this is simply an inherent limitation of the classic theory of distributions. In particular, there it has long been proved [Schwartz 1954] (see also the discussion in Ref. [Bottazzi 2017]) that there does not exist a differential algebra

$(A, +, \otimes, \delta)$ wherein the real distributions can be embedded, and: (i) $\otimes$ extends the product over $C^0(\mathbb{R})$; (ii) $\delta : A \to A$ extends the distributional derivative; (iii) $\forall u, v \in A$, the product rule $\delta(u \otimes v) = (\delta u) \otimes v + u \otimes (\delta v)$ holds. Attempts have been made to overcome this and create a sensible nonlinear theory of distributions by defining and working with more general objects dubbed "generalized functions" [Colombeau 2013]. Nonetheless, these have their own drawbacks (*e.g.* they sacrifice coherence between the product over $C^0(\mathbb{R})$ and that of the differential algebra), and different formulations are actively being investigated by mathematicians [Benci 2013; Bottazzi 2017]. A PwP method for nonlinear problems in the context of these formulations could be an interesting line of inquiry for future work.

Secondly, as we have seen, the PwP method as developed here is guaranteed to work only for those (linear) PDEs the source S of which is a distribution not on the entire problem domain $\mathscr{U}$, but only on a one-dimensional subspace $\mathscr{I}$ of that domain. One may sensibly wonder whether this situation can be improved, *i.e.* whether a similar procedure could succeed in tackling equations with sources involving (derivatives of) delta functions in *multiple* variables—yet, one may also immediately realize that such an attempted extension quickly leads to significant complications and potentially impassable problems. Let us suppose that the source contains (derivatives of) delta functions in $\bar{n} > 1$ variables. We still define $\mathscr{I}$ such that $\mathrm{supp}(S) \subset \mathscr{I}$, so now we have $\mathscr{I} \subseteq \mathbb{R}^{\bar{n}}$, and let us adapt the rest of our notation accordingly so that an arbitrary function on $\mathscr{U}$ is

$$\begin{aligned} f : \mathscr{U} \ = \ \mathscr{I} \times \mathscr{V} \ \subseteq \ \mathbb{R}^{\bar{n}} \times \mathbb{R}^{n-\bar{n}} \ = \ \mathbb{R}^n \to \mathbb{R} \\ \mathbf{x} = (\bar{\mathbf{x}}, \mathbf{y}) = (\bar{x}_1, ..., \bar{x}_{\bar{n}}, y_1, ..., y_{n-\bar{n}}) \mapsto f(\mathbf{x}) \ . \end{aligned} \tag{B.3.43}$$

We also adapt the multi-index notation to $\alpha = (\bar{\alpha}_1, \bar{\alpha}_2, \ldots, \bar{\alpha}_{\bar{n}}, \alpha_1, \alpha_2, \ldots, \alpha_{n-\bar{n}})$. We can still write the most general linear partial differential operator, just as we did earlier, as $\mathcal{L} = \sum_{|\alpha| \leq m} \xi^\alpha(\mathbf{x}) D^\alpha$ where now $D^\alpha = \partial^{|\alpha|} / \partial \bar{x}_1^{\bar{\alpha}_1} \cdots \partial \bar{x}_{\bar{n}}^{\bar{\alpha}_{\bar{n}}} \partial y_1^{\alpha_1} \cdots \partial y_{n-\bar{n}}^{\alpha_{n-\bar{n}}}$. Moreover, in general, we use the barred boldface notation $\bar{\mathbf{v}}$ for any vector in $\mathscr{I}$, $\bar{\mathbf{v}} = (\bar{v}_1, \ldots, \bar{v}_{\bar{n}}) \in \mathscr{I} \subseteq \mathbb{R}^{\bar{n}}$.

One may first ask whether a PwP-type method could be used to handle "point" sources in $\mathscr{I} \subseteq \mathbb{R}^{\bar{n}}$. In other words, can we find a decomposition of u which could be useful for a problem of the form

$$\mathcal{L} u(\mathbf{x}) = f(\mathbf{x}) \, \delta(\bar{\mathbf{x}} - \bar{\mathbf{x}}_p(\mathbf{y})) + \bar{\mathbf{g}}(\mathbf{x}) \cdot \bar{\boldsymbol{\nabla}} \delta(\bar{\mathbf{x}} - \bar{\mathbf{x}}_p(\mathbf{y})) + \cdots \, , \tag{B.3.44}$$

(assuming for simplicity a *single* point source at $\bar{\mathbf{x}}_p \in \mathscr{I}$) with $\bar{\boldsymbol{\nabla}} = \partial / \partial \bar{\mathbf{x}}$ and given functions $f : \mathscr{U} \to \mathbb{R}$, $\bar{\mathbf{g}} : \mathscr{U}^{\bar{n}} \to \mathbb{R}$ etc.? Intuitively, in order to match the delta function (derivatives) on the RHS, we might expect u to contain the $\bar{n}$-dimensional Heaviside function $\Theta : \mathcal{D}(\mathscr{I}) \to \mathbb{R}$. Thus, in the same vein as (B.3.17), a possible attempt (for $K < m$)

might be to try a splitting such as

$$u(\mathbf{x}) = \sum_{\bar{\boldsymbol{\sigma}}=\Pi^{\bar{n}}(\pm)} u^{\bar{\boldsymbol{\sigma}}}(\mathbf{x})\,\Theta\left(\bar{\boldsymbol{\sigma}}\odot\left(\bar{\mathbf{x}}-\bar{\mathbf{x}}_p(\mathbf{y})\right)\right)\,, \tag{B.3.45}$$

where Π is here the Cartesian product and $\odot$ the entrywise product; *but* whether or not this will work depends completely upon the detailed form of $\mathcal{L}$. For example, the procedure *might* work in the case where $\mathcal{L}$ contains a nonvanishing $D^{(1,1,\dots,1,\alpha_1,\dots,\alpha_{n-\bar{n}})}$ term, so as to produce a $\delta(\bar{\mathbf{x}}-\bar{\mathbf{x}}_p)$ term upon its action on u (in the form (B.3.45)), needed to match the $f(\mathbf{x})$ term on the RHS of (B.3.44). However, this still does not guarantee that *all* the distributional terms can in the end be appropriately matched, and so in general, one should *not* expect that such an approach in these sorts of problems will yield a workable strategy.

To render the above discussion a little less abstract, let us illustrate what we mean by way of a very simple example. Consider a two-dimensional Poisson equation on $\mathscr{U} = \{(x,y)\} \subseteq \mathbb{R}^2$: $(\partial_x^2+\partial_y^2)u = \delta_2(x,y)$, where the RHS is the two-dimensional delta function supported at the origin. An attempt to solve this via our method would begin by decomposing the solution into a form $u=\sum_j u^j\Theta^j$, for some suitably-defined Heaviside functions Θ^j— supported, for example, on positive/negative half-planes in each of the two coordinates, or perhaps on each quadrant of $\mathbb{R}^2$. However, the RHS of this problem is, by definition, $\delta_2(x,y)=\delta(x)\delta(y)=(\partial_x\Theta^+(x))(\partial_y\Theta^+(y))$, and there is no way to get such a term from the operator $\mathcal{L}=\partial_x^2+\partial_y^2$ acting on any linear combination of Heaviside functions. The unconvinced reader is invited to try a few attempts for themselves, and the difficulties with this will quickly become apparent.

That said, *one* case in which a PwP-type procedure could work is when the source contains (one-dimensional) "string"-like singularities (instead of $\bar{n}$-dimensional "point"-like ones) in each of the $\bar{\mathbf{x}}$ variables—in other words, when our problem is of the form

$$\mathcal{L}u(\mathbf{x}) = \sum_{a=1}^{\bar{n}} f^a(\mathbf{x})\,\delta\left(\bar{x}_a-\bar{x}_{a,p}(\mathbf{y})\right) + \sum_{a=1}^{\bar{n}} g^a(\mathbf{x})\,\delta'\left(\bar{x}_a-\bar{x}_{a,p}(\mathbf{y})\right) + \cdots\,, \tag{B.3.46}$$

with $f^a:\mathscr{U}\to\mathbb{R}$, $g^a:\mathscr{U}\to\mathbb{R}$ etc. Then, a decomposition of u which can be tried in such situations (for $K<m$) is

$$u(\mathbf{x}) = \sum_{a=1}^{\bar{n}}\sum_{\sigma_a=\pm} u^{a,\sigma_a}(\mathbf{x})\,\Theta\left(\sigma_a\left(\bar{x}_a-\bar{x}_{a,p}(\mathbf{y})\right)\right)\,. \tag{B.3.47}$$

B.4. First order hyperbolic PDEs

We now move on to applications of the PwP method, beginning with first order hyperbolic equations. First we look at the standard advection equation, and then a simple neural population model from neuroscience. Finally, we consider another popular advection-type problem with a distributional source—namely, the shallow water equations with discontinuous bottom topography—and briefly explain why the PwP method cannot be used in that case.

B.4.1. Advection equation. As a first very elementary illustration of our method, let us consider the $(1+1)$-dimensional advection equation for $u(x,t)$ with a time function singular point source at some $x = x_*$:

$$\begin{cases} \partial_t u + \partial_x u = g\left(t\right)\delta\left(x - x_*\right) \,, & x \in \mathscr{I} = [0, L]\,, \quad t > 0\,, \\ u\left(x, 0\right) = 0\,, & u\left(0, t\right) = u\left(L, t\right)\,, \end{cases} \tag{B.4.1}$$

where we assume that the source time function $g(t)$ is smooth and vanishes at $t = 0$. On an unbounded spatial domain (*i.e.* $x \in \mathbb{R}$), the exact solution of this problem is

$$u_{\text{ex}}\left(x,t\right) = \left[\Theta\left(x - x_*\right) - \Theta\left(x - x_* - t\right)\right] g\left(t - \left(x - x_*\right)\right)\,, \tag{B.4.2}$$

i.e. the forward-translated source function in the right half of the future light cone emanating from x_*. If we suppose that the source location satisfies $x_* \in (0, L/2]$, then (B.4.2) is also a solution of our problem (B.4.1) for $t \in [0, L - x_*]$.

This precise problem is treated in Ref. [Petersson, O'Reilly, et al. 2016] using a (polynomial) delta function approximation procedure, with the following: $g(t) = \mathrm{e}^{-(t-t_0)^2/2}$, $t_0 = 8$, $L = 40$ and $x_* = 10 + \pi$. We numerically implement the exact same setup, but using our PwP method: that is, we decompose $u = u^- \Theta^- + u^+ \Theta^+$ where $\Theta^{\pm} = \Theta(\pm(x - x_*))$. Inserting this into (B.4.1), we get homogeneous PDEs $\partial_t u^{\pm} + \partial_x u^{\pm} = 0$ to the left and right of the singularity, *i.e.* on $x \in \mathscr{D}^- = [0, x_*]$ and $x \in \mathscr{D}^+ = [x_*, L]$ respectively, along with a jump in the solution $[u]_* = g(t)$ at the point of the source singularity.

The details of all our numerical schemes in this work are described in an appendix, Section B.9. In particular, for the present problem, see Subsection B.9.2. We also offer in Subsection B.9.1 a brief description of the PSC methods and notation used therein.

The solution for zero initial data is displayed in Figure B.1, and the numerical convergence in Figure B.2. For the latter, we plot—for the numerical solution $\boldsymbol{u}$ at $t = T/2$—both the absolute error (in the l^2 norm on the CL grids, as in Ref. [Petersson, O'Reilly, et al. 2016]), $\epsilon_{\text{abs}} = ||\boldsymbol{u} - \boldsymbol{u}_{\text{ex}}||_2$, as well as the truncation error in the right CL domain $\mathscr{D}^+$ given simply the absolute value of the last spectral coefficient a_N of $\boldsymbol{u}^+$. We see that the truncation error exhibits typical (exponential) spectral convergence; the absolute error converges at the same rate until $N \approx 40$, after which it converges more slowly because it becomes dominated by the $\mathcal{O}(\Delta t) = \mathcal{O}(N^{-2})$ error in the finite difference time evolution scheme. Nevertheless, for the same number of grid points, our procedure still yields a lower order of magnitude of the l^2 error as was obtained in Ref. [Petersson, O'Reilly, et al. 2016] with a *sixth* order finite difference scheme (relying on a a source discretization with

6 moment conditions and 6 smoothness conditions); we present a simple comparison of these in the following table:

ϵ_{abs}	$N = 80$	$N = 160$
[Petersson, O'Reilly, et al. 2016]	$\mathcal{O}(10^{-2})$	$\mathcal{O}(10^{-3})$
PwP method	$\mathcal{O}(10^{-3})$	$\mathcal{O}(10^{-4})$

B.4.2. Advection-type equations in neuroscience. Advection-type equations with distributional sources arise in practice, for example, in the modeling of neural populations. In particular, among the simplest of these are the so-called "integrate-and-fire" models. For some of the earlier work on such models from a neuroscience perspective, see for example [Casti et al. 2002; Haskell et al. 2001] and references therein; for more recent work focusing on mathematical aspects, see [Cáceres, Carrillo, et al. 2011; Cáceres and Schneider 2016]. Their aim is to describe the probability density $\rho\,(\boldsymbol{v}, t)$ of neurons as a function of certain state variables $\boldsymbol{v}$ and time t. Often the detailed construction of these models can be quite involved and dependent on a large number of parameters, so to simply illustrate the principle of our method we here consider the simple case where the single state variable is the voltage V. Then, generally speaking, the dynamics of $\rho(V, t)$ takes the form of a Fokker-Planck-type equation on $V \in (-\infty, L]$ with a singular source at some fixed $V = V_* < L$,

$$\partial_t \rho + \partial_V \left(f\left(V, N\left(t\right)\right)\rho\right) - \frac{\sigma^2}{2}\partial_V^2 \rho = N\left(t\right)\delta\left(V - V_*\right)\,. \tag{B.4.3}$$

The source time function $N(t)$ must be such that conservation of probability, *i.e.* $\partial_t \int \mathrm{d}V\,\rho = 0$, is guaranteed under homogeneous Dirichlet BCs.

As a simplification of this problem, let us suppose, as is sometimes done, that the diffusive part (the second derivative term on the LHS) of (B.4.3) is negligible. Moreover, in simple cases, the velocity function f in the advection term has the form $f = -V + \text{constant}$, and we just work with the constant set equal to 1. We restrict ourselves to a bounded domain for V which for illustrative purposes we just choose to be $\mathscr{I} = [0, L]$. Demanding homogeneous Dirichlet BCs at the left boundary in conjunction with conservation of probability fixes the source time function to be $N(t) = (1-L)\rho(L, t)$. Thus, we are going to tackle the following problem:

$$\begin{cases} \partial_t \rho + \partial_V \left(\left(1 - V\right)\rho\right) = \left(1 - L\right)\rho\left(L, t\right)\delta\left(V - V_*\right)\,, & V \in \mathscr{I} = \left[0, L\right], t > 0\,, \\ \rho\left(V, 0\right) = \rho_0\left(V\right)\,, \;\; \int_{\mathscr{I}} \mathrm{d}V \rho_0\left(V\right) = 1\,, & \rho\left(0, t\right) = 0\,. \end{cases} \tag{B.4.4}$$

We now implement the PwP decomposition: $\rho = \rho^- \Theta^- + \rho^+ \Theta^+$ with $\Theta^{\pm} = \Theta(\pm(V - V_*))$. Inserting this into the PDE (B.4.4), we get the homogeneous problems

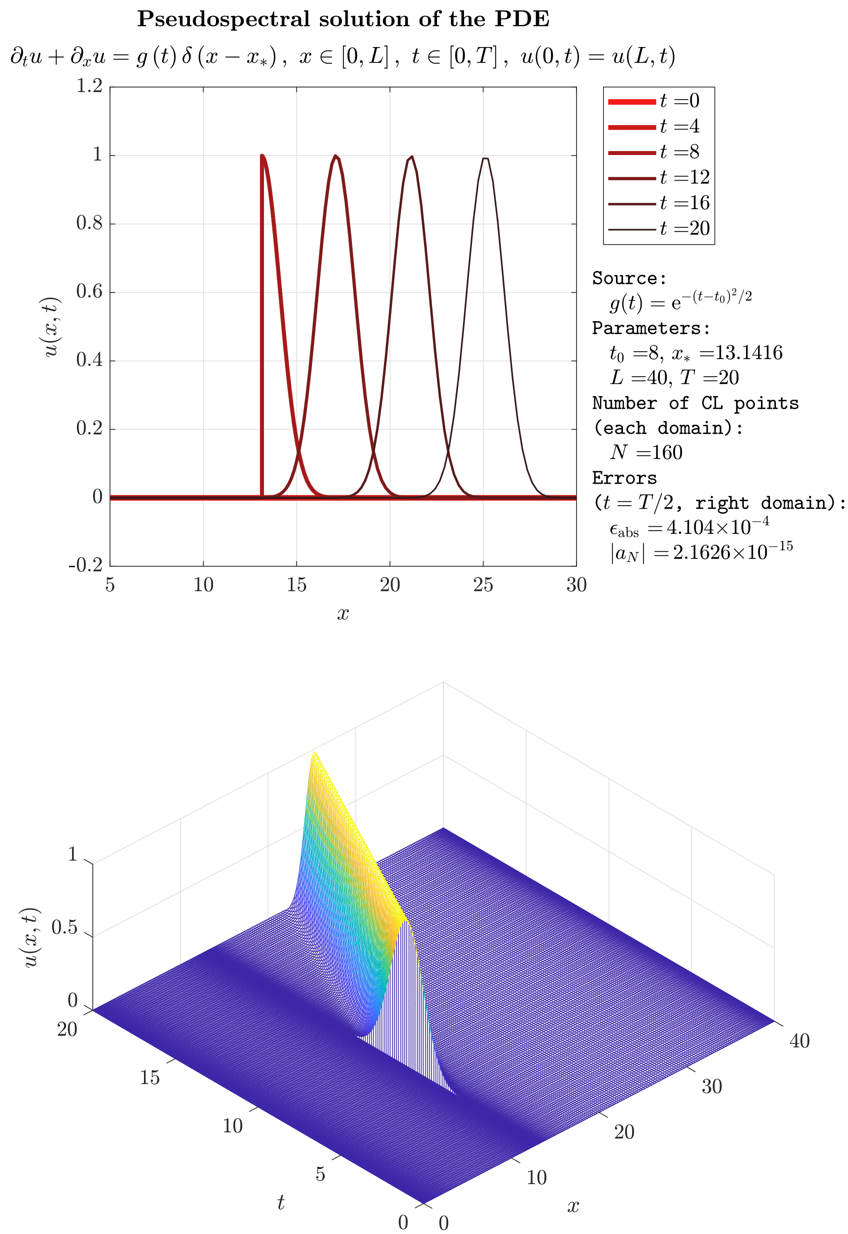

Figure B.1. Solution of the problem (B.4.1) with zero initial data.

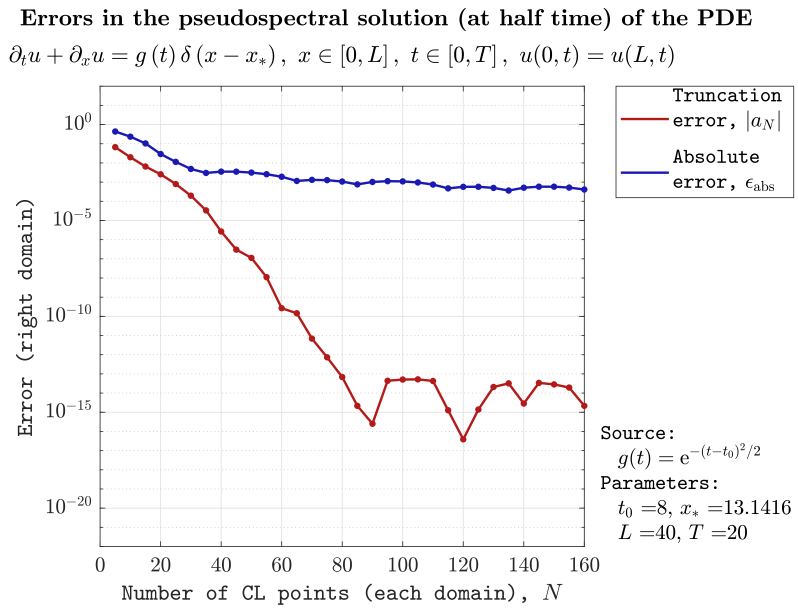

FIGURE B.2. Convergence of the numerical scheme for the problem (B.4.1).

$\partial_t \rho^{\pm} + \partial_V\left((1 - V)\,\rho^{\pm}\right) = 0$ on $\mathscr{D}^{\pm}$, with $\mathscr{D}^{-} = [0, V_*]$ and $\mathscr{D}^{+} = [V_*, L]$, along with the JC $[\rho]_* = \frac{1-L}{1-V_*}\rho(L,t)$.

An example solution for Gaussian initial data centered at $V = 0.3$ is displayed in Figure B.3, and the numerical convergence in Figure B.4. In the latter, we plot—again for the numerical solution $\boldsymbol{\rho}$ at the final time—the truncation error as well as (in the absence of an exact solution) what we refer to as the conservation error, $\epsilon_{\rm cons} = |1 - \int_{\mathscr{I}} {\rm d}V\,\rho(V,t)|$, which simply measures how far we are from exact conservation of probability. Both of these exhibit exponential convergence. The integral in $\epsilon_{\rm cons}$ is computed as a sum over both domains, $\int_{\mathscr{I}} {\rm d}V\,\rho = \int_{\mathscr{D}^-} {\rm d}V\,\rho + \int_{\mathscr{D}^+} {\rm d}V\,\rho$, and numerically performed on each using a standard pseudospectral quadrature method (as in, *e.g.*, Chapter 12 of Ref. [Trefethen 2001]).

This procedure can readily be complexified with the inclusion of a diffusion term, and indeed we will shortly turn to purely diffusion (heat-type equation) problems in the following section.

B.4.3. Advection-type equations in other applications. Another advection-type application in which one may be tempted to try applying some form the PwP method is

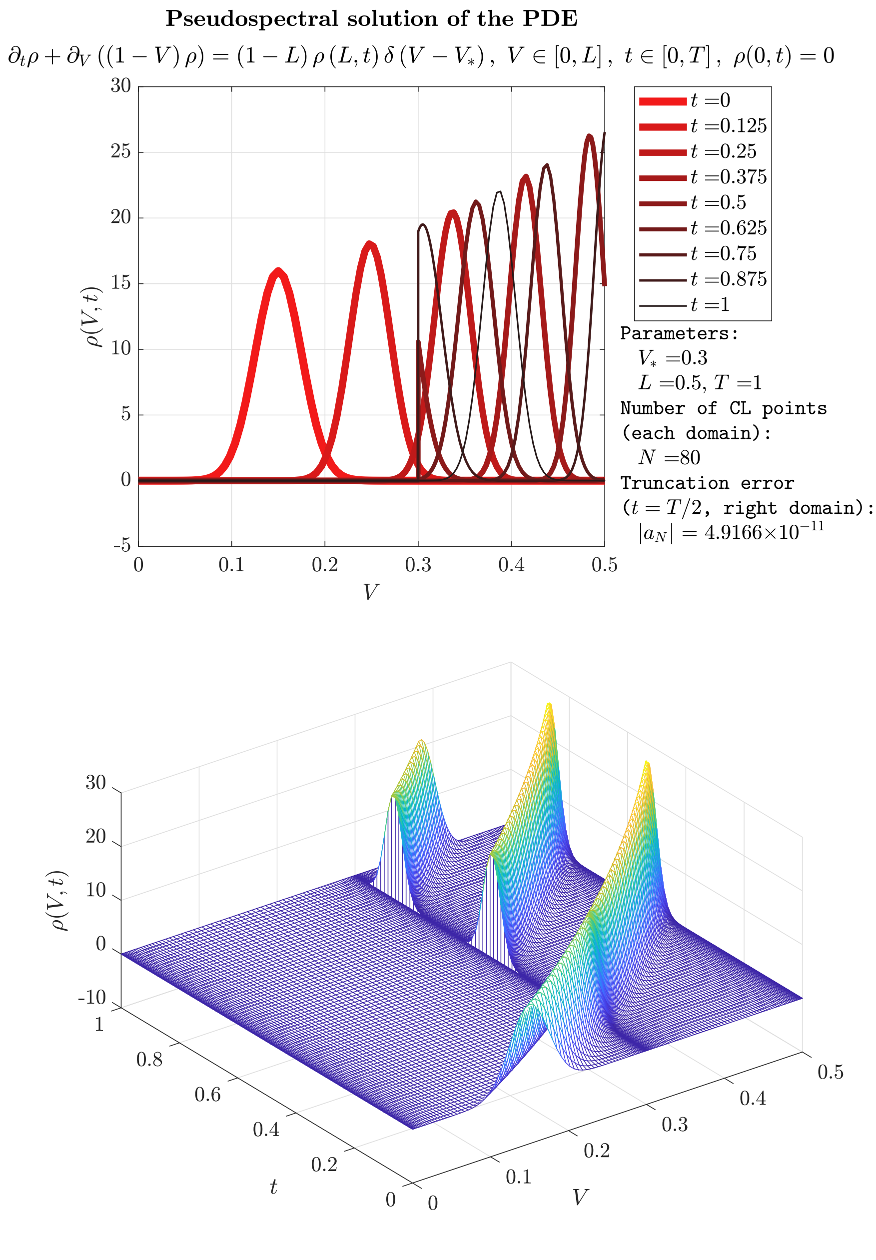

Figure B.3. Solution of (B.4.4) with (normalized) Gaussian initial data centered at $V = 0.3$.

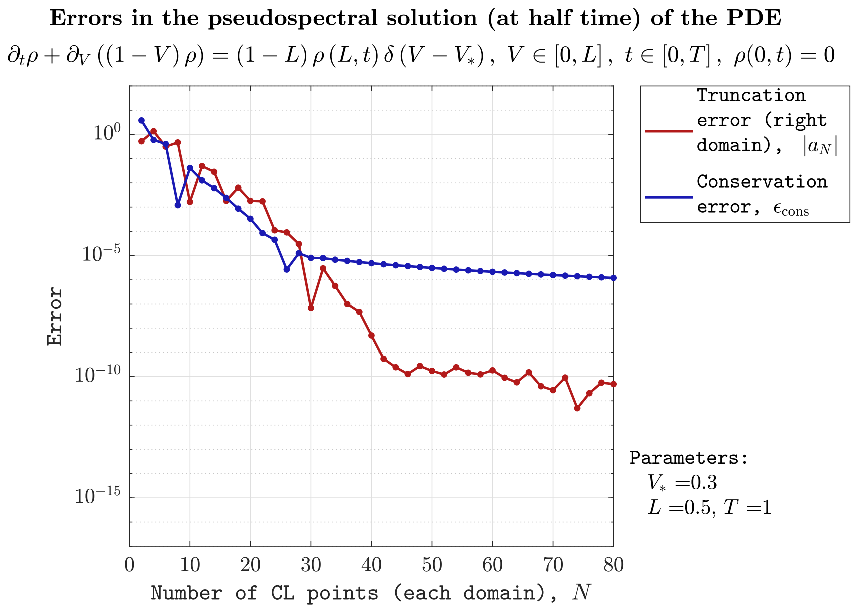

FIGURE B.4. Convergence of the numerical scheme for the problem (B.4.4).

the shallow water equations. Setting the gravitational acceleration to 1, these read:

$$\partial_t \begin{bmatrix} h \\ hu \end{bmatrix} + \partial_x \begin{bmatrix} hu \\ hu^2 + \frac{1}{2}h^2 \end{bmatrix} = \begin{bmatrix} 0 \\ -h\partial_x B \end{bmatrix}, \tag{B.4.5}$$

where $B(x)$ is the elevation of the bottom topography, $h(x,t)$ is the fluid depth above the bottom and $u(x,t)$ is the velocity. If the topography is discontinuous, *i.e.* if $B \notin C^0(\mathbb{R})$, then the RHS of (B.4.5) will be distributional; this can happen, *e.g.*, if the bottom is a step (if $B = \Theta$, then the RHS is $\left[\begin{smallmatrix} 0 \\ -h\delta \end{smallmatrix}\right]$), a wall etc. However, the problem with applying the PwP method here is that (B.4.5) is nonlinear, and so one encounters precisely the sorts of issues detailed at the end of the preceding section. Indeed, explicit numerical solutions that have been obtained for (B.4.5) in the literature [Bernstein et al. 2016; Zhou et al. 2002] qualitatively indicate that a PwP-type decomposition as described here would be inadequate (and, anyway, nonsensical mathematically) for such problems.

B.5. Parabolic PDEs

We begin by analyzing the standard heat equation and then move on to an application in finance which includes two (time-dependent) singular source terms.

B.5.1. Heat equation. Let us consider now the $(1+1)$-dimensional heat equation for $u(x,t)$ with a constant point source at a time-dependent location $x = x_p(t)$, with Dirichlet boundary conditions:

$$\begin{cases} \partial_t u - \partial_x^2 u = \lambda\delta\left(x - x_p\left(t\right)\right) \,, & x \in \mathscr{I} = [a,b] \,, \ \ t > 0 \,, \\ u\left(x,0\right) = 0 \,, & u\left(a,t\right) = \alpha, \ u\left(b,t\right) = \beta \,. \end{cases} \tag{B.5.1}$$

In this case, we do not have the exact solution.

This problem is treated in [Tornberg and Engquist 2004] using a delta function approximation procedure, with the following setup: $\mathscr{I} = [0,1]$, $\alpha = 0 = \beta$ and $\lambda = 10$; constant-valued and sinusoidal point source locations $x_p(t)$ are considered. We implement here the same, using our PwP method: we decompose $u = u^-\Theta^- + u^+\Theta^+$ where $\Theta^\pm = \Theta(\pm(x - x_p(t)))$. Inserting this into (B.5.1), we get homogeneous PDEs $\partial_t u^\pm - \partial_x^2 u^\pm = 0$ to the left and right of the singularity, $x \in \mathscr{D}^- = [0, x_p(t)]$ and $x \in \mathscr{D}^+ = [x_p(t), 1]$ respectively; additionally, we have the following JCs: $[u]_p = 0$ and $[\partial_x u]_p = -\lambda$.

The details of the numerical scheme are given in Subsection B.9.3, and results for zero initial data in Figures B.5 and B.6.

B.5.2. Heat-type equations in finance. We consider a model of price formation initially proposed in Ref. [Lasry and Lions 2007]; see also Refs. [Achdou et al. 2014; Burger et al. 2013; Caffarelli et al. 2011; Markowich et al. 2009; Pietschmann 2012]. This model describes the density of buyers $f_{\mathrm{B}}(x,t)$ and the density of vendors $f_{\mathrm{V}}(x,t)$ in a system, as functions of of the bid or, respectively, ask price $x \in \mathbb{R}$ for a certain good being traded between them, and time $t \in [0,\infty)$.

The idea is that when a buyer and vendor agree on a price, the transaction takes place; the buyer then becomes a vendor, and vice-versa. However, it is also assumed that there exists a fixed transaction fee $a \in \mathbb{R}$. Consequently, the actual buying price is $x + a$, and so the (former) buyer will try to sell the good at the next trading event not for the price x, but for $x + a$. Similarly, the profit for the vendor is actually $x - a$, and so he/she would not be willing to pay more than $x - a$ for the good at the next trading event. In time, this system should achieve an equilibrium.

Mathematically, the dynamics of the buyer/vendor densities is assumed to be governed by the heat equation with a certain source term. The source term in each case is simply the (time-dependent) transaction rate $\lambda(t)$, corresponding to the flux of buyers and vendors, at the particular price where the trading event occurs, shifted accordingly by the transaction cost. Thus the system is described by

$$\begin{cases} \left(\partial_t - \partial_x^2\right) f_{\mathrm{B}} = \lambda\left(t\right)\delta\left(x - \left(x_p\left(t\right) - a\right)\right) \,, & \text{for } x < x_p\left(t\right) \,, \\ f_{\mathrm{B}} = 0 \,, & \text{for } x > x_p\left(t\right) \,, \end{cases} \tag{B.5.2}$$

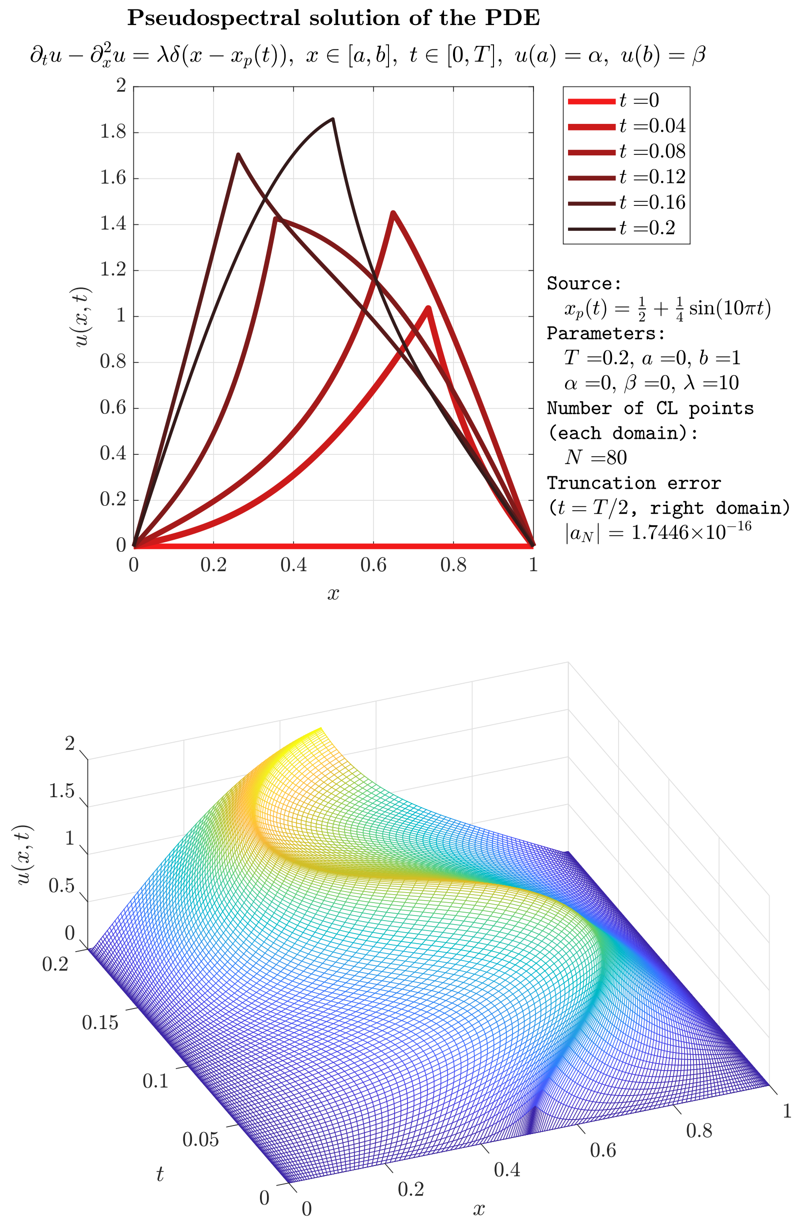

Figure B.5. Solution of the problem (B.5.1) with zero initial data.

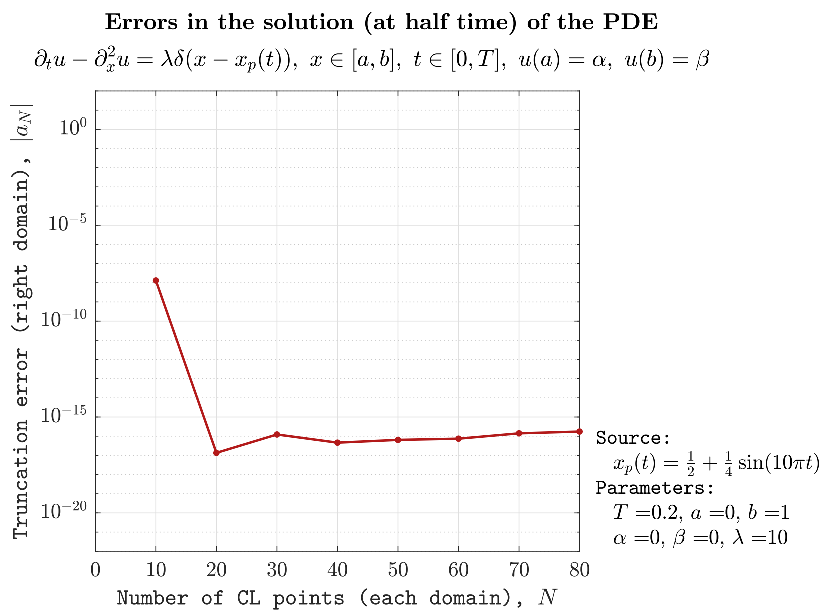

FIGURE B.6. Convergence of the numerical scheme for the problem (B.5.1).

and

$$\begin{cases} \left(\partial_t - \partial_x^2\right) f_{\mathrm{V}} = \lambda\left(t\right)\delta\left(x - \left(x_p\left(t\right) + a\right)\right) , & \text{for } x > x_p\left(t\right) , \\ f_{\mathrm{V}} = 0 , & \text{for } x < x_p\left(t\right) , \end{cases} \tag{B.5.3}$$

where the free boundary $x_p(t)$ represents the agreed price of trading at time t, and the transaction rate is $\lambda(t) = -\partial_x f_{\mathrm{B}}(x_p(t), t) = \partial_x f_{\mathrm{V}}(x_p(t), t)$. (NB: The functional form of $\lambda(t)$ is uniquely fixed simply by the requirement that the two densities are conserved, *i.e.* $\partial_t \int \mathrm{d}x f_{\mathrm{B}} = 0 = \partial_t \int \mathrm{d}x f_{\mathrm{V}}$, under the assumption that we have homogeneous Neumann BCs at the left and right boundaries respectively.) Now, we can actually combine this system into a single problem for the difference between buyer and vendor densities,

$$f = f_{\mathrm{B}}\Theta\left(-\left(x - x_p\left(t\right)\right)\right) - f_{\mathrm{V}}\Theta\left(x - x_p\left(t\right)\right) . \tag{B.5.4}$$

The "spatial" (*i.e.* price) domain can be taken to be bounded, and homogeneous Neumann BCs are assumed at the boundaries. Thus the problem we are interested in is:

$$\begin{cases} \partial_t f - \partial_x^2 f = \lambda\left(t\right)\left(\delta\left(x - x_{p_-}\left(t\right)\right) - \delta\left(x - x_{p_+}\left(t\right)\right)\right) , & x \in \mathscr{I} = \left[0, 1\right] , \ t > 0 , \\ f\left(x, 0\right) = f_{\mathrm{I}}\left(x\right) , \ \ f \gtrless 0 \ \text{ for } \ x \lessgtr x_p\left(t\right) , & \partial_x f\left(0, t\right) = \partial_x f\left(1, t\right) = 0 . \end{cases} \tag{B.5.5}$$

where $\lambda(t) = -\partial_x f(x_p(t), t)$, and we have defined $x_{p\pm}(t) = x_p(t) \pm a$. Moreover, one can show that from this setup, it follows that the free boundary evolves via

$$\dot{x}_p\left(t\right) = \frac{\partial_x^2 f\left(x_p\left(t\right), t\right)}{\lambda\left(t\right)} . \tag{B.5.6}$$

In this case, we have not one but two singular source locations on the RHS of the PDE. Hence, in order to implement the PwP method, we must here divide the spatial domain $\mathscr{I}$ into three disjoint regions, with the two singularity locations at their interfaces: $\mathscr{I} = \mathscr{D}^- \cup \mathscr{D}^0 \cup \mathscr{D}^+$ with $\mathscr{D}^- = [0, x_{p_-}(t)]$, $\mathscr{D}^0 = [x_{p_-}(t), x_{p_+}(t)]$ and $\mathscr{D}^+ = [x_{p_+}(t), 1]$. Then, we decompose $f = f^-\Theta^- + f^0\Theta^0 + f^+\Theta^+$ with $\Theta^- = \Theta(-(x - x_{p_-}(t)))$, $\Theta^0 = \Theta(x - x_{p_-}(t)) - \Theta(x - x_{p_+}(t))$ and $\Theta^+ = \Theta(x - x_{p_+}(t))$. Inserting this into the PDE (B.5.5), we get homogeneous problems $(\partial_t - \partial_x^2)f^\sigma = 0$ on $\mathscr{D}^\sigma$ for $\sigma \in \{0, \pm\}$, along with the JCs $[f]_{p_\pm} = 0$ and $[\partial_x f]_{p_\pm} = \pm\lambda(t)$.

Before proceeding to the numerical implementation, we note that it is possible to derive an exact stationary (*i.e.* $t \to \infty$) solution of the problem (B.5.5). In particular, denoting the (time-conserved) number of buyers and vendors, respectively, by $N_{\rm B} = \int_0^{x_p} \mathrm{d}x\, f$ and $N_{\rm V} = -\int_{x_p}^1 \mathrm{d}x\, f$, one can show that in the stationary ($t \to \infty$) limit,

$$\begin{cases} N_{\rm B} = & -\lambda^{\rm stat} a (x_p^{\rm stat} - a/2) , \\ N_{\rm V} = & -\lambda^{\rm stat} a (1 - x_p^{\rm stat} - a/2) , \end{cases} \tag{B.5.7}$$

$$\Leftrightarrow \begin{cases} \lambda^{\rm stat} = & \left[-\left(N_{\rm B} + N_{\rm V}\right)\right] / \left[a\left(1-a\right)\right] , \\ x_p^{\rm stat} = & \left[2N_{\rm B} + a\left(N_{\rm V} - N_{\rm B}\right)\right] / \left[2\left(N_{\rm B} + N_{\rm V}\right)\right] , \end{cases} \tag{B.5.8}$$

which we can use to determine the exact stationary solution

$$\lim_{t\to\infty} f\left(x, t\right) = f^{\rm stat}\left(x\right) = \begin{cases} -\lambda^{\rm stat} a , & \text{for } 0 \leq x < x_{p_-}^{\rm stat} , \\ \lambda^{\rm stat}\left(x - x_p^{\rm stat}\right) , & \text{for } x_{p_-}^{\rm stat} \leq x \leq x_{p_+}^{\rm stat} , \\ \lambda^{\rm stat} a , & \text{for } x_{p_+}^{\rm stat} < x \leq 1 . \end{cases} \tag{B.5.9}$$

The problem (B.5.5) is solved numerically in Ref. [Markowich et al. 2009] (see also section 2.5.2 of [Pietschmann 2012]) using (Gaussian) delta function approximations for the source on an equispaced computational grid. We implement here using our PwP method the exact same setup: in particular, we take a transaction fee of $a = 0.1$ and initial data $f_{\rm I}(x) = \frac{875}{6}x^3 - \frac{700}{3}x^2 + \frac{175}{2}x$. (NB: Despite the fact that this does not actually satisfy homogeneous Neumann BCs, the numerical evolution will force it to.) Analytically, we have $x_p(0) = \frac{3}{5}$ and $\lambda(0) = 35$. Also, using (B.5.8), we have $\lambda^{\rm stat} = -\frac{8855}{162}$ and $x_p^{\rm stat} = \frac{731}{1012} \approx 0.7223$. As we evolve forward in time, we use Chebyshev polynomial interpolation to determine the transaction rate $\lambda(t)$ (*i.e.* the negative of the spatial derivative of the solution at $x_p(t)$) as well as the evolution of $x_p(t)$ via (B.5.6).

The numerical scheme is given in Subsection B.9.3, and results in Figures B.7 and B.8. In particular, in Figure B.7 we show the numerical solution for f, and in Figure B.8, the price as a function of time as well as the numerical convergence rates. For the latter, we plot not only the truncation error but also the absolute error with the stationary solution (B.5.9), in this case, using the infinity norm: $\epsilon_{\text{abs}} = ||\boldsymbol{f} - \boldsymbol{f}^{\text{stat}}||_\infty$. Of course, since we can only evolve the solution up to a finite time (which we choose to be $t = T = 1$), we should not expect this to converge to zero; however, its decline with increasing N nevertheless serves to illustrate a good validation of our results.

We remark that our numerical implementation here not only requires an order of magnitude fewer grid points than that of Ref. [Markowich et al. 2009], but in fact yields convergence to the *correct* stationary solution while that of Ref. [Markowich et al. 2009] *does not*. Indeed, in the latter, not only are more points required (essentially due to the necessity of resolving well enough the Gaussian-approximated delta functions) but the scheme actually fails, even so, to approach (B.5.9) as well as ours by the same finite time, $t = T = 1$. (To wit, Ref. [Markowich et al. 2009] obtains $x_p \to 0.71$ in the large t limit, instead of the correct value, 0.7223, which we achieve with our PwP method as shown in Figure B.8.)

B.6. Second order hyperbolic PDEs

We move on to consider in this section second order hyperbolic problems. In particular, we first solve the standard $(1+1)$-dimensional elastic wave equation, taking a delta derivative source. Afterwards, we discuss possible physical applications of this and obstacles thereto—including problems in gravitational physics and seismology.

B.6.1. Wave equation. Let us consider the the elastic wave equation, in the form of the following simplified $(1+1)$-dimensional problem for $u(x,t)$ with a delta function derivative source at a fixed point $x_* \in \mathscr{I} = [0, L]$, and homogeneous Dirichlet boundary conditions:

$$\begin{cases} \partial_t^2 u - \partial_x^2 u = g\left(t\right)\delta'\left(x - x_*\right)\,, & x \in \mathscr{I} = [0, L]\,,\ \ t > 0\,, \\ u\left(x, 0\right) = 0\,,\ \ \partial_t u\left(x, 0\right) = 0\,, & u\left(0, t\right) = 0 = u\left(L, t\right)\,. \end{cases} \tag{B.6.1}$$

It is actually possible to derive an exact solution for this problem on an unbounded domain $\mathscr{I} = \mathbb{R}$. For the interested reader, the procedure is explained in Appendix D of [Oltean, Sopuerta, et al. 2019]. For concreteness we take a simple sinusoidal source time

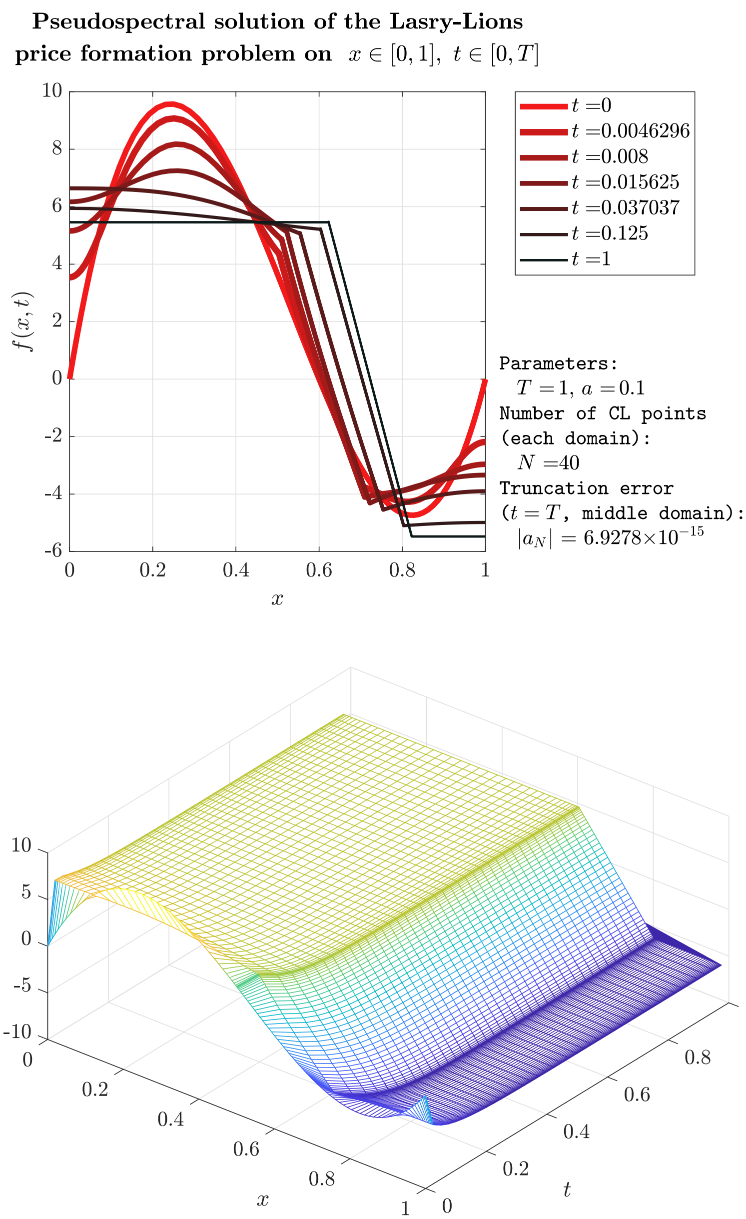

FIGURE B.7. Solution of the problem (B.5.5).

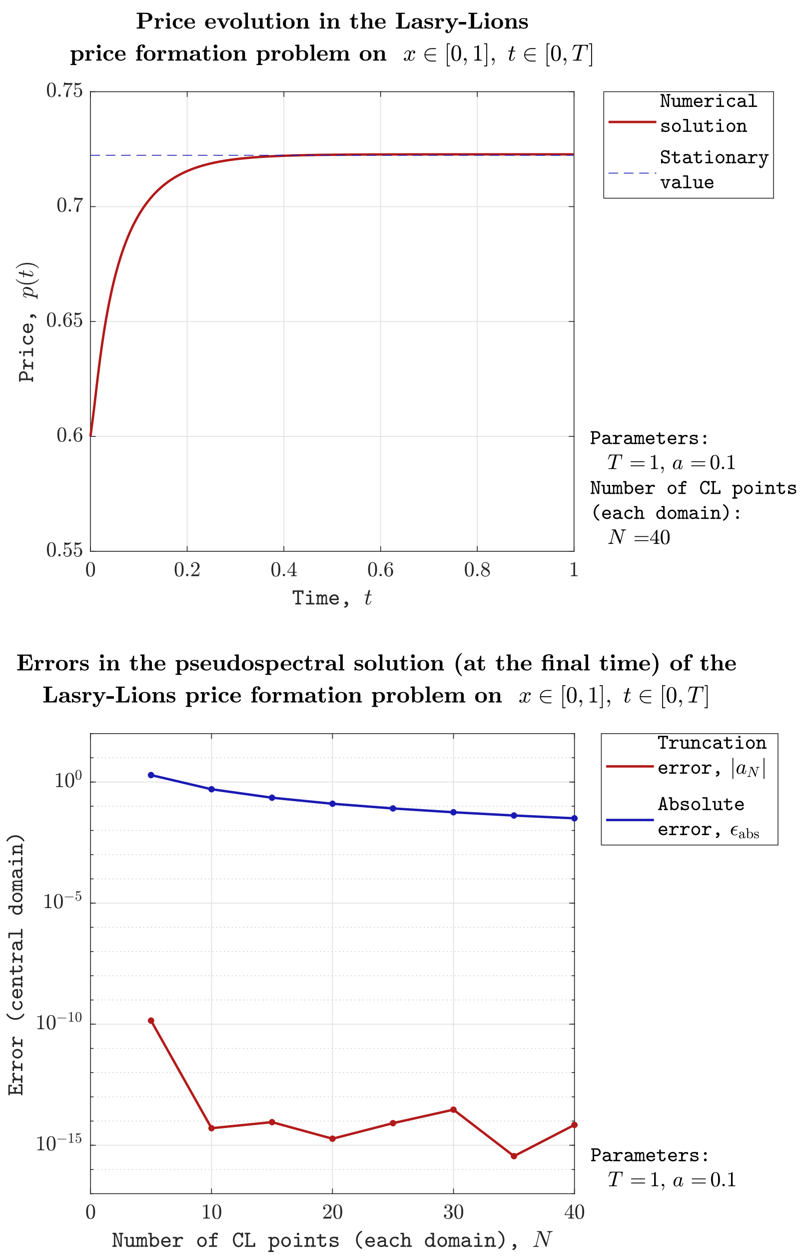

FIGURE B.8. Price evolution and convergence of the numerical scheme for the problem (B.5.5).

function $g(t) = \kappa \sin(\omega t)$, in which case the exact solution reads:

$$u_{\text{ex}}(x,t) = \kappa \Bigg[\frac{1}{4} \sum_{\sigma=\pm} \sigma \operatorname{sgn}(x - x_* + \sigma t) \sin(\omega(x - x_* + \sigma t)) - \frac{1}{2} \operatorname{sgn}(x - x_*) \cos(\omega(x - x_*)) \sin(\omega t) \Bigg], \tag{B.6.2}$$

where $\operatorname{sgn}(\cdot)$ is the sign function, with the property $\mathrm{d}(\operatorname{sgn}(x))/\mathrm{d}x = 2\mathrm{d}\Theta(x)/\mathrm{d}x = 2\delta(x)$.

To solve (B.6.1) numerically, we implement the now familiar PwP decomposition: $u = u^-\Theta^- + u^+\Theta^+$ where $\Theta^\pm = \Theta(\pm(x - x_*))$. Inserting this into (B.6.1), we get homogeneous PDEs $\partial_t^2 u^\pm - \partial_x^2 u^\pm = 0$ to the left and right of the singularity, $x \in \mathscr{D}^- = [0, x_*]$ and $x \in \mathscr{D}^+ = [x_*, L]$ respectively, along with the JCs $[u]_p = -g(t)$ and $[\partial_x u]_p = 0$. We now proceed by recasting (B.6.1) as a first-order hyperbolic system for $\vec{U} = [u \;\; v \;\; w]^{\mathrm{T}}$ with $v = \partial_x u$ and $w = \partial_t u$, as

$$\partial_t \vec{U} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \partial_x \vec{U} + \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \vec{U} \quad \text{on } \mathscr{D}^\pm, \quad \left[\vec{U}\right]_p = \begin{bmatrix} -g \\ 0 \\ -\dot{g} \end{bmatrix}. \tag{B.6.3}$$

The numerical scheme is given in Subsection B.9.4, and results in Figures B.9 and B.10. The absolute error is again computed in the infinity norm on the CL grids: $\epsilon_{\text{abs}} = ||\boldsymbol{u} - \boldsymbol{u}_{\text{ex}}||_\infty$.

The same problem (B.6.1) is considered numerically in Ref. [Petersson and Sjogreen 2010], but using a different (polynomial) source function $g(t)$, and a discretization procedure for the delta function (derivatives) on the computational grid (carried out in such a way that the distributional action thereof yields the expected result on polynomials up to a given degree). With our PwP method here, we obtain the same order of magnitude of the (absolute) error in the numerical solution as that in Ref. [Petersson and Sjogreen 2010] for the same (order of magnitude of) number of grid points; however the drawback of the “discretized delta” method of Ref. [Petersson and Sjogreen 2010], in contrast to the PwP method, is that the solution in the former is visibly quite poorly resolved close to the singularity.

We add that we have also carried out the solution to the problem shown in Figure B.9 using higher-order (from second up to eighth order) finite-difference time evolution schemes. These yield no visible improvement (at any order tried) in either the absolute or the truncation error relative to the first-order time evolution results. Thus the spacial pseudospectral grid appears to control the total level of the error, with a higher-order scheme for the time evolution producing, at least in this case, no greater benefits.

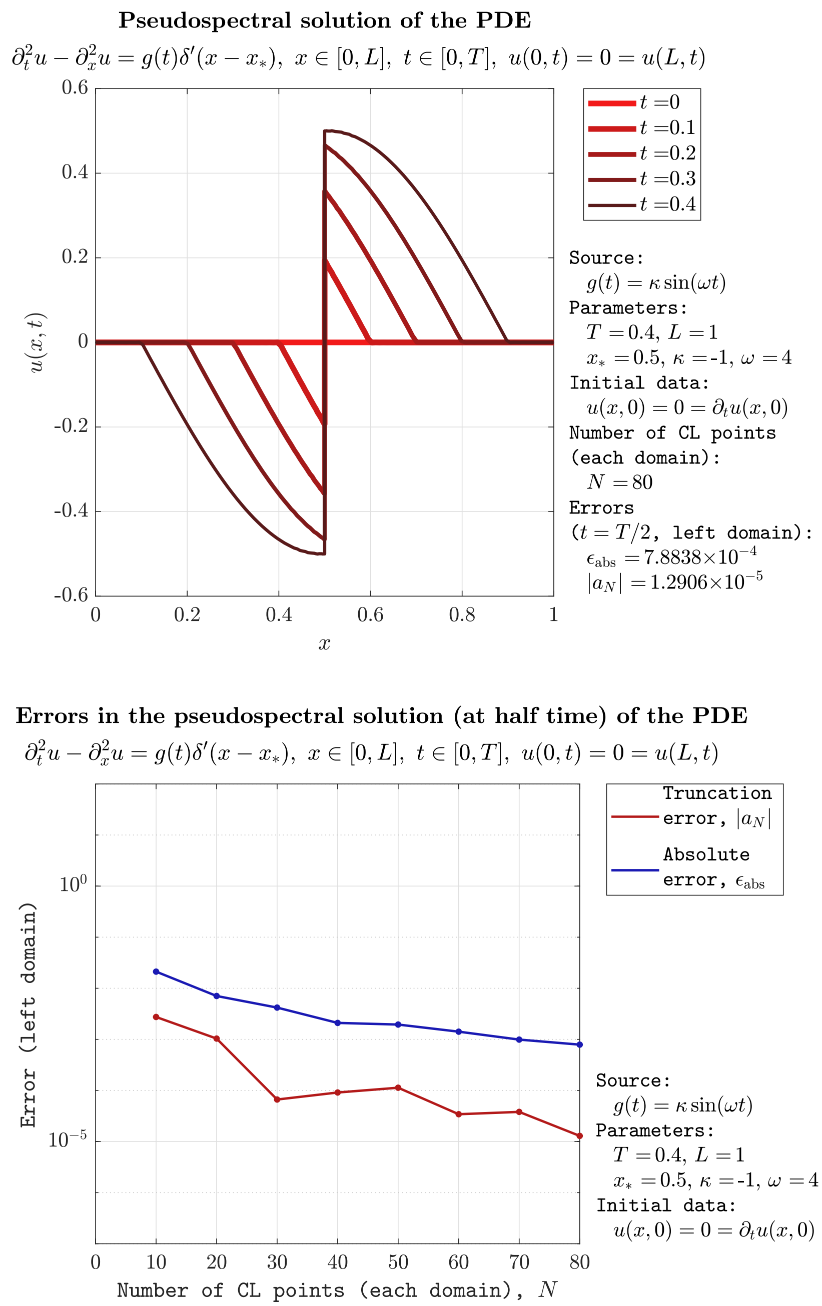

FIGURE B.9. Solution and convergence of the numerical scheme for the problem (B.6.1).

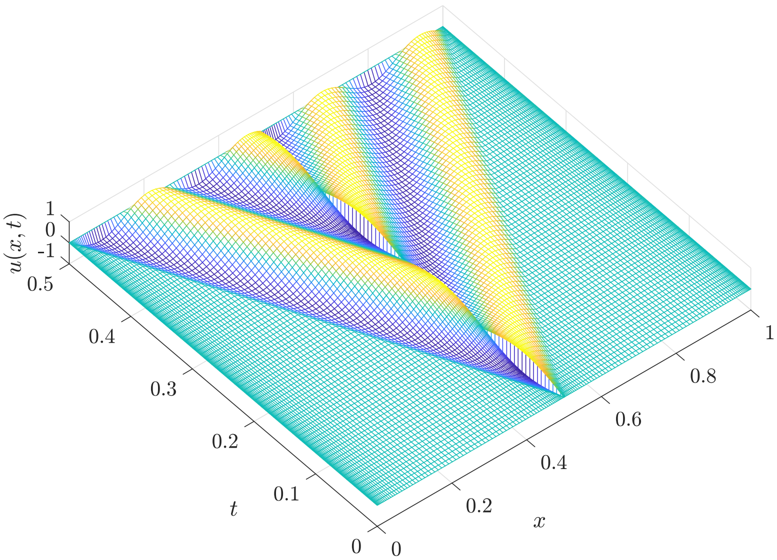

FIGURE B.10. Solution (with $N = 80$) of the same problem as in Figure B.9 but using $\omega = 24$.

B.6.2. Wave-type equations in physical applications. As we have mentioned, our numerical studies in this paper are largely motivated by their applicability to the computation of the self-force in gravitational physics. There one encounters different levels of complexity of this problem, the simplest being that of the self-force due to a scalar field—as a conceptual testbed for the more complicated and realistic problem of the full self-force of the gravitational field—in a fixed (non-dynamical) black hole spacetime. This can refer to a non-spinning (Schwarzschild-Droste) black hole, where the problem is of the form (B.1.1) where $\mathcal{L} = \partial_t^2 - \partial_x^2 + V$ is just a simple $(1+1)$-dimensional wave operator with some known potential V and a source $S = f\delta_{(p)}$, with $\dim(\mathscr{I}) = 1$. We recognize this now as quite typical for the application of our PwP method, and indeed this has been done with success in the past [Canizares 2011; Canizares and Sopuerta 2009b, 2014, 2011a,b; Canizares, Sopuerta, and Jaramillo 2010b; Jaramillo, Sopuerta, et al. 2011; Oltean, Sopuerta, et al. 2017]. As we briefly remarked in the Introduction, the main difference between most of these works and our numerical schemes throughout this paper is that for the time evolution, rather than relying on finite-difference methods, the former made use of the method of lines. This can be quite well-suited especially for these types of $(1+1)$-dimensional hyperbolic problems, which can be formulated in terms of characteristic fields propagating along the two lightcone directions ($t \pm x =$ const.). The imposition of the JCs is then achieved quite simply in this setting by just evolving, in the left domain, the characteristic field propagating towards the right and relating it (via the JC) to the value of the characteristic field propagating towards the left in the right domain. For the interested reader, this kind of procedure is described in detail in Chapter 3 of Ref. [Canizares 2011].

We could also consider the scalar self-force problem in a spinning (Kerr) black hole spacetime, however the issue there—owing to the existence of fewer symmetries in the problem than in the non-spinning case—is that $\dim(\mathscr{I}) = 2$ in the time domain (with a more complicated second-order hyperbolic operator $\mathcal{L}$); however, this could be remedied for a possible PwP implementation by passing to the frequency domain, which transforms (B.1.1) to an ODE (with $\dim(\mathscr{I}) = 1$, $\mathscr{V} = \emptyset$, and again, a simple source $S = f\delta_{(p)}$).

The application of the PwP method to the full gravitational self-force is a subject of ongoing work, however (modulo certain technical problems relating to the gauge choice, which we will not elaborate upon here) in the Schwarzschild-Droste case it essentially reduces to solving the same type of problem (B.1.1) with $\dim(\mathscr{I}) = 1$ and $S = f\delta_{(p)} + g\delta'_{(p)}$. The equivalent problem in the Kerr case once again suffers from the issue that $\dim(\mathscr{I}) = 2$ in the time domain, so the PwP method cannot be applied there except after a transformation to the frequency domain (which produces $\dim(\mathscr{I}) = 1$ and $S = f\delta_{(p)} + g\delta'_{(p)} + h\delta''_{(p)}$ in this case).

Outside of gravitational physics, another setting where the PwP technique could also possibly prove useful is in seismology. There, however, the modeling of seismic waves [Aki and Richards 2009; Madariaga 2007; Petersson and Sjogreen 2010; Romanowicz and Dziewonski 2007; Shearer 2009] typically involves equations of the form (B.3.44) with 3-dimensional delta functions (*i.e.* $\dim(\mathscr{I}) = \bar{n} = 3$, usually referring to the 3 dimensions of ordinary space) which, as we have amply discussed in relation thereto, are not directly amenable to a PwP-type approach as such. However, the methods outlined in this paper might be of some use if symmetries or other simplifying assumptions can, in a situation of interest, reduce the dimension of the distributional source to 1 (as an alternative to delta function approximation procedures, which are common practice in this area as well).

B.7. Elliptic PDEs

Finally, we consider in this section the elliptical problem appearing in section 4.3 of Ref. [Tornberg and Engquist 2004]: namely, the Poisson equation on a square of side length 2 centered on the origin in $\mathbb{R}^2$, with a simple (negative) one-dimensional delta function source supported on the circle of radius $r_* = \frac{1}{2}$,

$$\begin{cases} \triangle_{\mathbb{R}^2} u = -\delta\left(r - r_*\right) \,, & \text{on } \mathscr{U} = [-1,1]\times[-1,1] \subset \mathbb{R}^2 \,, \\ u = 1 - \frac{1}{2}\log\left(2r\right) \,, & \text{on } \partial\mathscr{U} \,. \end{cases} \tag{B.7.1}$$

In this case, the polar symmetry of the PDE entails that the solution will only depend on the radial coordinate r (which in this case notationally substitutes the x coordinate in antecedent sections). Indeed, (B.7.1) has an exact solution which is simply given by

$$u_{\text{ex}} = 1 - \tfrac{1}{2}\log(2r)\,\Theta\,(r - r_*)\ . \tag{B.7.2}$$

We can use the fact that in polar coordinates, $\triangle_{\mathbb{R}^2} = \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\theta^2$, and so numerically all we need to do is solve $(\partial_r^2 + \frac{1}{r}\partial_r)u(r) = -\delta(r - r_*)$ for a given $\theta \in [0, 2\pi]$, where the value of θ will determine $\{r\} = \mathscr{I} = [0, L]$ and hence the BC at $u(L)$ (that is, on $\partial\mathscr{U}$), and repeat over some set of discrete θ values in case the entire numerical solution on the (r, θ)-plane is desired.

Thus, we simply implement the PwP method here by writing $u = u^-\Theta^- + u^+\Theta^+$ for $\Theta^\pm = \Theta(r - r_*)$, whereby we obtain the homogeneous equations $(\partial_r^2 + \frac{1}{r}\partial_r)u^\pm = 0$ along with the JCs $[u]_* = 0$ and $[\partial_r u]_* = -1$.

The detailed numerical scheme is given in Subsection B.9.5, and results in Figure B.11. In this case, we simply plot the errors along the positive x-axis in $\mathbb{R}^2$ on the CL grids: in addition to the right-domain truncation error, we also show (as is done in Ref. [Tornberg and Engquist 2004]) the absolute error in both the l^1-norm, $\epsilon_{\text{abs}}^{(1)} = ||\boldsymbol{u} - \boldsymbol{u}_{\text{ex}}||_1$, as well as in the infinity norm, $\epsilon_{\text{abs}}^{(\infty)} = ||\boldsymbol{u} - \boldsymbol{u}_{\text{ex}}||_\infty$. Up to $N \approx 20$, we observe the typical (exponential) spectral convergence of all three errors, with a significant (by a few orders of magnitude) improvement over the results of Ref. [Tornberg and Engquist 2004] (using delta function approximations) for the latter two.

B.8. Conclusions

We have expounded in this paper a practical approach—the “Particle-without-Particle” (PwP) method—for numerically solving differential equations with distributional sources; to summarize, one does this by breaking up the solution into (regular function) pieces supported between—plus, if necessary, at—singularity (“particle”) locations, solving sourceless (homogeneous) problems for these pieces, and then matching them via the appropriate “jump” (boundary) conditions effectively substituting the original singular source. Building upon its successful prior application in the specific context of the self-force problem in general relativity, we have here generalized this method and have shown it to be viable for any linear partial differential equation of arbitrary order, with the provision that the distributional source is supported only on a one-dimensional subspace of the total problem domain. Accordingly, we have demonstrated its usefulness by solving first and second order hyperbolic problems, with applications in neuroscience and acoustics, respectively; parabolic problems, with applications in finance; and finally a simple elliptic problem. In particular, the numerical schemes we have employed for carrying these out have been based on pseudospectral collocation methods on Chebyshev-Lobatto grids. Generally speaking, our results have yielded varying degrees of improvement in the numerical convergence

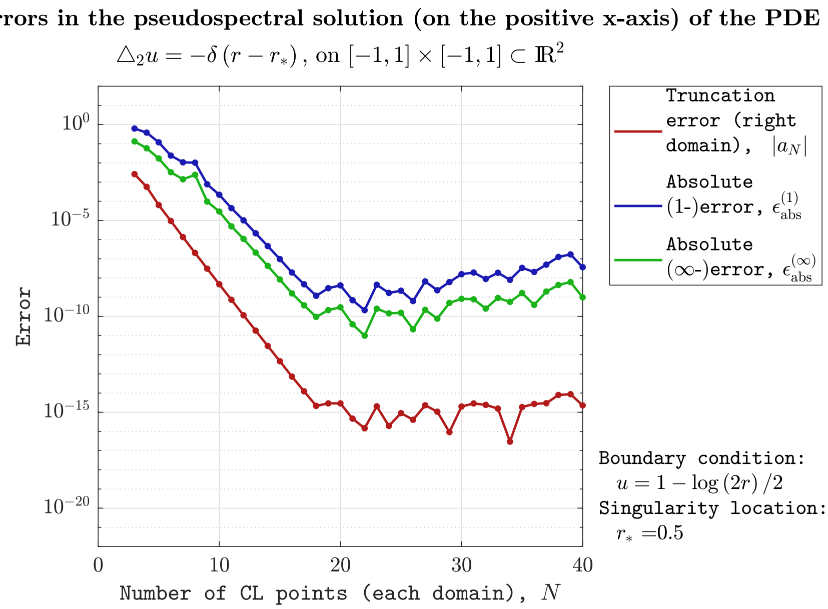

FIGURE B.11. Convergence of the pseudospectral numerical scheme for the problem (B.7.1).

rates relative to other methods in the literature that have been attempted for solving these problems (typically relying on delta function approximation procedures on the computational grid).

We stress once more that the main limitations of the our PwP method as developed here are that it is only applicable to *linear* problems with *one*-dimensionally supported distributional sources. Thus, interesting lines of inquiry for future work might be to explore—however/if at all possible—extensions or adaptations of these ideas (a) to nonlinear PDEs, which would require working with nonlinear theories of distributions (having potential applicability to problems such as, *e.g.*, the shallow-water equations with discontinuous bottom topography); (b) to more complicated sources than the sorts considered in this paper, perhaps even containing higher-dimensional distributions but possibly also requiring additional assumptions, such as symmetries (which might be useful for problems such as, *e.g.*, seismology models with three-dimensional delta function sources).

B.9. Appendix: pseudospectral numerical schemes

B.9.1. Pseudospecteal collocation methods. We use this subsection to describe very cursorily the PSC methods used for the numerical schemes in this work and to introduce some notation in relation thereto. For good detailed expositions see, for example, Refs. Boyd 2001; Peyret 2002; Trefethen 2001.

We work on Chebyshev-Lobatto (CL) computational grids. On any domain $[a, b] = \mathscr{D} \subseteq \mathscr{I}$, these comprise the (*non-uniformly* spaced) set of N points $\{X_i\}_{i=0}^{N} \subset \mathscr{D}$ obtained by projecting onto $\mathscr{D}$ those points located at equal angles on a hypothetical semicircle having $\mathscr{D}$ as its diameter. That is to say, the CL grid on the "standard" spectral domain $\mathscr{D}^{\mathrm{s}} = [-1, 1]$ is given by

$$X_i^{\mathrm{s}} = -\cos\left(\frac{\pi i}{N}\right), \quad \forall 0 \leq i \leq N, \tag{B.9.1}$$

which can straightforwardly be transformed (by shifting and stretching) to the desired grid on $\mathscr{D}$. For any function $f : \mathscr{D} \to \mathbb{R}$ we denote via a subscript its value at the i-th CL point, $f(X_i) = f_i$, and in slanted boldface the vector containing all such values,

$$\boldsymbol{f} = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_N \end{bmatrix}. \tag{B.9.2}$$

There exists an $(N+1) \times (N+1)$ matrix $\mathbb{D}$, the so-called CL differentiation matrix, such that the derivative values of f can be approximated simply by applying it to (B.9.2), *i.e.* $\boldsymbol{f}' = \mathbb{D}\boldsymbol{f}$. For convenience, we also employ the notation $\mathbb{M}(r_{\mathrm{i}} : r_{\mathrm{f}}, c_{\mathrm{i}} : c_{\mathrm{f}})$ to refer to the part of any matrix $\mathbb{M}$ from the r_{i}-th to the r_{f}-th row and from the c_{i}-th to the c_{f}-th column. (A simple ":" indicates taking all rows/columns.)

B.9.2. First-order hyperbolic PDEs. We apply a first order in time finite difference scheme to the homogeneous PDEs; thus, prior to imposing BCs/JCs, the equations become $\frac{1}{\Delta t}(\boldsymbol{u}_{k+1}^{\pm} - \boldsymbol{u}_k^{\pm}) = -\mathbb{D}^{\pm}\boldsymbol{u}_k^{\pm}$, where the vectors $\boldsymbol{u}_k^{\pm}$ contain the values of the solutions on the CL grids at the k-th time step, $\mathbb{D}^{\pm}$ is the CL differentiation matrix on the respective domains, and Δt is our time step. We can rewrite the discretized PDE as $\boldsymbol{u}_{k+1}^{\pm} = \boldsymbol{u}_k^{\pm} - \Delta t \mathbb{D}^{\pm}\boldsymbol{u}_k^{\pm} = \boldsymbol{s}_k^{\pm}$. To impose the BC and JC, we modify the equations as follows:

$$\left[\begin{array}{c} \boldsymbol{u}_{k+1}^{-} \\ \hline \boldsymbol{u}_{k+1}^{+} \end{array}\right] = \left[\begin{array}{c} u_{N,k}^{+} \\ \boldsymbol{s}_k^{-}(2:N+1) \\ \hline u_{N,k}^{-} + g_k \\ \boldsymbol{s}_k^{+}(2:N+1) \end{array}\right]. \tag{B.9.3}$$

Similarly, for our neuroscience application, we discretize the PDE using a first order finite difference scheme: $\frac{1}{\Delta t}(\boldsymbol{\rho}^{\pm}_{k+1} - \boldsymbol{\rho}^{\pm}_{k}) = -\mathbb{D}^{\pm}\boldsymbol{R}^{\pm}_{k}$ where $R^{\pm}_{i,k} = (1 - V^{\pm}_{i})\rho^{\pm}_{i,k}$. Hence, prior to imposing the BC/JC, we have $\boldsymbol{\rho}^{\pm}_{k+1} = \boldsymbol{\rho}^{\pm}_{k} - \Delta t\mathbb{D}^{\pm}\boldsymbol{R}^{\pm}_{k} = \boldsymbol{s}^{\pm}_{k}$. To impose the BC/JC, we just modify the equations accordingly:

$$\left[\begin{array}{c} \boldsymbol{\rho}^{-}_{k+1} \\ \hline \boldsymbol{\rho}^{+}_{k+1} \end{array}\right] = \left[\begin{array}{c} 0 \\ \boldsymbol{s}^{-}_{k}(2:N+1) \\ \hline \rho^{-}_{N,k} + \frac{1-L}{1-V_*}\rho^{+}_{N,k} \\ \boldsymbol{s}^{+}_{k}(2:N+1) \end{array}\right]. \tag{B.9.4}$$

B.9.3. Parabolic PDEs. In these problems, we have moving boundaries for the CL grids (since the location of the singular source is time-dependent). The mapping for transforming the standard (fixed) spectral domain $[-1, 1]$ into an arbitrary (time-dependent) one, say $\mathscr{D} = [a(t), b(t)]$, is given by

$$\mathscr{V} \times [0,1] \to \mathscr{V} \times \mathscr{D} \tag{B.9.5}$$

$$(T, X) \mapsto (t\,(T)\,, x\,(T, X))\ , \tag{B.9.6}$$

where

$$t\,(T) = T\,, \tag{B.9.7}$$

$$x\,(T, X) = \frac{b-a}{2}X + \frac{a+b}{2}\,. \tag{B.9.8}$$

For transforming back, we have

$$\mathscr{V} \times \mathscr{D} \to \mathscr{V} \times [0,1] \tag{B.9.9}$$

$$(t, x) \mapsto (T\,(t)\,, X\,(t, x))\ , \tag{B.9.10}$$

where

$$T\,(t) = t\,, \tag{B.9.11}$$

$$X\,(t, x) = \frac{2x - a - b}{b - a}\,. \tag{B.9.12}$$

Thus, for any function $f(t, x)$ in these problems, we must take care to express the time partial using the chain rule as

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial T}\frac{\partial T}{\partial t} + \frac{\partial f}{\partial X}\frac{\partial X}{\partial t} \tag{B.9.13}$$

$$= \frac{\partial f}{\partial T} - \frac{2}{(b-a)^2}\left[(b-x)\,\dot{a} + (x-a)\,\dot{b}\right]\frac{\partial f}{\partial X}\,, \tag{B.9.14}$$

where in the second line we have used (B.9.11)-(B.9.12).

Now, let us use this to formulate the numerical schemes for our problems—first, for the heat equation. Let $\mathbb{D}_k^{\pm}$ denote the CL differentiation matrices on each of the two domains at the k-th time step. Then, using (B.9.14), we have here the following finite difference formula for the homogeneous PDEs prior to imposing BCs/JCs: $\frac{1}{\Delta t}(\boldsymbol{u}_{k+1}^{\pm} - \boldsymbol{u}_k^{\pm}) = (\mathbb{D}_k^{\pm})^2\boldsymbol{u}_k^{\pm} - \mathbb{C}_k^{\pm}\mathbb{D}\boldsymbol{u}_k^{\pm}$, where $\mathbb{D}$ is the CL differentiation matrix on $[-1,1]$ and $\mathbb{C}_k^{-} = \mathrm{diag}([2/(x_p(t_k))^2][(-x_i^{-})\dot{x}_p(t_k)])$, $\mathbb{C}_k^{+} = \mathrm{diag}([2/(1-x_p(t_k))^2][(x_i^{+}-1)\dot{x}_p(t_k)])$. Thus $\boldsymbol{u}_{k+1}^{\pm} = \boldsymbol{u}_k^{\pm} + \Delta t[(\mathbb{D}_k^{\pm})^2 - \mathbb{C}_k^{\pm}\mathbb{D}]\boldsymbol{u}_k^{\pm} = \boldsymbol{s}_k^{\pm}$. We can implement the BCs and JCs, by modifying the first and last equations on each domain:

$$\left[\begin{array}{cccc|cccc}
1 & 0 & \cdots & 0 & 0 & & & \\
0 & 1 & \cdots & 0 & & 0 & & \\
\vdots & \vdots & \ddots & \vdots & & & \ddots & \\
0 & 0 & \cdots & 1 & & & & 0 \\
\hline
0 & & & & & \mathbb{D}_k^{+}(1,:) & & \\
 & 0 & & & 0 & 1 & \cdots & 0 \\
 & & \ddots & & \vdots & \vdots & \ddots & \vdots \\
 & & & 0 & 0 & 0 & \cdots & 1
\end{array}\right]
\left[\begin{array}{c}
\boldsymbol{u}_{k+1}^{-} \\
\hline
\boldsymbol{u}_{k+1}^{+}
\end{array}\right]
=
\left[\begin{array}{c}
0 \\
\boldsymbol{s}_k^{-}(2:N) \\
u_{0,k}^{+} \\
\hline
\mathbb{D}_k^{-}(N,:)\boldsymbol{u}_k^{-} - \lambda \\
\boldsymbol{s}_k^{+}(2:N) \\
0
\end{array}\right]. \quad \text{(B.9.15)}$$

Note that we are actually introducing an error by using (for convenience and ease of adaptability) $\mathbb{D}_k^{+}$ instead of $\mathbb{D}_{k+1}^{+}$ on the LHS (in the equation for $u_{0,k+1}^{+}$). However, one can easily convince oneself that $\mathbb{D}_{k+1}^{+} - \mathbb{D}_k^{+} = \mathcal{O}(\Delta t)$, which is already the order of the error of the finite difference scheme, so we are not actually introducing any new error in this way. Furthermore, because we use up the last equation for $\boldsymbol{u}_k^{-}$ to impose the JC on u (*i.e.* we do not have an equation for $u_{N,k}^{-}$), we must use the derivative at the previous point (*i.e.*, at $u_{N-1,k}^{-}$) in order to impose the derivative JC. Hence on the RHS, we use $\mathbb{D}_k^{-}(N,:)$ instead of $\mathbb{D}_k^{-}(N+1,:)$.

The scheme for the finance model is analogous. We use again the first-order finite-difference method for the homogeneous equations, $\frac{1}{\Delta t}(\boldsymbol{f}_{k+1}^{\sigma} - \boldsymbol{f}_k^{\sigma}) = (\mathbb{D}_k^{\sigma})^2\boldsymbol{f}_k^{\sigma} - \mathbb{C}_k^{\sigma}\mathbb{D}\boldsymbol{f}_k^{\sigma}$ with the matrices $\mathbb{C}_k^{\sigma}$ defined similarly to those in the heat equation problem (again using (B.9.14)); thus $\boldsymbol{f}_{k+1}^{\sigma} = \boldsymbol{f}_k^{\sigma} + \Delta t[(\mathbb{D}_k^{\sigma})^2 - \mathbb{C}_k^{\sigma}\mathbb{D}]\boldsymbol{f}_k^{\sigma} = \boldsymbol{s}_k^{\sigma}$. To impose the BCs/JCs, we modify the equations appropriately:

$$\begin{bmatrix} & & \mathbb{D}_k^-(1,:) & & \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \\ \\ \boldsymbol{f}_{k+1}^- \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ s_{1,k}^- \\ \vdots \\ s_{N,k}^- \\ f_{0,k}^0 \end{bmatrix}, \tag{B.9.16}$$

$$\begin{bmatrix} & & \mathbb{D}_k^0(1,:) & & \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \\ \\ \boldsymbol{f}_{k+1}^0 \\ \\ \\ \end{bmatrix} = \begin{bmatrix} \mathbb{D}_k^-(N,:)\boldsymbol{f}_k^- - \lambda_k \\ s_{1,k}^0 \\ \vdots \\ s_{N,k}^0 \\ f_{0,k}^+ \end{bmatrix}, \tag{B.9.17}$$

$$\begin{bmatrix} & & \mathbb{D}_k^+(1,:) & & \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ & & \mathbb{D}_k^+(N+1,:) & & \end{bmatrix} \begin{bmatrix} \\ \\ \boldsymbol{f}_{k+1}^+ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} \mathbb{D}_k^0(N,:)\boldsymbol{f}_k^0 + \lambda_k \\ s_{1,k}^+ \\ \vdots \\ s_{N,k}^+ \\ 0 \end{bmatrix}. \tag{B.9.18}$$

B.9.4. Second-order hyperbolic PDEs. We again apply a first order in time finite difference scheme to the homogeneous PDEs; prior to imposing BCs/JCs, the equations become

$$\frac{1}{\Delta t}\left(\begin{bmatrix} \boldsymbol{u}_{k+1}^\pm \\ \boldsymbol{v}_{k+1}^\pm \\ \boldsymbol{w}_{k+1}^\pm \end{bmatrix} - \begin{bmatrix} \boldsymbol{u}_k^\pm \\ \boldsymbol{v}_k^\pm \\ \boldsymbol{w}_k^\pm \end{bmatrix} \right) = \mathbb{C}^\pm \begin{bmatrix} \boldsymbol{u}_k^\pm \\ \boldsymbol{v}_k^\pm \\ \boldsymbol{w}_k^\pm \end{bmatrix}, \tag{B.9.19}$$

where

$$\mathbb{C}^\pm = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathbb{I} \\ 0 & \mathbb{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbb{D}^\pm & 0 & 0 \\ 0 & \mathbb{D}^\pm & 0 \\ 0 & 0 & \mathbb{D}^\pm \end{bmatrix} + \begin{bmatrix} 0 & 0 & \mathbb{I} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \mathbb{I} \\ 0 & 0 & \mathbb{D}^\pm \\ 0 & \mathbb{D}^\pm & 0 \end{bmatrix}. \tag{B.9.20}$$

We can rewrite the discretized PDE as

$$\begin{bmatrix} \boldsymbol{u}_{k+1}^{\pm} \\ \boldsymbol{v}_{k+1}^{\pm} \\ \boldsymbol{w}_{k+1}^{\pm} \end{bmatrix} = \left(\Delta t \mathbb{C}^{\pm} + \mathbb{I}\right) \begin{bmatrix} \boldsymbol{u}_{k}^{\pm} \\ \boldsymbol{v}_{k}^{\pm} \\ \boldsymbol{w}_{k}^{\pm} \end{bmatrix} = \begin{bmatrix} \boldsymbol{s}_{k}^{\pm} \\ \boldsymbol{y}_{k}^{\pm} \\ \boldsymbol{z}_{k}^{\pm} \end{bmatrix}. \tag{B.9.21}$$

To impose the BCs and JCs, we modify the equations as follows:

$$\begin{bmatrix} 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 1 & 0 \\ & \mathbb{D}^{-}(N+1,:) & & \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{u}_{k+1}^{-} \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{s}_{k}^{-}(2:N) \\ \mathbb{D}^{+}(1,:)\boldsymbol{u}_{k}^{+} \end{bmatrix}, \tag{B.9.22}$$

$$\begin{bmatrix} \\ \boldsymbol{u}_{k+1}^{+} \\ \\ \end{bmatrix} = \begin{bmatrix} u_{N-1,k+1}^{-} - g_k \\ \boldsymbol{s}_{k}^{+}(2:N) \\ 0 \end{bmatrix}, \tag{B.9.23}$$

$$\begin{bmatrix} \boldsymbol{v}_{k+1}^{-} \\ \hline \boldsymbol{v}_{k+1}^{+} \end{bmatrix} = \begin{bmatrix} \mathbb{D}^{-}\boldsymbol{u}_{k+1}^{-} \\ \hline \mathbb{D}^{+}\boldsymbol{u}_{k+1}^{+} \end{bmatrix}, \tag{B.9.24}$$

$$\begin{bmatrix} \boldsymbol{w}_{k+1}^{-} \\ \hline \boldsymbol{w}_{k+1}^{+} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{z}_{k}^{-}(2:N+1) \\ \hline w_{N,k+1}^{-} - \dot{g}_{k+1} \\ \boldsymbol{z}_{k}^{+}(2:N) \\ 0 \end{bmatrix}, \tag{B.9.25}$$

B.9.5. Elliptic PDEs. In this case we have no time evolution, and we simply need to solve $((\mathbb{D}^{\pm})^2 + \mathrm{diag}(1/X_i^{\pm})\mathbb{D}^{\pm})\boldsymbol{u}^{\pm} = \mathbb{M}^{\pm}\boldsymbol{u}^{\pm} = \boldsymbol{0}$, modified appropriately to account for the BCs and JCs. In particular, we first solve for $\boldsymbol{u}^{+}$ using the BCs, and then for $\boldsymbol{u}^{-}$

using the solution for $\boldsymbol{u}^+$ to implement the JCs:

$$\begin{bmatrix} \mathbb{M}^+(1:N-1,:) \\ 0\ 0\cdots 0\ 1 \\ \mathbb{D}^+(N+1,:) \end{bmatrix} \boldsymbol{u}^+ = \begin{bmatrix} \mathbf{0}(1:N-1) \\ 1-\frac{1}{2}\log(2L) \\ -\frac{1}{2L} \end{bmatrix}, \tag{B.9.26}$$

$$\begin{bmatrix} \mathbb{M}^-(1:N-1,:) \\ 0\ 0\cdots 0\ 1 \\ \mathbb{D}^-(N+1,:) \end{bmatrix} \boldsymbol{u}^- = \begin{bmatrix} \mathbf{0}(1:N-1) \\ u_0^+ \\ \mathbb{D}^+(1,:)\boldsymbol{u}^+ + 1 \end{bmatrix}. \tag{B.9.27}$$

Bibliography

[1] L. V. Abad, "Principles of classical statistical mechanics: A perspective from the notion of complementarity", Annals of Physics **327**, 1682 (2012).

[2] E. Abbena, S. Salamon, and A. Gray, *Modern Differential Geometry of Curves and Surfaces with Mathematica*, 3rd edition (Chapman and Hall/CRC, Boca Raton, 2006).

[3] B. P. Abbott et al., "GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs", arXiv:1811.12907 [astro-ph, physics:gr-qc] (2018).

[4] R. Abraham, J. E. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis, and Applications*, 3rd edition (Springer-Verlag, New York, 2001).

[5] Y. Achdou, F. J. Buera, J.-M. Lasry, P.-L. Lions, and B. Moll, "Partial differential equation models in macroeconomics", Philosophical Transactions of the Royal Society A **372**, 20130397 (2014).

[6] A. D. Aczel, *God's Equation: Einstein, Relativity, and the Expanding Universe* (Delta, New York, 2000).

[7] K. Aki and P. G. Richards, *Quantitative Seismology*, 2nd edition (University Science Books, Sausalito, 2009).

[8] M. Alcubierre, *Introduction to 3+1 Numerical Relativity* (Oxford University Press, Oxford, 2012).

[9] M. Alcubierre, G. Allen, B. Brügmann, E. Seidel, and W.-M. Suen, "Towards an understanding of the stability properties of the 3+1 evolution equations in general relativity", Physical Review D **62**, 124011 (2000).

[10] P. Amaro-Seoane et al., "Laser Interferometer Space Antenna", arXiv:1702.00786 [astro-ph] (2017).

[11] P. Amaro-Seoane et al., "The Gravitational Universe", arXiv:1305.5720 [astro-ph.CO] (2013).

[12] P. Amaro-Seoane, "Relativistic dynamics and extreme mass ratio inspirals", Living Reviews in Relativity **21**, 4 (2018).

[13] P. Amaro-Seoane, "X-MRIs: Extremely Large Mass-Ratio Inspirals", arXiv:1903.10871 [astro-ph, physics:gr-qc] (2019).

[14] J. L. Anderson and P. G. Bergmann, "Constraints in Covariant Field Theories", Physical Review **83**, 1018 (1951).

[15]L. Andersson, T. Bäckdahl, and P. Blue, "Geometry of black hole spacetimes", arXiv:1610.03540 [gr-qc] (2016).

[16]S. Aoudia and A. D. A. M. Spallicci, "Source-free integration method for black hole perturbations and self-force computation: Radial fall", Physical Review D **83**, 064029 (2011).

[17]M. Armano et al., "Beyond the Required LISA Free-Fall Performance: New LISA Pathfinder Results down to 20 muHz", Physical Review Letters **120**, 061101 (2018).

[18]M. Armano et al., "Sub-Femto-g Free Fall for Space-Based Gravitational Wave Observatories: LISA Pathfinder Results", Physical Review Letters **116**, 231101 (2016).

[19]V. I. Arnold, *Mathematical Methods of Classical Mechanics*, trans. by A. Weinstein and K. Vogtmann, 2nd edition (Springer, New York, 1997).

[20]R. Arnowitt, S. Deser, and C. W. Misner, "Dynamical Structure and Definition of Energy in General Relativity", Physical Review **116**, 1322 (1959).

[21]R. Arnowitt, S. Deser, and C. W. Misner, "The dynamics of general relativity", in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), p. 227.

[22]R. Arnowitt, S. Deser, and C. W. Misner, "Republication of: The dynamics of general relativity", General Relativity and Gravitation **40**, 1997 (2008).

[23]A. Ashtekar, "New Hamiltonian formulation of general relativity", Physical Review D **36**, 1587–1602 (1987).

[24]A. Ashtekar, M. Reuter, and C. Rovelli, "From General Relativity to Quantum Gravity", in *General Relativity and Gravitation: A Centennial Perspective*, edited by A. Ashtekar, B. K. Berger, J. Isenberg, and M. MacCallum (Cambridge University Press, Cambridge, 2015), p. 553.

[25]S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, "Science with the space-based interferometer LISA. V. Extreme mass-ratio inspirals", Physical Review D **95**, 103012 (2017).

[26]J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, "Gravitational-Wave Extraction from an Inspiraling Configuration of Merging Black Holes", Physical Review Letters **96**, 111102 (2006).

[27]L. Barack and D. A. Golbourn, "Scalar-field perturbations from a particle orbiting a black hole using numerical evolution in 2+1 dimensions", Physical Review D **76**, 044020 (2007).

[28]L. Barack, D. A. Golbourn, and N. Sago, "M-mode regularization scheme for the self-force in Kerr spacetime", Physical Review D **76**, 124036 (2007).

[29]L. Barack and C. O. Lousto, "Perturbations of Schwarzschild black holes in the Lorenz gauge: Formulation and numerical implementation", Physical Review D **72**, 104026 (2005).

[30]L. Barack, Y. Mino, H. Nakano, A. Ori, and M. Sasaki, "Calculating the gravitational self force in Schwarzschild spacetime", Physical Review Letters **88**, 091101 (2002).

[31]L. Barack and A. Ori, "Mode sum regularization approach for the self force in black hole spacetime", Physical Review D **61**, 061502 (2000).

[32]L. Barack and A. Ori, "Regularization parameters for the self force in Schwarzschild spacetime. I: Scalar case", Physical Review D **66**, 084022 (2002).

[33]L. Barack, A. Ori, and N. Sago, "Frequency-domain calculation of the self force: The High-frequency problem and its resolution", Physical Review D **78**, 084021 (2008).

[34]L. Barack and A. Pound, "Self-force and radiation reaction in general relativity", Reports on Progress in Physics **82**, 016904 (2018).

[35]J. F. Barbero G., "Real Ashtekar variables for Lorentzian signature space-times", Physical Review D **51**, 5507 (1995).

[36]J. Barbour, "Shape Dynamics. An Introduction", in *Quantum Field Theory and Gravity: Conceptual and Mathematical Advances in the Search for a Unified Framework*, edited by F. Finster, O. Müller, M. Nardmann, J. Tolksdorf, and E. Zeidler (Springer, Basel, 2012), p. 257.

[37]J. Barbour, T. Koslowski, and F. Mercati, "A Gravitational Origin of the Arrows of Time", arXiv:1310.5167 [gr-qc] (2013).

[38]J. Barbour, T. Koslowski, and F. Mercati, "Identification of a Gravitational Arrow of Time", Physical Review Letters **113**, 181101 (2014).

[39]J. M. Bardeen, "Gauge-invariant cosmological perturbations", Physical Review D **22**, 1882–1905 (1980).

[40]A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Dover Publications, New York, 1980).

[41]T. W. Baumgarte and S. L. Shapiro, "Numerical integration of Einstein's field equations", Physical Review D **59**, 024007 (1998).

[42]T. W. Baumgarte and S. L. Shapiro, *Numerical Relativity: Solving Einstein's Equations on the Computer* (Cambridge University Press, Cambridge, 2010).

[43]J. D. Bekenstein, "Black Holes and Entropy", Physical Review D **7**, 2333 (1973).

[44]J. S. Bell, "How to Teach Special Relativity", Progress in Scientific Culture **1**, 1 (1976).

[45]V. Benci, "Ultrafunctions and Generalized Solutions", Advanced Nonlinear Studies **13**, 461 (2013).

[46]A. Bernstein, A. Chertock, and A. Kurganov, "Central-upwind scheme for shallow water equations with discontinuous bottom topography", Bulletin of the Brazilian Mathematical Society, New Series **47**, 91 (2016).

[47]C. P. L. Berry, S. A. Hughes, C. F. Sopuerta, A. J. K. Chua, A. Heffernan, K. Holley-Bockelmann, D. P. Mihaylov, M. C. Miller, and A. Sesana, "The unique potential of extreme mass-ratio inspirals for gravitational-wave astronomy", arXiv:1903.03686 [astro-ph, physics:gr-qc] (2019).

[48]L. Blanchet, A. Spallicci, and B. Whiting, eds., *Mass and Motion in General Relativity*, Fundamental Theories of Physics 162 (Springer, New York, 2011).

[49]M. Bojowald, *Canonical Gravity and Applications: Cosmology, Black Holes, and Quantum Gravity* (Cambridge University Press, New York, 2011).

[50]L. Boltzmann, "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen", Sitzungsberichte der Akademie der Wissenschaften, Wien, II **66**, 275 (1872).

[51]E. Bottazzi, "Grid functions of nonstandard analysis in the theory of distributions and in partial differential equations", arXiv:1704.00470 [math] (2017).

[52]J. P. Boyd, *Chebyshev and Fourier Spectral Methods*, 2nd edition (Dover Publications, Mineola, 2001).

[53]R. H. Brandenberger, "Lectures on the Theory of Cosmological Perturbations", in *The Early Universe and Observational Cosmology*, edited by N. Bretón, J. L. Cervantes-Cota, and M. Salgado, Lecture Notes in Physics (Springer, Berlin, 2004), p. 127.

[54]H. R. Brown, W. Myrvold, and J. Uffink, "Boltzmann's H-theorem, its discontents, and the birth of statistical mechanics", Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **40**, 174 (2009).

[55]J. D. Brown, S. R. Lau, and J. W. York, "Action and Energy of the Gravitational Field", Annals of Physics **297**, 175 (2002).

[56]J. D. Brown and J. W. York, "Quasilocal energy and conserved charges derived from the gravitational action", Physical Review D **47**, 1407 (1993).

[57]M. Bruni, S. Matarrese, S. Mollerach, and S. Sonego, "Perturbations of spacetime: gauge transformations and gauge invariance at second order and beyond", Classical and Quantum Gravity **14**, 2585 (1997).

[58]M. Burger, L. Caffarelli, P. A. Markowich, and M.-T. Wolfram, "On a Boltzmann-type price formation model", Proceedings of the Royal Society A **469**, 20130126 (2013).

[59]M. J. Cáceres, J. A. Carrillo, and B. Perthame, "Analysis of nonlinear noisy integrate & fire neuron models: blow-up and steady states", The Journal of Mathematical Neuroscience **1**, 7 (2011).

[60]M. J. Cáceres and R. Schneider, "Blow-up, steady states and long time behaviour of excitatory-inhibitory nonlinear neuron models", Kinetic and Related Models **10**, 587 (2016).

[61]L. A. Caffarelli, P. A. Markowich, and J.-F. Pietschmann, "On a price formation free boundary model by Lasry and Lions", Comptes Rendus Mathematique **349**, 621 (2011).

[62]M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, "Accurate Evolutions of Orbiting Black-Hole Binaries without Excision", Physical Review Letters **96**, 111101 (2006).

[63]P. Canizares, "Extreme-Mass-Ratio Inspirals", PhD thesis (Universitat Autònoma de Barcelona, 2011).

[64]P. Canizares and C. F. Sopuerta, "An Efficient Pseudospectral Method for the Computation of the Self-force on a Charged Particle: Circular Geodesics around a Schwarzschild Black Hole", Physical Review D **79**, 084020 (2009).

[65]P. Canizares and C. F. Sopuerta, "Efficient pseudospectral method for the computation of the self-force on a charged particle: Circular geodesics around a Schwarzschild black hole", Physical Review D **79**, 084020 (2009).

[66]P. Canizares and C. F. Sopuerta, "Overcoming the Gauge Problem for the Gravitational Self-Force", arXiv:1406.7154 [gr-qc] (2014).

[67]P. Canizares and C. F. Sopuerta, "Simulations of Extreme-Mass-Ratio Inspirals Using Pseudospectral Methods", Journal of Physics: Conference Series **154**, 012053 (2009).

[68]P. Canizares and C. F. Sopuerta, "Time-domain modelling of Extreme-Mass-Ratio Inspirals for the Laser Interferometer Space Antenna", Journal of Physics: Conference Series **314**, 012075 (2011).

[69]P. Canizares and C. F. Sopuerta, "Tuning time-domain pseudospectral computations of the self-force on a charged scalar particle", Classical and Quantum Gravity **28**, 134011 (2011).

[70]P. Canizares, C. F. Sopuerta, and J. L. Jaramillo, "Pseudospectral Collocation Methods for the Computation of the Self-Force on a Charged Particle: Generic Orbits around a Schwarzschild Black Hole", Physical Review D **82**, 044023 (2010).

[71]P. Canizares, C. F. Sopuerta, and J. L. Jaramillo, "Pseudospectral collocation methods for the computation of the self-force on a charged particle: Generic orbits around a Schwarzschild black hole", Physical Review D **82**, 044023 (2010).

[72]V. Cardoso, L. Gualtieri, C. Herdeiro, U. Sperhake, P. M. Chesler, L. Lehner, S. C. Park, H. S. Reall, C. F. Sopuerta, D. Alic, O. J. C. Dias, R. Emparan, V. Ferrari, S. B. Giddings, M. Godazgar, R. Gregory, V. E. Hubeny, A. Ishibashi, G. Landsberg, C. O. Lousto, D. Mateos, V. Moeller, H. Okawa, P. Pani, M. A. Parker, F. Pretorius, M. Shibata, H. Sotani, T. Wiseman, H. Witek, N. Yunes, and M. Zilhão, "NR/HEP: roadmap for the future", Classical and Quantum Gravity **29**, 244001 (2012).

[73]S. M. Carroll, *Spacetime and Geometry: An Introduction to General Relativity* (Pearson, San Francisco, 2003).

[74]S. M. Carroll and G. N. Remmen, "What is the Entropy in Entropic Gravity?", Physical Review D **93**, 124052 (2016).

[75]É. Cartan, *Les systèmes différentiels extérieurs et leurs applications géométriques* (Hermann et Cie, Paris, 1945).

[76]É. Cartan, "Sur certaines expressions différentielles et le problème de Pfaff", Annales scientifiques de l'École Normale Supérieure **16**, 239 (1899).

[77]A. R. R. Casti, A. Omurtag, A. Sornborger, E. Kaplan, B. Knight, J. Victor, and L. Sirovich, "A population study of integrate-and-fire-or-burst neurons", Neural Computation **14**, 957 (2002).

[78]M. Celoria, R. Oliveri, A. Sesana, and M. Mapelli, "Lecture notes on black hole binary astrophysics", arXiv:1807.11489 [astro-ph] (2018).

[79]S. Chandrasekhar, *Newton's Principia for the Common Reader* (Clarendon Press, Oxford, 2003).

[80]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford University Press, New York, 1998).

[81]S. Chandrasekhar and S. Detweiler, "The quasi-normal modes of the Schwarzschild black hole", Royal Society of London Proceedings Series A **344**, 441 (1975).

[82]Y. Choquet-Bruhat, *General Relativity and the Einstein Equations* (Oxford University Press, Oxford, 2009).

[83]Y. Choquet-Bruhat and R. Geroch, "Global aspects of the Cauchy problem in general relativity", Communications in Mathematical Physics **14**, 329–335 (1969).

[84]D. Christodoulou and S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space* (Princeton University Press, Princeton, 1993).

[85]D. Codazzi, "Sulle coordinate curvilinee d'una superficie dello spazio", Annali di Matematica Pura ed Applicata **2**, 101 (1868).

[86]A. A. Coley, "Mathematical General Relativity", arXiv:1807.08628 [gr-qc] (2018).

[87]A. A. Coley, "Open problems in mathematical physics", Physica Scripta **92**, 093003 (2017).

[88]J. F. Colombeau, "Nonlinear Generalized Functions: their origin, some developments and recent advances", The São Paulo Journal of Mathematical Sciences **7**, 201 (2013).

[89]M. Colpi and A. Sesana, "Gravitational Wave Sources in the Era of Multi-Band Gravitational Wave Astronomy", in *An Overview of Gravitational Waves: Theory, Sources and Detection*, edited by A. Gerard and E. Plagnol (World Scientific, 2016), p. 43.

[90]G. Constantine and T. Savits, "A Multivariate Faa di Bruno Formula with Applications", Transactions of the American Mathematical Society **348**, 503 (1996).

[91]S. F. Cortizo, "On Dirac's Delta Calculus", arXiv:funct-an/9510004 (1995).

[92]J. Corvino and R. M. Schoen, "On the Asymptotics for the Vacuum Einstein Constraint Equations", Journal of Differential Geometry **73**, 185 (2006).

[93]Č. Crnković, "Symplectic geometry and (super-)Poincaré algebra in geometrical theories", Nuclear Physics B **288**, 419 (1987).

[94]Č. Crnković and E. Witten, "Covariant description of canonical formalism in geometrical theories", in *Three Hundred Years of Gravitation*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1989), p. 676.

[95]C. T. Cunningham, R. H. Price, and V. Moncrief, "Radiation from collapsing relativistic stars. I - Linearized odd-parity radiation", The Astrophysical Journal **224**, 643 (1978).

[96]P. C. W. Davies, *The Physics of Time Asymmetry* (University of California Press, Berkeley, 1977).

[97]S. Detweiler and B. F. Whiting, "Self-force via a Green's function decomposition", Physical Review D **67**, 024025 (2003).

[98]E. Dewan and M. Beran, "Note on Stress Effects due to Relativistic Contraction", American Journal of Physics **27**, 517 (1959).

[99]B. S. DeWitt and R. W. Brehme, "Radiation damping in a gravitational field", Annals of Physics **9**, 220 (1960).

[100]L. M. Diaz-Rivera, E. Messaritaki, B. F. Whiting, and S. Detweiler, "Scalar field self-force effects on orbits about a Schwarzschild black hole", Physical Review D **70**, 124018 (2004).

[101]P. A. M. Dirac, "Classical Theory of Radiating Electrons", Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **167**, 148 (1938).

[102]P. A. M. Dirac, "Generalized Hamiltonian dynamics", Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **246**, 326 (1958).

[103]P. A. M. Dirac, "The theory of gravitation in Hamiltonian form", Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **246**, 333 (1958).

[104]R. M. Dreizler and E. Engel, *Density Functional Theory* (Springer, Berlin, 2011).

[105]J. Droste, "Het veld van een enkel centrum in Einstein's theorie der zwaartekracht, en de beweging van een stoffelijk punt in dat veld", Koninklijke Akademie van Wetenschappen Amsterdam **25**, 163 (1916).

[106]J. Droste, "Het zwaartekrachtsveld van een of meer lichamen volgens de theorie van Einstein", Doctorate thesis (Rijksuniversiteit Leiden, 1916).

[107]M. D. Duez and Y. Zlochower, "Numerical relativity of compact binaries in the 21st century", Reports on Progress in Physics **82**, 016902 (2018).

[108]J. Ehlers and R. Geroch, "Equation of motion of small bodies in relativity", Annals of Physics **309**, 232 (2004).

[109]A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie", Annalen der Physik **354**, 769 (1916).

[110]A. Einstein, "Zur Theorie des statischen Gravitationsfeldes", Annalen der Physik **343**, 443 (1912).

[111]A. Einstein and N. Rosen, "The Particle Problem in the General Theory of Relativity", Physical Review **48**, 73 (1935).

[112]A. Einstein, "Autobiographical Notes", in *Albert Einstein, Philosopher-Scientist*, Vol. VII, edited by P. A. Schilpp, The Library of Living Philosophers (Open Court, La Salle, 1949).

[113]A. Einstein, "Die Feldgleichungen der Gravitation", Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 844 (1915).

[114]A. Einstein, *Four Lectures on the Theory of Relativity, Held at Princeton University in May 1921*, trans. by E. P. Adams (Methuen, London, 1922).

[115]A. Einstein, "Four Lectures on the Theory of Relativity, Held at Princeton University in May 1921", in *The Collected Papers of Albert Einstein, Volume 7: The Berlin Years: Writings, 1918-1921 (English translation supplement)*, edited by M. Janssen, R. Schulmann, J. Illy, C. Lehner, and D. Kormos Buchwald (Princeton University Press, Princeton, 2002), p. 261.

[116]A. Einstein, *Geometrie Und Erfahrung* (Erweiterte Fassung des Festvortrages Gehalten an der Preussischen Akademie der Wissenschaften, Berlin, 1921).

[117]A. Einstein, “The Field Equations of Gravitation”, in *The Collected Papers of Albert Einstein, Volume 6: The Berlin Years: Writings, 1914-1917 (English translation supplement)*, edited by A. J. Kox, M. J. Klein, and R. Schulmann (Princeton University Press, Princeton, 1996), p. 117.

[118]A. Einstein, “The Foundation of the General Theory of Relativity”, in *The Collected Papers of Albert Einstein, Volume 6: The Berlin Years: Writings, 1914-1917 (English translation supplement)*, edited by A. J. Kox, M. J. Klein, and R. Schulmann (Princeton University Press, Princeton, 1996), p. 146.

[119]A. Einstein, *The Meaning of Relativity* (Routledge, London, 2003).

[120]A. Einstein, “What Is the Theory of Relativity?”, in *Ideas and Opinions*, edited by C. Seelig (Crown Publishers, Inc., New York, 1954).

[121]A. Einstein, “Zum gegenwärtigen Stande des Gravitationsproblems”, Physikalische Zeitschrift **14**, 1249 (1913).

[122]A. Einstein and L. Infeld, *The Evolution of Physics* (Cambridge University Press, London, 1938).

[123]A. Einstein, L. Infeld, and B. Hoffmann, “The Gravitational Equations and the Problem of Motion”, Annals of Mathematics **39**, 65 (1938).

[124]R. J. Epp, R. B. Mann, and P. L. McGrath, “On the Existence and Utility of Rigid Quasilocal Frames”, in *Classical and Quantum Gravity: Theory, Analysis and Applications*, edited by V. R. Frignanni, Physics Research and Technology (Nova Science Publishers, New York, 2012).

[125]R. J. Epp, R. B. Mann, and P. L. McGrath, “Rigid motion revisited: rigid quasilocal frames”, Classical and Quantum Gravity **26**, 035015 (2009).

[126]R. J. Epp, P. L. McGrath, and R. B. Mann, “Momentum in general relativity: local versus quasilocal conservation laws”, Classical and Quantum Gravity **30**, 195019 (2013).

[127]L. C. Evans, *Partial Differential Equations* (American Mathematical Society, Providence, 1998).

[128]B. Felsager, *Geometry, Particles, and Fields* (Springer, New York, 2012).

[129]S. E. Field, J. S. Hesthaven, and S. R. Lau, “Discontinuous Galerkin method for computing gravitational waveforms from extreme mass ratio binaries”, Classical and Quantum Gravity **26**, 165010 (2009).

[130]Y. Fourès-Bruhat, “Théorème d’existence pour certains systèmes d’équations aux dérivées partielles non linéaires”, Acta Mathematica **88**, 141 (1952).

[131]L. Freidel, “Gravitational energy, local holography and non-equilibrium thermodynamics”, Classical and Quantum Gravity **32**, 055005 (2015).

[132]L. Freidel and Y. Yokokura, “Non-equilibrium thermodynamics of gravitational screens”, Classical and Quantum Gravity **32**, arXiv: 1405.4881, 215002 (2015).

[133]V. Frolov and I. Novikov, *Black Hole Physics: Basic Concepts and New Developments* (Springer, Dordrecht, 1998).

[134]P. Galison, M. Gordin, and D. Kaiser, eds., *The Roots of Special Relativity: Science and Society* (Routledge, New York, 2001).

[135]C. F. Gauss, "Disquisitiones Generales circa Superficies Curvas", Comm. Soc. Reg. Sc. Gott. Rec. **6**, 99 (1827).

[136]R. Geroch and P. S. Jang, "Motion of a body in general relativity", Journal of Mathematical Physics **16**, 65 (1975).

[137]R. Geroch and J. Traschen, "Strings and other distributional sources in general relativity", Physical Review D **36**, 1017 (1987).

[138]G. W. Gibbons and S. W. Hawking, "Action integrals and partition functions in quantum gravity", Physical Review D **15**, 2752 (1977).

[139]J. W. Gibbs, "Letter to the Editor", Nature **59**, 606 (1899).

[140]D. Giulini, "Dynamical and Hamiltonian Formulation of General Relativity", in *Springer Handbook of Spacetime*, edited by A. Ashtekar and V. Petkov (Springer, Dordrecht, 2014), p. 323.

[141]J. Gleick, *Isaac Newton* (Vintage Books, New York, 2004).

[142]H. Goenner, J. Renn, J. Ritter, and T. Sauer, eds., *The Expanding Worlds of General Relativity*, Einstein Studies (Birkhäuser, Basel, 1999).

[143]S. E. Gralla, "Gauge and averaging in gravitational self-force", Physical Review D **84**, 084050 (2011).

[144]S. E. Gralla, "Second-order gravitational self-force", Physical Review D **85**, 124011 (2012).

[145]S. E. Gralla, A. I. Harte, and R. M. Wald, "Rigorous derivation of electromagnetic self-force", Physical Review D **80**, 024031 (2009).

[146]S. E. Gralla and R. M. Wald, "A note on the coordinate freedom in describing the motion of particles in general relativity", Classical and Quantum Gravity **28**, 177001 (2011).

[147]S. E. Gralla and R. M. Wald, "A rigorous derivation of gravitational self-force", Classical and Quantum Gravity **25**, 205009 (2008).

[148]P. Grandclément and J. Novak, "Spectral Methods for Numerical Relativity", Living Reviews in Relativity **12**, 1 (2009).

[149]R. Haas and E. Poisson, "Mode-sum regularization of the scalar self-force: Formulation in terms of a tetrad decomposition of the singular field", Physical Review D **74**, 044009 (2006).

[150]H. Hankel, *Zur allgemeinen Theorie der Bewegung der Flüssigkeiten. Eine von der philosophischen Facultät der Georgia Augusta am 4.Juni 1861 gekrönte Preisschrift* (Dieterische Univ.-Buchdruckerei (W.F. Kaestner), Göttingen, 1861).

[151]A. I. Harte, "Electromagnetic self-forces and generalized Killing fields", Classical and Quantum Gravity **26**, 155015 (2009).

[152]A. I. Harte, “Self-forces from generalized Killing fields”, Classical and Quantum Gravity **25**, 235020 (2008).

[153]E. Haskell, D. Q. Nykamp, and D. Tranchina, “Population density methods for large-scale modelling of neuronal networks with realistic synaptic kinetics: cutting the dimension down to size”, Network (Bristol, England) **12**, 141 (2001).

[154]S. W. Hawking, “Particle creation by black holes”, Communications in Mathematical Physics **43**, 199 (1975).

[155]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1975).

[156]D. Hilbert, “Die Grundlagen der Physik”, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen – Mathematisch-Physikalische Klasse **3**, 395 (1915).

[157]D. Hilditch, “An introduction to well-posedness and free-evolution”, International Journal of Modern Physics A **28**, 1340015 (2013).

[158]H. Hofer and E. Zehnder, *Symplectic Invariants and Hamiltonian Dynamics* (Springer Basel, Basel, 2011).

[159]S. Holst, “Barbero’s Hamiltonian derived from a generalized Hilbert-Palatini action”, Physical Review D **53**, 5966 (1996).

[160]G. Horwitz, “Steepest descent path for the microcanonical ensemble-resolution of an ambiguity”, Communications in Mathematical Physics **89**, 117 (1983).

[161]G. Immirzi, “Real and complex connections for canonical gravity”, Classical and Quantum Gravity **14**, L177 (1997).

[162]J. L. Jaramillo and E. Gourgoulhon, “Mass and Angular Momentum in General Relativity”, in *Mass and Motion in General Relativity*, edited by L. Blanchet, A. Spallicci, and B. Whiting, Fundamental Theories of Physics (Springer, Dordrecht, 2011), p. 87.

[163]J. L. Jaramillo, C. F. Sopuerta, and P. Canizares, “Are time-domain self-force calculations contaminated by Jost solutions?”, Physical Review D **83**, 061503 (2011).

[164]J. Jezierski, “Energy and Angular Momentum ofthe Weak Gravitational Waves on the Schwarzschild Background — Quasilocal Gauge-invariant Formulation”, General Relativity and Gravitation **31**, 1855 (1999).

[165]J.-H. Jung, “A Note on the Spectral Collocation Approximation of Some Differential Equations with Singular Source Terms”, Journal of Scientific Computing **39**, 49 (2009).

[166]J.-H. Jung and W. S. Don, “Collocation Methods for Hyperbolic Partial Differential Equations with Singular Sources”, Advances in Applied Mathematics and Mechanics, 769 (2009).

[167]M. Kaltenbacher, ed., *Computational Acoustics* (Springer, New York, 2017).

[168]V. J. Katz, “The History of Stokes’ Theorem”, Mathematics Magazine **52**, 146 (1979).

[169]J. Kepler, *Astronomia Nova aitiologetos [romanized] : seu physica coelestis, tradita commentariis de motibus stellae Martis ex observationibus G.V. Tychonis Brahe* (Voegelin, Heidelberg, 1609).

[170] J. Kepler, *Prodromus dissertationum cosmographicarum, continens mysterium cosmographicum, de admirabili proportione orbium coelestium, deque causis coelorum numeri, magnitudinis, motuumque periodicorum genuinis & propriis, demonstratum, per quinque regularia corpora geometrica* (Excudebat Georgius Gruppenbachius, Tuebingen, 1596).

[171] R. P. Kerr, "Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics", Physical Review Letters **11**, 237 (1963).

[172] D.-H. Kim, "Radiation reaction in curved spacetime", PhD thesis (University of Florida, 2005).

[173] D.-H. Kim, "Regularization parameters for the self-force of a scalar particle in a general orbit about a Schwarzschild black hole", arXiv:gr-qc/0402014 (2004).

[174] W. Kinnersley, "Type D Vacuum Metrics", Journal of Mathematical Physics **10**, 1195 (1969).

[175] A. Koestler, *The Sleepwalkers* (Hutchinson, London, 1959).

[176] C. Lanczos, *The Variational Principles of Mechanics* (University of Toronto Press, Toronto, 1949).

[177] J. Laskar, "Is the Solar System Stable?", in *Chaos: Poincaré Seminar 2010*, edited by B. Duplantier, S. Nonnenmacher, and V. Rivasseau, Progress in Mathematical Physics (Springer Basel, Basel, 2013), p. 239.

[178] J.-M. Lasry and P.-L. Lions, "Mean field games", Japanese Journal of Mathematics **2**, 229 (2007).

[179] J. M. Lee, *Introduction to Smooth Manifolds* (Springer, New York, 2002).

[180] D. Lehmkuhl, "Why Einstein did not believe that general relativity geometrizes gravity", Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **46**, 316 (2014).

[181] L. Lehner and F. Pretorius, "Numerical Relativity and Astrophysics", Annual Review of Astronomy and Astrophysics **52**, 661 (2014).

[182] J. Leray, *Hyperbolic Differential Equations* (Institute for Advanced Studies, Princeton, 1952).

[183] C. K. Li, "A review on the products of distributions", in *Mathematical Methods in Engineering*, edited by K. Taş, J. A. Tenreiro Machado, and D. Baleanu (Springer, Dordrecht, 2007), p. 71.

[184] R. López-Alemán, G. Khanna, and J. Pullin, "Perturbative evolution of particle orbits around Kerr black holes: time-domain calculation", Classical and Quantum Gravity **20**, 3259 (2003).

[185] J. J. Loschmidt, "Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft", Sitz. Akad. Wiss. Wien **73**, 128 (1876).

[186] Luis Barreira, "Poincaré recurrence: old and new", in *XIVth International Congress on Mathematical Physics*, edited by J.-C. Zambrini (World Scientific, 2006), pp. 415–422.

[187]R. Madariaga, “Seismic Source Theory”, in *Earthquake Seismology*, edited by H. Kanamori, Treatise on Geophysics: (Elsevier, 2007), p. 59.

[188]M. Maggiore, *Gravitational Waves: Volume 1: Theory and Experiments* (Oxford University Press, Oxford, 2007).

[189]G. Mainardi, “Sulle coordinate curvilinee d'una superficie dello spazio”, Giornale del R. Istituto Lombardo **9**, 385 (1856).

[190]P. A. Markowich, N. Matevosyan, J.-F. Pietschmann, and M.-T. Wolfram, “On a parabolic free boundary equation modeling price formation”, Mathematical Models and Methods in Applied Sciences **19**, 1929 (2009).

[191]K. Martel and E. Poisson, “Gravitational perturbations of the Schwarzschild spacetime: A practical covariant and gauge-invariant formalism”, Physical Review D **71**, 104003 (2005).

[192]P. McGrath, “Rigid Quasilocal Frames”, Doctoral thesis (University of Waterloo, 2014).

[193]P. L. McGrath, M. Chanona, R. J. Epp, M. J. Koop, and R. B. Mann, “Post-Newtonian conservation laws in rigid quasilocal frames”, Classical and Quantum Gravity **31**, 095006 (2014).

[194]P. L. McGrath, R. J. Epp, and R. B. Mann, “Quasilocal conservation laws: why we need them”, Classical and Quantum Gravity **29**, 215012 (2012).

[195]Y. Mino, M. Sasaki, and T. Tanaka, “Gravitational radiation reaction to a particle motion”, Physical Review D **55**, 3457 (1997).

[196]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973).

[197]V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, “Theory of cosmological perturbations”, Physics Reports **215**, 203 (1992).

[198]A. Nagar and L. Rezzolla, “Gauge-invariant non-spherical metric perturbations of Schwarzschild black-hole spacetimes”, Classical and Quantum Gravity **22**, R167 (2005).

[199]M. Nakahara, *Geometry, Topology and Physics*, 2nd edition (CRC Press, Bristol, 2003).

[200]T. Nakamura, K. Oohara, and Y. Kojima, “General Relativistic Collapse to Black Holes and Gravitational Waves from Black Holes”, Progress of Theoretical Physics Supplement **90**, 1 (1987).

[201]E. Newman and R. Penrose, “An Approach to Gravitational Radiation by a Method of Spin Coefficients”, Journal of Mathematical Physics **3**, 566 (1962).

[202]R. P. A. C. Newman, “Compact space-times and the no-return theorem”, General Relativity and Gravitation **18**, 1181 (1986).

[203]I. Newton, *Philosophiae naturalis principia mathematica.* (Jussu Societatis Regiæ ac Typis Josephi Streater, London, 1687).

[204]M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[205]E. T. Olsen, "Classical mechanics and entropy", Foundations of Physics Letters **6**, 327 (1993).

[206]M. Oltean, L. Bonetti, A. D. A. M. Spallicci, and C. F. Sopuerta, "Entropy theorems in classical mechanics, general relativity, and the gravitational two-body problem", Physical Review D **94**, 064049 (2016).

[207]M. Oltean, R. J. Epp, P. L. McGrath, and R. B. Mann, "Geoids in general relativity: Geoid quasilocal frames", in *The Fourteenth Marcel Grossmann Meeting on General Relativity*, edited by M. Bianchi, R. T. Jantzen, and R. Ruffini (World Scientific, Singapore, 2017), p. 3682.

[208]M. Oltean, R. J. Epp, P. L. McGrath, and R. B. Mann, "Geoids in general relativity: geoid quasilocal frames", Classical and Quantum Gravity **33**, 105001 (2016).

[209]M. Oltean, R. J. Epp, C. F. Sopuerta, A. D. A. M. Spallicci, and R. B. Mann, "The motion of localized sources in general relativity: gravitational self-force from quasilocal conservation laws", arXiv:1907.03012 [astro-ph, physics:gr-qc, physics:hep-th] (2019).

[210]M. Oltean, C. F. Sopuerta, and A. D. A. M. Spallicci, "A frequency-domain implementation of the particle-without-particle approach to EMRIs", Journal of Physics: Conference Series **840**, 012056 (2017).

[211]M. Oltean, C. F. Sopuerta, and A. D. A. M. Spallicci, "Particle-without-Particle: A Practical Pseudospectral Collocation Method for Linear Partial Differential Equations with Distributional Sources", Journal of Scientific Computing **79**, 827 (2019).

[212]T. Padmanabhan, *Gravitation: Foundations and Frontiers* (Cambridge University Press, Cambridge, 2010).

[213]K. M. Peterson, "Über die Biegung der Flächen", Doctoral thesis (Dorpat University, 1853).

[214]N. A. Petersson, O. O'Reilly, B. Sjögreen, and S. Bydlon, "Discretizing singular point sources in hyperbolic wave propagation problems", Journal of Computational Physics **321**, 532–555 (2016).

[215]N. A. Petersson and B. Sjogreen, "Stable Grid Refinement and Singular Source Discretization for Seismic Wave Simulations", Commun. Comput. Phys. **8**, 1074 (2010).

[216]R. Peyret, "Chebyshev method", in *Spectral Methods for Incompressible Viscous Flow*, Applied Mathematical Sciences (Springer, New York, NY, 2002), pp. 39–100.

[217]J.-F. Pietschmann, "On some partial differential equation models in socio-economic contexts - analysis and numerical simulations", Doctoral thesis (University of Cambridge, 2012).

[218]F. A. E. Pirani, "On the Quantization of the Gravitational Field of General Relativity", PhD thesis (Carnegie Institute of Technology, 1951).

[219]F. A. E. Pirani and A. Schild, "On the Quantization of Einstein's Gravitational Field Equations", Physical Review **79**, 986 (1950).

[220] H. Poincaré, “Sur le problème des trois corps et les équations de la dynamique”, Acta mathematica **13**, 1 (1890).

[221] H. Poincaré, “Sur les tentatives d’explication mecanique des principes de la thermodynamique”, C. R. Acad. Sci. (Paris) **108**, 550 (1889).

[222] E. Poisson, *A Relativist’s Toolkit: The Mathematics of Black-Hole Mechanics* (Cambridge University Press, Cambridge, 2007).

[223] E. Poisson, “An introduction to the Lorentz-Dirac equation”, arXiv:gr-qc/9912045 (1999).

[224] E. Poisson, A. Pound, and I. Vega, “The Motion of Point Particles in Curved Spacetime”, Living Reviews in Relativity **14**, 7 (2011).

[225] A. Pound, “Gauge and motion in perturbation theory”, Physical Review D **92**, 044021 (2015).

[226] A. Pound, “Motion of Small Objects in Curved Spacetimes: An Introduction to Gravitational Self-Force”, in *Equations of Motion in Relativistic Gravity*, edited by D. Puetzfeld, C. Lämmerzahl, and B. Schutz, Fundamental Theories of Physics 179 (Springer, Cham, 2015), p. 399.

[227] A. Pound, “Nonlinear gravitational self-force: Second-order equation of motion”, Physical Review D **95**, 104056 (2017).

[228] A. Pound, “Second-Order Gravitational Self-Force”, Physical Review Letters **109**, 051101 (2012).

[229] A. Pound, “Self-consistent gravitational self-force”, Physical Review D **81**, 024023 (2010).

[230] C. Prescod-Weinstein and E. Bertschinger, “An extension of the Faddeev–Jackiw technique to fields in curved spacetimes”, Classical and Quantum Gravity **32**, 075011 (2014).

[231] F. Pretorius, “Evolution of Binary Black-Hole Spacetimes”, Physical Review Letters **95**, 121101 (2005).

[232] L. R. Price, “Developments in the perturbation theory of algebraically special spacetimes”, Doctoral thesis (University of Florida, 2007).

[233] D. Puetzfeld and Y. N. Obukhov, “Generalized deviation equation and determination of the curvature in general relativity”, Physical Review D **93**, 044073 (2016).

[234] M. H. P. M. van Putten, “Entropic force in black hole binaries and its Newtonian limits”, Physical Review D **85**, 064046 (2012).

[235] T. C. Quinn and R. M. Wald, “Axiomatic approach to electromagnetic and gravitational radiation reaction of particles in curved spacetime”, Physical Review D **56**, 3381 (1997).

[236] T. Regge and C. Teitelboim, “Role of surface integrals in the Hamiltonian formulation of general relativity”, Annals of Physics **88**, 286 (1974).

[237] T. Regge and J. A. Wheeler, “Stability of a Schwarzschild Singularity”, Physical Review **108**, 1063 (1957).

[238] J. Renn, ed., *The Genesis of General Relativity*, Vol. 1-4, Boston Studies in the Philosophy and History of Science (Springer Netherlands, 2007).

[239]P. Ritter, S. Aoudia, A. D. A. M. Spallicci, and S. Cordier, “Indirect (source-free) integration method. I. Wave-forms from geodesic generic orbits of EMRIs”, International Journal of Geometric Methods in Modern Physics **13**, 1650021 (2015).

[240]P. Ritter, S. Aoudia, A. D. A. M. Spallicci, and S. Cordier, “Indirect (source-free) integration method. II. Self-force consistent radial fall”, International Journal of Geometric Methods in Modern Physics **13**, 1650019 (2015).

[241]P. Ritter, S. Aoudia, A. D. A. M. Spallicci, and S. Cordier, “Indirect (source-free) integration method. II. Self-force consistent radial fall”, International Journal of Geometric Methods in Modern Physics **13**, 1650019 (2016).

[242]P. Ritter, A. D. A. M. Spallicci, S. Aoudia, and S. Cordier, “A fourth-order indirect integration method for black hole perturbations: even modes”, Classical and Quantum Gravity **28**, 134012 (2011).

[243]B. Romanowicz and A. Dziewonski, eds., *Seismology and Structure of the Earth*, Treatise on Geophysics (Elsevier, 2007).

[244]T. Rothman, “Editor’s note: the field of a single centre in Einstein’s theory of gravitation, and the motion of a particle in that field”, General Relativity and Gravitation **34**, 1541 (2002).

[245]C. Rovelli, “General relativistic statistical mechanics”, Physical Review D **87**, 084055 (2012).

[246]C. Rovelli, *Quantum Gravity* (Cambridge University Press, Cambridge, 2007).

[247]C. Sagan, *Cosmos* (Random House, New York, 1980).

[248]V. Sahni, Y. Shtanov, and A. Toporensky, “Arrow of time in dissipationless cosmology”, Classical and Quantum Gravity **32**, 182001 (2015).

[249]V. Sahni and A. Toporensky, “Cosmological hysteresis and the cyclic universe”, Physical Review D **85**, 123542 (2012).

[250]D. Santos-Oliván and C. F. Sopuerta, “Pseudo-Spectral Collocation Methods for Hyperbolic Equations with Arbitrary Precision: Applications to Relativistic Gravitation”, arXiv:1803.00858 [gr-qc, physics:physics] (2018).

[251]O. Sarbach and M. Tiglio, “Continuum and Discrete Initial-Boundary Value Problems and Einstein’s Field Equations”, Living Reviews in Relativity **15**, 9 (2012).

[252]S.-i. Sasa and Y. Yokokura, “Thermodynamic Entropy as a Noether Invariant”, Physical Review Letters **116**, 140601 (2016).

[253]M. Sasaki and H. Tagoshi, “Analytic Black Hole Perturbation Approach to Gravitational Radiation”, Living Reviews in Relativity **6**, 6 (2003).

[254]J. S. Schiffrin and R. M. Wald, “Measure and Probability in Cosmology”, Physical Review D **86**, 023521 (2012).

[255]L. Schwartz, “Sur l’impossibilité de la multiplication des distributions”, C. R. Acad. Sci. Paris **29**, 847–848 (1954).

[256]L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1957).

[257]K. Schwarzschild, “Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie”, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 189 (1916).

[258]P. M. Shearer, *Introduction to Seismology*, 2nd edition (Cambridge University Press, Cambridge; New York, 2009).

[259]M. Shibata, *Numerical Relativity* (World Scientific, Singapore, 2015).

[260]M. Shibata and T. Nakamura, “Evolution of three-dimensional gravitational waves: Harmonic slicing case”, Physical Review D **52**, 5428 (1995).

[261]B.-C. Shin and J.-H. Jung, “Spectral collocation and radial basis function methods for one-dimensional interface problems”, Applied Numerical Mathematics **61**, 911 (2011).

[262]A. C. d. Silva, *Lectures on Symplectic Geometry* (Springer, Berlin, 2008).

[263]A. C. d. Silva, “Symplectic geometry”, in *Handbook of Differential Geometry*, Vol. 2, edited by F. E. Dillen and L. A. Verstraelen (North-Holland, 2006), p. 79.

[264]L. Sklar, *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics* (Cambridge University Press, Cambridge, 1995).

[265]A. D. A. M. Spallicci and P. Ritter, “A fully relativistic radial fall”, International Journal of Geometric Methods in Modern Physics **11**, 1450090 (2014).

[266]A. D. A. M. Spallicci, P. Ritter, and S. Aoudia, “Self-force driven motion in curved spacetime”, International Journal of Geometric Methods in Modern Physics **11**, 1450072 (2014).

[267]A. D. Spallicci, P. Ritter, S. Jubertie, S. Cordier, and S. Aoudia, “Towards a Self-consistent Orbital Evolution for EMRIs”, Astronomical Society of the Pacific Conference Series **467**, 221 (2012).

[268]I. Stakgold and M. J. Holst, *Green’s Functions and Boundary Value Problems*, 3rd edition (Wiley, Hoboken, 2011).

[269]J. M. Stewart and M. Walker, “Perturbations of space-times in general relativity”, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **341**, 49 (1974).

[270]J. Stewart, *Advanced General Relativity* (Cambridge University Press, Cambridge, 1993).

[271]S. Suhr and K. Zehmisch, “Linking and closed orbits”, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg **86**, 133 (2016).

[272]L. B. Szabados, “Quasi-Local Energy-Momentum and Angular Momentum in GR: A Review Article”, Living Reviews in Relativity **7**, 4 (2004).

[273]S. A. Teukolsky, “Perturbations of a Rotating Black Hole. I. Fundamental Equations for Gravitational, Electromagnetic, and Neutrino-Field Perturbations”, The Astrophysical Journal **185**, 635 (1973).

[274]S. A. Teukolsky, “Rotating Black Holes: Separable Wave Equations for Gravitational and Electromagnetic Perturbations”, Physical Review Letters **29**, 1114 (1972).

[275]S. A. Teukolsky, “The Kerr metric”, Classical and Quantum Gravity **32**, 124006 (2015).

[276] F. J. Tipler, “General Relativity and the Eternal Return”, in *Essays in General Relativity*, edited by F. J. Tipler (Academic Press, 1980), pp. 21–37.

[277] F. J. Tipler, “General relativity, thermodynamics, and the Poincaré cycle”, Nature **280**, 203 (1979).

[278] A.-K. Tornberg and B. Engquist, “Numerical approximations of singular source terms in differential equations”, Journal of Computational Physics **200**, 462 (2004).

[279] L. N. Trefethen, *Spectral Methods in MATLAB* (SIAM: Society for Industrial and Applied Mathematics, Philadelphia, 2001).

[280] I. Vega and S. Detweiler, “Regularization of fields for self-force problems in curved spacetime: Foundations and a time-domain application”, Physical Review D **77**, 084008 (2008).

[281] E. P. Verlinde, “On the Origin of Gravity and the Laws of Newton”, Journal of High Energy Physics **1104**, 029 (2010).

[282] V. Villani Cedric, “H-Theorem and beyond: Boltzmann’s entropy in today’s mathematics”, in *Boltzmann’s Legacy, ESI Lectures in Mathematics and Physics*, edited by G. Gallavotti, W. L. Reiter, and J. Yngvason (European Mathematical Society, Zurich, 2008), p. 129.

[283] C. V. Vishveshwara, “Stability of the Schwarzschild Metric”, Physical Review D **1**, 2870 (1970).

[284] R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).

[285] R. M. Wald, “The Thermodynamics of Black Holes”, Living Reviews in Relativity **4**, 6 (2001).

[286] B. Wardell, “Self-force: Computational Strategies”, in *Equations of Motion in Relativistic Gravity*, edited by D. Puetzfeld, C. Lämmerzahl, and B. Schutz, Fundamental Theories of Physics 179 (Springer, Cham, 2015), p. 487.

[287] J. O. Weatherall, “Geometry and Motion in General Relativity”, arXiv:1810.09046 [gr-qc, physics:physics] (2018).

[288] H. Wilbraham, “On a certain periodic function”, The Cambridge and Dublin Mathematical Journal **3**, 198 (1848).

[289] D. C. Wilkins, “Bound Geodesics in the Kerr Metric”, Physical Review D **5**, 814 (1972).

[290] H.-T. Yau, “The work of Cedric Villani”, in *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010)*, edited by R. Bhatia, A. Pal, G. Rangarajan, V. Srinivas, and M. Vanninathan (World Scientific, 2011), p. 87.

[291] J. W. York, “Role of Conformal Three-Geometry in the Dynamics of Gravitation”, Physical Review Letters **28**, 1082 (1972).

[292] A. Zenginoglu, “Hyperboloidal layers for hyperbolic equations on unbounded domains”, Journal of Computational Physics **230**, 2286 (2011).

[293] F. J. Zerilli, “Gravitational Field of a Particle Falling in a Schwarzschild Geometry Analyzed in Tensor Harmonics”, Physical Review D **2**, 2141 (1970).

[294]J. G. Zhou, D. M. Causon, D. M. Ingram, and C. G. Mingham, “Numerical solutions of the shallow water equations with discontinuous bed topography”, International Journal for Numerical Methods in Fluids **38**, 769 (2002).