Calculus of Variations

A positive quasilocal mass for causal variational principles

Felix Finster[1] · **Niky Kamran[2]**

Abstract

A new inequality for a nonlinear surface layer integral is proved for minimizers of causal variational principles. This inequality is applied to obtain a new proof of the positive mass theorem with volume constraint. Next, a positive mass theorem without volume constraint is stated and proved by introducing and using the concept of asymptotic alignment. Moreover, a positive quasilocal mass and a synthetic definition of scalar curvature are introduced in the setting of causal variational principles. Our notions and results are illustrated by the explicit examples of causal fermion systems constructed in ultrastatic spacetimes and the Schwarzschild spacetime. In these examples, the correspondence to the ADM mass and similarities to the Brown–York mass are worked out.

Mathematics Subject Classification 49S05 · 49Q20 · 58C35 · 28A33

Contents

Communicated by A. Mondino.

✉ Felix Finster
   finster@ur.de

   Niky Kamran
   niky.kamran@mcgill.ca

[1] Fakultät für Mathematik, Universität Regensburg, 93040 Regensburg, Germany

[2] Department of Mathematics and Statistics, McGill University, Montréal, Canada

  Springer

1 Introduction

The theory of *causal fermion systems* is a recent approach to fundamental physics where spacetime is no longer modelled by a Lorentzian manifold but may instead have a more general, possibly discrete structure on a microscopic length scale (which can be thought of as the Planck scale). In the setting of causal fermion systems, the physical equations are formulated via a variational principle, the *causal action principle* (for the general context see the reviews [12, 23], the textbooks [11, 22] or the website [36]). *Causal variational principles* were introduced in [10] as a mathematical generalization of the causal action principle. From the mathematical perspective, causal fermion systems and causal variational principles are of interest because they provide a setting for describing and analyzing non-smooth geometries and singular spaces (see [17, 18] for more details on different aspects).

In general terms, given a manifold $\mathcal{G}$ together with a non-negative function $\mathcal{L} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}_0^+$, in a causal variational principle one minimizes the action $\mathcal{S}$ defined by

$$\mathcal{S}(\mu) = \int_{\mathcal{G}} d\mu(x) \int_{\mathcal{G}} d\mu(y) \, \mathcal{L}(x, y)$$

under variations of the measure μ on $\mathcal{G}$, keeping the total volume $\mu(\mathcal{G})$ fixed (for the precise mathematical setup see Sect. 2.1 below). The support of the measure μ denoted by

$$N := \operatorname{supp} \mu \subset \mathcal{G} \tag{1.1}$$

has the interpretation as the underlying *space* or *spacetime*.

In order to clarify the nature of the interaction described by the causal variational principle, it is an important task to investigate whether quantities which are defined geometrically and which have been extensively studied in the context of classical general relativity and differential geometry can be formulated and analyzed in the broader setting of causal variational principles. In [30] a notion of *total mass* $\mathfrak{M}^{\mathrm{tot}}$ was introduced for a class of causal variational principles which include the causal action principle for static causal fermion systems. To this end, one considers two critical measures μ (describing the vacuum) and $\tilde{\mu}$ (describing the gravitating system) and takes the limit of double integrals

$$\mathfrak{M}^{\mathrm{tot}} = \lim_{\Omega \nearrow N, \; \tilde{\Omega} \nearrow \tilde{N}}$$
$$\times \left(\int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) - \int_{\Omega} d\mu(\mathbf{x}) \int_{\tilde{N} \setminus \tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) \right), \quad (1.2)$$

where $\Omega \nearrow N$ and $\tilde{\Omega} \nearrow \tilde{N}$ denote exhaustions under the

$$\text{volume constraint} \, \mu(\Omega) = \tilde{\mu}(\tilde{\Omega}) < \infty \,. \tag{1.3}$$

The limit in (1.2) is well-defined provided that the measures μ and $\tilde{\mu}$ are *asymptotically close* (for the precise definition and more details see Sect. 2.4). Moreover, a *positive mass theorem* was proven which states that the total mass is non-negative, provided that a suitable local energy condition holds. The proof of this theorem was inspired by and bears some similarity with Witten's spinor proof of the positive mass theorem [38, 42]. Moreover, in [30] it was shown how the ADM mass is recovered in a limiting case in which the measure $\tilde{\mu}$ is constructed from a static asymptotically flat Lorentzian spacetime.

In the present paper we shall improve and generalize the concepts, methods and results in [30] in various ways. A key ingredient is a novel *positivity argument for a surface layer integral* (see Sect. 3). In the simplest case, our positivity argument yields that, again for critical measures μ and $\tilde{\mu}$ and any subsets $\tilde{\Omega} \subset N$ and $\Omega \subset N$ satisfying the volume constraint (1.3), the following combination of double integrals is non-negative,

$$0 \le \mathfrak{M}_{\tilde{\mu}, \mu}(\tilde{\Omega}, \Omega) := 2 \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$- \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{N} \setminus \tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) - \int_{\Omega} \mu(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \,. \tag{1.4}$$

Asymptotically as $\tilde{\Omega} \nearrow \tilde{N}$ and $\Omega \nearrow N$, this goes over to (1.2), giving a new and much simpler proof of the above-mentioned positive mass theorem (see Sect. 4). It is remarkable that this new proof does not require a local energy condition. Instead, it is a general consequence of the assumption that the vacuum measure μ is a minimizer of the causal variational principle.

Another main concern of the present paper is the proper *treatment of the volume constraint* (1.3). Our motivation comes from the fact that this volume constraint is not fully convincing because it has no counterpart in general relativity, where the ADM mass is defined purely in terms of the asymptotic geometry at infinity. This raises the important question of if and how the volume constraint can be removed. In order to address these questions, we introduce the concept of *alignment*. The idea is to remove the freedom in identifying N and $\tilde{N}$ by imposing a condition which can be evaluated locally in a neighborhood of each point $\boldsymbol{\zeta} \in N$ (see (5.9) in Definition 5.4). It turns out that the resulting identification of Ω with $\tilde{\Omega}$ satisfies the volume inequality $\tilde{\mu}(\tilde{\Omega}) \ge \mu(\Omega)$, provided that the so-called *local volume condition* holds (see Definition 5.6). In this way, we obtain a positive mass theorem without volume constraint (see Theorem 5.8). As a further generalization, we define an *equivariant* positive mass theorem, where we minimize over a group $\mathcal{G}$ of diffeomorphisms acting on $\mathcal{G}$ which describe symmetries of the Lagrangian (see Definition 6.2 and Theorem 6.3).

Starting from the above concepts and results, we also succeed in introducing a *quasilocal mass*. Given a subset $\tilde{\Omega} \subset \tilde{N}$, it is defined by

$$\mathfrak{M}(\tilde{\Omega}) := \inf_{\Phi, \Omega} \mathfrak{M}_{\tilde{\mu}, \Phi_* \mu}(\tilde{\Omega}, \Omega) \,,$$

where we take the infimum over symmetry transformations $\Phi \in \mathcal{G}$ and subsets $\Omega \subset \Phi(N)$ which must satisfy certain admissibility conditions (see Definition 7.2). This quasilocal mass

is shown to be non-negative (Theorem 7.3). Moreover, it bounds the total mass from below (see Theorem 8.1). In the example of causal fermion systems constructed in ultrastatic spacetimes, our quasilocal mass has surprising similarities with the Brown–York mass [4, 37, 39] (see Theorem 10.4). Finally, we propose a notion of *synthetic scalar curvature* (see Definition 9.1). Again in the example of ultrastatic spacetimes, this definition reduces, up to a constant, to the scalar curvature of the spatial metric (see Theorem 10.6). Our findings and results are illustrated by detailed computations and examples in ultrastatic spacetimes and the Schwarzschild spacetime.

A major advantage of describing the concept of mass in the setting of causal variational principles is that the *regularity and smoothness assumptions can be weakened* drastically. Indeed, the definition of mass (1.2) does not require the spaces N, $\tilde{N} \subset \mathcal{G}$ to be smooth or even topological manifolds. Instead, all we need are certain boundedness and decay assumptions on the measures μ and $\tilde{\mu}$ near infinity (for details see Sect. 2.4). Clearly, in order to compare our notions to the classical concepts of total and quasilocal mass, we need to go back to the smooth setting. Moreover, local notions like the alignment, the local volume condition and scalar curvature only make sense in smooth spaces. For clarity in presentation, we begin as general as possible and specify the regularity and smoothness assumptions on the way whenever needed. More precisely, starting from Sect. 5 we assume that the vacuum N is smooth and translation invariant (see (5.1)) and that the gravitating spacetime $\tilde{N}$ has a smooth manifold structure in the asymptotic end. Beginning from Sect. 5.3, we need to assume that $\tilde{N}$ has a smooth manifold structure everywhere. An exception is the bound of the total mass in terms of the quasilocal mass (see Theorem 8.1), were it suffices to assume smoothness in the exterior region $\tilde{N} \backslash \tilde{U}$.

The paper is organized as follows. After the necessary preliminaries on causal variational principles and the total mass (Sect. 2), the general positivity statement for a nonlinear surface layer integral is introduced in various versions (Sect. 3). As an application, we give the new proof of the positive mass theorem (Sect. 4). Then the volume constraint is removed by working with asymptotic alignment (Sect. 5). An equivariant version of the positive mass theorem is stated and proven, where we minimize over a group of isometries of the Lagrangian (Sect. 6). We proceed by defining a quasilocal mass (Sect. 7) and studying its relation to the total mass (Sect. 8). Moreover, a notion of *synthetic scalar curvature* is introduced (Sect. 9). Finally, in Sect. 10 it is worked out that, for causal fermion systems constructed in ultrastatic spacetimes, our synthetic scalar curvature reduces to the scalar curvature of the Riemannian metric. Moreover, we work out an interesting similarity of our quasilocal mass with the Brown–York mass. We conclude the paper with detailed computations in the example of causal fermion systems in the Schwarzschild spacetime (Appendix A).

2 Preliminaries

This section provides the necessary background on causal variational principles and the definition of the total mass. More details can be found in [30].

2.1 Causal variational principles in the non-compact setting

We consider causal variational principles in the non-compact setting as introduced in [25, Section 2]. Thus we let $\mathcal{G}$ be a (possibly non-compact) smooth manifold of dimension $m \geq 1$

and μ a (positive) Borel measure on $\mathcal{G}$ (the *universal measure*). Moreover, we are given a non-negative function $\mathcal{L} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}_0^+$ (the *Lagrangian*) with the following properties:

(i) $\mathcal{L}$ is symmetric: $\mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathcal{L}(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{G}$.
(ii) $\mathcal{L}$ is lower semi-continuous, i.e. for all sequences $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{y}_{n'} \to \mathbf{y}$,

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) \leq \liminf_{n, n' \to \infty} \mathcal{L}(\mathbf{x}_n, \mathbf{y}_{n'}) .$$

The *causal variational principle* is to minimize the action

$$\mathcal{S}(\mu) = \int_{\mathcal{G}} d\mu(\mathbf{x}) \int_{\mathcal{G}} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \tag{2.1}$$

under variations of the measure μ, keeping the total volume $\mu(\mathcal{G})$ fixed (*volume constraint*).

If the total volume $\mu(\mathcal{G})$ is finite, one minimizes (2.1) over all regular Borel measures with the same total volume. If the total volume $\mu(\mathcal{G})$ is infinite, however, it is not obvious how to implement the volume constraint, making it necessary to proceed as follows. We make the following additional assumptions:

(iii) The measure μ is *locally finite* (meaning that any $\mathbf{x} \in \mathcal{G}$ has an open neighborhood U with $\mu(U) < \infty$) and *regular* (meaning that the measure of a set can be recovered by approximation from inside with compact and from outside with open sets).
(iv) The function $\mathcal{L}(\mathbf{x}, .)$ is μ-integrable for all $\mathbf{x} \in \mathcal{G}$, giving a lower semi-continuous and bounded function on $\mathcal{G}$.

Given a regular Borel measure μ on $\mathcal{G}$, we vary over all regular Borel measures $\tilde{\mu}$ with

$$|\tilde{\mu} - \mu|(\mathcal{G}) < \infty \quad \text{and} \quad (\tilde{\mu} - \mu)(\mathcal{G}) = 0 \tag{2.2}$$

(where |.| denotes the total variation of a measure). We then consider the difference of the actions defined by

$$\begin{aligned}
\left(\mathcal{S}(\tilde{\mu}) - \mathcal{S}(\mu) \right) := &\int_{\mathcal{G}} d(\tilde{\mu} - \mu)(\mathbf{x}) \int_{\mathcal{G}} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \\
&+ \int_{\mathcal{G}} d\mu(\mathbf{x}) \int_{\mathcal{G}} d(\tilde{\mu} - \mu)(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) + \int_{\mathcal{G}} d(\tilde{\mu} - \mu)(\mathbf{x}) \int_{\mathcal{G}} d(\tilde{\mu} - \mu)(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) .
\end{aligned} \tag{2.3}$$

The measure μ is said to be a *minimizer* of the causal action with respect to *variations of finite volume* if this difference is non-negative for all $\tilde{\mu}$ satisfying (2.2),

$$\left(\mathcal{S}(\tilde{\mu}) - \mathcal{S}(\mu) \right) \geq 0 . \tag{2.4}$$

These variations of the causal action are well-defined. The existence theory for minimizers is developed in [27]. It is shown in [25, Lemma 2.3] that a minimizer satisfies the *Euler–Lagrange (EL) equations* which state that for a suitable value of the parameter $\mathfrak{s} > 0$, the lower semi-continuous function $\ell : \mathcal{G} \to \mathbb{R}_0^+$ defined by

$$\ell(\mathbf{x}) := \int_{\mathcal{G}} \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) - \mathfrak{s} \tag{2.5}$$

is minimal and vanishes on the support (1.1) of the measure,

$$\ell|_N \equiv \inf_{\mathcal{G}} \ell = 0 . \tag{2.6}$$

For further details we refer to [25, Section 2].

2.2 The restricted Euler–Lagrange equations and jets

The EL equations (2.6) are nonlocal in the sense that they make a statement on ℓ even for points $\mathbf{x} \in \mathcal{G}$ which are far away from N. It turns out that for the applications in this paper, it is preferable to evaluate the EL equations locally in a neighborhood of N. This leads to the *restricted EL equations* introduced in [25, Section 4]. We here give a slightly less general version of these equations which is sufficient for our purposes. In order to explain how the restricted EL equations come about, we begin with the simplified situation that the function ℓ is smooth. In this case, the minimality of ℓ implies that the derivative of ℓ vanishes on N, i.e.

$$\ell|_N \equiv 0 \quad \text{and} \quad D\ell|_N \equiv 0 \tag{2.7}$$

(where $D\ell(p) : T_p\mathcal{G} \to \mathbb{R}$ is the derivative). In order to combine these two equations in a compact form, it is convenient to consider a pair $\mathfrak{u} := (a, \mathbf{u})$ consisting of a real-valued function a on N and a vector field $\mathbf{u}$ on $T\mathcal{G}$ along N, and to denote the combination of multiplication and directional derivative by

$$\nabla_\mathfrak{u}\ell(\mathbf{x}) := a(\mathbf{x})\,\ell(\mathbf{x}) + \big(D_\mathbf{u}\ell\big)(\mathbf{x}) \,. \tag{2.8}$$

Then the Eqs. (2.7) imply that $\nabla_\mathfrak{u}\ell(\mathbf{x})$ vanishes for all $\mathbf{x} \in N$. The pair $\mathfrak{u} = (a, \mathbf{u})$ is referred to as a *jet*.

In the general lower-continuous setting, one must be careful because the directional derivative $D_\mathbf{u}\ell$ in (2.8) need not exist. Our method for dealing with this problem is to restrict attention to vector fields for which the directional derivative is well-defined. Moreover, we must specify the regularity assumptions on a and u. To begin with, we always assume that a and u are *smooth* in the sense that they have a smooth extension to the manifold $\mathcal{G}$. Thus the jet $\mathfrak{u}$ should be an element of the jet space

$$\mathfrak{J} := \big\{ \mathfrak{u} = (a, \mathbf{u}) \text{ with } a \in C^\infty(N, \mathbb{R}) \text{ and } \mathbf{u} \in \Gamma(N, T\mathcal{G}) \big\} \,,$$

where $C^\infty(N, \mathbb{R})$ and $\Gamma(N, T\mathcal{G})$ denote the space of real-valued functions and vector fields on N, respectively, which admit smooth extensions to $\mathcal{G}$.

Clearly, the fact that a jet $\mathfrak{u}$ is smooth does not imply that the functions ℓ or $\mathcal{L}$ are differentiable in the direction of $\mathfrak{u}$. This must be ensured by additional conditions which are satisfied by suitable subspaces of $\mathfrak{J}$ which we now introduce. First, we let Γ^{diff} be those vector fields for which the directional derivative of the function ℓ exists,

$$\Gamma^{\text{diff}} = \big\{ \mathbf{u} \in C^\infty(N, T\mathcal{G}) \,\big|\, D_\mathbf{u}\ell(\mathbf{x}) \text{ exists for all } \mathbf{x} \in N \big\} \,.$$

This gives rise to the jet space

$$\mathfrak{J}^{\text{diff}} := C^\infty(N, \mathbb{R}) \oplus \Gamma^{\text{diff}} \subset \mathfrak{J} \,.$$

For the jets in $\mathfrak{J}^{\text{diff}}$, the combination of multiplication and directional derivative in (2.8) is well-defined. We choose a linear subspace $\mathfrak{J}^{\text{test}} \subset \mathfrak{J}^{\text{diff}}$ with the property that its scalar and vector components are both vector spaces,

$$\mathfrak{J}^{\text{test}} = C^{\text{test}}(N, \mathbb{R}) \oplus \Gamma^{\text{test}} \subseteq \mathfrak{J}^{\text{diff}} \,,$$

and the scalar component is nowhere trivial in the sense that

$$\text{for all } \mathbf{x} \in N \text{ there is } a \in C^{\text{test}}(N, \mathbb{R}) \text{ with } a(\mathbf{x}) \neq 0 \,.$$

Then the *restricted EL equations* read (for details cf. [25, (eq. (4.10)]])

$$\nabla_{\mathfrak{u}} \ell|_N = 0 \qquad \text{for all } \mathfrak{u} \in \mathfrak{J}^{\text{test}} . \tag{2.9}$$

The purpose of introducing $\mathfrak{J}^{\text{test}}$ is that it gives the freedom to restrict attention to the portion of information in the EL equations which is relevant for the application in mind. For example, if one is interested only in the macroscopic dynamics, one can choose $\mathfrak{J}^{\text{test}}$ to be composed of jets pointing in directions where the microscopic fluctuations of ℓ are disregarded.

We finally point out that the restricted EL equations (2.9) do not hold only for minimizers, but also for critical points of the causal action. For brevity, we also refer to a measure which satisfies the restricted EL equations (2.9) as a *critical measure*.

2.3 The linearized field equations and inner solutions

In words, the *linearized field equations* describe variations of the measure μ which preserve the EL equations. In order to make this statement mathematically precise, we consider variations where we multiply μ by a weight function f_τ and then take the push-forward with respect to a mapping F_τ from N to $\mathcal{G}$. More precisely, we consider the ansatz

$$\tilde{\mu}_\tau = (F_\tau)_* \big(f_\tau \, \mu \big) , \tag{2.10}$$

where $f_\tau \in C^\infty(N, \mathbb{R}^+)$ and $F_\tau \in C^\infty(N, \mathcal{G})$ are smooth mappings, and $(F_\tau)_* \mu$ denotes the push-forward (defined for a subset $\Omega \subset \mathcal{G}$ by $((F_\tau)_* \mu)(\Omega) = \mu(F_\tau^{-1}(\Omega))$; see for example [2, Section 3.6]). Demanding that the family of measures (2.10) is critical for all τ implies that the jet $\mathfrak{v}$ defined by

$$\mathfrak{v}(x) := \frac{d}{d\tau} \big(f_\tau(x), F_\tau(x) \big) \Big|_{\tau=0}$$

satisfies the linearized field equations

$$\langle \mathfrak{u}, \Delta \mathfrak{v} \rangle|_N = 0 \qquad \text{for all } \mathfrak{u} \in \mathfrak{J}^{\text{test}} , \tag{2.11}$$

where

$$\langle \mathfrak{u}, \Delta \mathfrak{v} \rangle(\mathbf{x}) := \nabla_{\mathfrak{u}} \bigg(\int_N \big(\nabla_{1,\mathfrak{v}} + \nabla_{2,\mathfrak{v}} \big) \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) - \nabla_{\mathfrak{v}} \, \mathfrak{s} \bigg) ,$$

where ∇_1 and ∇_2 refer to derivatives acting on the first and second argument of the Lagrangian, respectively. Here we do not enter the details but refer instead to the general derivation in [14, Section 3.3] or to the simplified presentation in the smooth setting in the textbook [22, Chapter 6]. We denote the vector space of all solutions of the linearized field equations by $\mathfrak{J}^{\text{lin}}$.

A specific class of linearized solutions are described by vector fields on N. In order to make sense of the notion of a vector field, we need to assume that N has the following smoothness property (we restrict attention to the three-dimensional case throughout).

Definition 2.1 Space $N := \operatorname{supp} \mu$ has a **smooth manifold structure** if the following conditions hold:

(i) N is a three-dimensional, smooth, oriented and connected submanifold of $\mathcal{G}$.
(ii) In a chart $(\mathbf{x}, U)$ of N, the measure μ is absolutely continuous with respect to the Lebesgue measure with a smooth, strictly positive weight function,

$$d\mu = h(\mathbf{x}) \, d^3\mathbf{x} \qquad \text{with} \quad h \in C^\infty(N, \mathbb{R}^+) .$$

Let $\mathbf{v} \in \Gamma(N, TN)$ be a vector field. Then, under the above assumptions, its *divergence* $\operatorname{div} \mathbf{v} \in C^\infty(N, \mathbb{R})$ can be defined by the relation

$$\int_N \operatorname{div} \mathbf{v}(\mathbf{x})\, \eta(\mathbf{x})\, d\mu(\mathbf{x}) = -\int_N D_{\mathbf{v}} \eta(\mathbf{x})\, d\mu(\mathbf{x})\,,$$

to be satisfied by all test functions $\eta \in C_0^\infty(N, \mathbb{R})$. In a local chart $(\mathbf{x}, U)$, the divergence is computed by

$$\operatorname{div} \mathbf{v} = \frac{1}{h}\, \partial_\alpha\big(h\, \mathbf{v}^\alpha\big)$$

(where, using the Einstein summation convention, we sum over $\alpha = 1, 2, 3$).

Definition 2.2 An **inner solution** is a jet $\mathfrak{v}$ of the form

$$\mathfrak{v} = (\operatorname{div} \mathbf{v}, \mathbf{v}) \qquad \text{with} \qquad \mathbf{v} \in \Gamma(N, TN)\,. \tag{2.12}$$

The name "inner *solution*" stems from the fact that a jet $\mathfrak{v}$ of the form (2.12) satisfies the linearized field equations (2.11). The reason is that, applying the Gauss divergence theorem, integrating its jet derivative of a compactly supported function gives zero, i.e. for for every $f \in C_0^1(N, \mathbb{R})$

$$\int_N \nabla_\mathfrak{v} f\, d\mu = \int_N \big(\operatorname{div} \mathbf{v}\, f + D_{\mathbf{v}} f\big)\, d\mu = \int_N \operatorname{div}\big(f\mathbf{v}\big)\, d\mu = 0\,.$$

Integrating by parts formally, one finds that

$$\begin{aligned}
\langle \mathfrak{u}, \Delta\mathfrak{v}\rangle_N &= \nabla_\mathfrak{u}\bigg(\int_N \big(\nabla_{1,\mathfrak{v}} + \nabla_{2,\mathfrak{v}}\big)\mathcal{L}(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y}) - \nabla_\mathfrak{v}\, \mathfrak{s}\bigg) \\
&= \nabla_\mathfrak{u}\bigg(\int_N \nabla_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y}) - \nabla_\mathfrak{v}\, \mathfrak{s}\bigg) \\
&= \nabla_\mathfrak{u} \nabla_\mathfrak{v} \ell(\mathbf{x}) = \nabla_\mathfrak{v}\big(\nabla_\mathfrak{u} \ell(\mathbf{x})\big) - \nabla_{D_\mathbf{v}\mathfrak{u}} \ell(\mathbf{x}) = 0\,.
\end{aligned}$$

Here $D_\mathbf{v}\mathfrak{u}$ denotes the partial derivative computed in given charts. Here we do not need to introduce a connection and work with covariant derivatives, simply because the corresponding summand $\nabla_{D_\mathbf{v}\mathfrak{u}} \ell(\mathbf{x})$ vanishes by the restricted EL equations. Moreover, the function $\nabla_\mathfrak{u} \ell$ vanishes identically on N in view of the restricted EL equations. Therefore, it is differentiable in the direction of every vector field on N, and this directional derivative is zero.

This formal computation can be made rigorous by imposing suitable regularity and decay assumptions of the vector field $\mathbf{v}$ near infinity. In order to avoid excessive overlap with previous works, we here omit the details and refer instead for example to [19, Section 3] or [30, Section 2.1.3].

2.4 The total mass as a surface layer integral

In [30] the total mass was defined for causal variational principles. Moreover, it was shown that in the example of asymptotically flat, static causal fermion system, this total mass gives back the familiar ADM mass. We now recall the basic definitions. We consider two measures: A measure μ which describes the vacuum spacetime, and another measure $\tilde{\mu}$ which typically describes a gravitating spacetime. In order to compare the measures μ and $\tilde{\mu}$, we introduce

the functions

$$\begin{cases} n \,:\, N \to \mathbb{R}_0^+ \cup \{\infty\} \,, & n(\mathbf{x}) = \displaystyle\int_{\tilde{N}} \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\tilde{\mu}(\mathbf{y}) \\[2mm] \tilde{n} \,:\, \tilde{N} \to \mathbb{R}_0^+ \cup \{\infty\} \,, & \tilde{n}(\mathbf{x}) = \displaystyle\int_{N} \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) \,. \end{cases}$$

For simplicity, throughout this paper we shall restrict attention to one asymptotic end (but all our methods and results could be extended in a straightforward way to several asymptotic ends). Then we need to make the following assumption.

Definition 2.3 The measures $\tilde{\mu}$ and μ are **asymptotically close** if they are both σ-finite with infinite total volume,

$$\tilde{\mu}(\tilde{N}) = \mu(N) = \infty \,,$$

but

$$\int_N \big| n(\mathbf{x}) - \mathfrak{s} \big| \, d\mu(\mathbf{x}) < \infty \quad \text{and} \quad \int_{\tilde{N}} \big| \tilde{n}(\mathbf{x}) - \mathfrak{s} \big| \, d\tilde{\mu}(\mathbf{x}) < \infty \,.$$

We now state the most general definition of the total mass.

Definition 2.4 Assume that μ and $\tilde{\mu}$ are asymptotically close. Then the **total mass** $\mathfrak{M}^{\mathrm{tot}}$ of $\tilde{\mu}$ relative to μ is defined by

$$\begin{aligned} \mathfrak{M}^{\mathrm{tot}} := \lim_{\Omega \nearrow N} \lim_{\tilde{\Omega} \nearrow \tilde{N}} \Bigg(&- \mathfrak{s}\Big(\tilde{\mu}(\tilde{\Omega}) - \mu(\Omega) \Big) \\ &+ \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) - \int_{\Omega} d\mu(\mathbf{x}) \int_{\tilde{N} \setminus \tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \Bigg) \,, \end{aligned} \qquad (2.13)$$

where the notation $\Omega \nearrow N$ means that we take an exhaustion of N by sets of finite μ-measure.

Restricting attention to sets $\tilde{\Omega}$ and Ω which satisfy the volume constraint (1.3), one gets back the definition of the mass (1.2) stated in the introduction.

For what follows, it is important that, when evaluating the double integrals in (2.13) asymptotically near infinity, one may *linearize* in the sense that $\tilde{\mu}$ may be described in the asymptotic end by a first order perturbation of the vacuum measure μ. The needed technical assumptions are subsumed in the following definition, which is a simplified and weakened version of [30, Definition 2.3].

Definition 2.5 The measure $\tilde{\mu}$ is **asymptotically flat** (with respect to μ) if it is asymptotically close to μ (see Definition 2.3) and has the following additional properties: There is a function $f \in C^\infty(N, \mathbb{R}_0^+)$ and a mapping $F \in C^\infty(N, \mathcal{G})$ such that

$$\tilde{\mu} = F_*\big(f\mu\big)$$

(where $F_*\mu$ is the push-forward measure defined by $(F_*\mu)(\tilde{\Omega}) = \mu(F^{-1}(\tilde{\Omega}))$). This transformation tends to the identity at infinity in the sense that there is a jet $\mathfrak{v} \in \mathfrak{J}^{\mathrm{vary}}$ with

$$\lim_{\Omega \nearrow N} \int_\Omega d\mu(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \Big(f(\mathbf{x}) \, \mathcal{L}\big(F(\mathbf{x}), \mathbf{y}\big) - \mathcal{L}(\mathbf{x}, \mathbf{y}) - \nabla_{1, \mathfrak{v}} \mathcal{L}(\mathbf{x}, \mathbf{y}) \Big) = 0 \qquad (2.14)$$

$$\lim_{\Omega \nearrow N} \int_\Omega d\mu(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \Big(\mathcal{L}\big(\mathbf{x}, F(\mathbf{y})\big) \, f(\mathbf{y}) - \mathcal{L}(\mathbf{x}, \mathbf{y}) - \nabla_{2, \mathfrak{v}} \mathcal{L}(\mathbf{x}, \mathbf{y}) \Big) = 0 \,. \qquad (2.15)$$

If this condition holds, the surface layer integral in (1.2) can be linearized to obtain

$$\mathfrak{M}^{\text{tot}} = \lim_{\Omega \nearrow N} \mathfrak{M}(\Omega, \mathfrak{v}) \,, \tag{2.16}$$

where the *linearized quasilocal mass* $\mathfrak{M}(\Omega, \mathfrak{v})$ is defined by

$$\mathfrak{M}(\Omega, \mathfrak{v}) := \int_{\Omega} d\mu(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \left(\nabla_{1, \mathfrak{v}} - \nabla_{2, \mathfrak{v}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \,. \tag{2.17}$$

The double integrals in (2.13) and (2.17) (and similarly in (1.2) and (1.4) in the introduction) have the structure of a *surface layer integral*, as we now briefly explain (for more details see [24, Section 2.3] and [19, Section 4]). The main point is that in the double integral (2.17) the two variables $\mathbf{x}$ and $\mathbf{y}$ are integrated over Ω and its complement $N \setminus \Omega$, respectively. In typical situations, the Lagrangian and its derivatives are very small if $\mathbf{x}$ and $\mathbf{y}$ are far apart. Therefore, we only get a relevant contribution to the double integral if both $\mathbf{x}$ and $\mathbf{y}$ are near the boundary of Ω. This picture can be made more quantitative if one assumes that the Lagrangian is of *short range* in the following sense. We let $d \in C^0(N \times N, \mathbb{R}_0^+)$ be a distance function on N. The assumption of short range means that $\mathcal{L}$ vanishes on distances larger than δ, i.e.

$$d(\mathbf{x}, \mathbf{y}) > \delta \quad \Longrightarrow \quad \mathcal{L}(\mathbf{x}, \mathbf{y}) = 0 = \nabla \mathcal{L}(\mathbf{x}, \mathbf{y}) \tag{2.18}$$

Then the integrand in (2.17) vanishes unless both arguments lie in a layer around the boundary of Ω of width δ, i.e.

$$\mathbf{x}, \mathbf{y} \in B_\delta(\partial \Omega) \,.$$

For the purpose of this paper, we do not need to assume that the Lagrangian is of short range. It suffices that it *decays on the scale δ*. Nevertheless, the reader who wants to avoid scaling arguments may assume the stronger assumption (2.18). We refer to δ as the *range* of the Lagrangian. The double integral in (2.13) can be regarded as a nonlinear version of the linear surface layer integral. Our above considerations again apply, provided that the measures μ and $\tilde{\mu}$ are close to each other near the boundary of $\tilde{\Omega}$ and Ω. This statement could be quantified in straightforward way a strict sense similar to (2.18); we here omit the details for brevity.

We finally comment on the positive mass theorem as proven in [30, Section 5]. Inspired by Witten's spinor proof of the positive mass theorem, in this theorem one works with a linear equation for the wave functions. This linear equation is obtained by linearizing the EL equations with respect to a parameter κ, being the Lagrange parameter of the so-called boundedness constraint. This linearization makes it necessary to assume the existence of a whole family (μ_κ) of minimizing measures. It is then shown under certain technical assumptions that if a suitable local energy condition holds, then the total mass is non-negative. We shall see that, in contrast to this result, the positive mass theorem that will be proved in Sect. 4 will be quite different in nature. In particular, it does not require a local energy condition. Instead, positivity of the total mass will be a direct consequence of the fact that the vacuum measure is a minimizer.
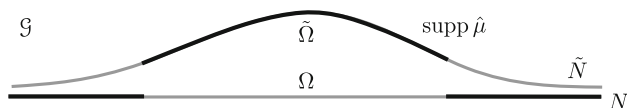
Fig. 1 The measure $\hat{\mu}$

3 A positive nonlinear surface layer integral

3.1 A positivity argument under a volume constraint

We consider two measures: A measure μ which describes the vacuum spacetime, and another measure $\tilde{\mu}$ which typically describes an interacting spacetime. We assume that the vacuum measure is a *minimizer with respect to variations of finite volume* as defined in Sect. 2.1. We choose subsets $\Omega \subset N$ and $\tilde{\Omega} \subset \tilde{N}$ having the same finite volume,

$$\mu(\Omega) = \tilde{\mu}(\tilde{\Omega}) < \infty . \tag{3.1}$$

In order to construct an admissible test measure $\hat{\mu}$, we "cut out" Ω from μ and "glue in" the set $\tilde{\Omega}$, i.e.

$$\hat{\mu} := \chi_{\tilde{\Omega}}\, \tilde{\mu} + \chi_{N\setminus\Omega}\, \mu$$

(see Fig. 1).

The measure $\hat{\mu}$ differs from μ only on a set of finite volume and preserves the volume constraint (see (2.2)).

Therefore, we obtain from (2.4) and (2.3) (with $\tilde{\mu}$ replaced by $\hat{\mu}$) that

$$
\begin{aligned}
0 \leq &\big(\mathcal{S}(\hat{\mu}) - \mathcal{S}(\mu)\big) \\
= &\, 2 \int_{\mathcal{G}} d(\hat{\mu} - \mu)(\mathbf{x}) \int_{N} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) + \int_{\mathcal{G}} d(\hat{\mu} - \mu)(\mathbf{x}) \int_{\mathcal{G}} d(\hat{\mu} - \mu)(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \\
= &\, 2 \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) - 2 \int_{\Omega} d\mu(\mathbf{x}) \int_{N} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \\
&+ \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) - 2 \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\Omega} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \\
&+ \int_{\Omega} d\mu(\mathbf{x}) \int_{\Omega} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \\
= &\, 2 \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N\setminus\Omega} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) - 2 \int_{\Omega} d\mu(\mathbf{x}) \int_{N\setminus\Omega} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \\
&+ \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) - \int_{\Omega} d\mu(\mathbf{x}) \int_{\Omega} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) .
\end{aligned}
$$

Our findings are summarized as follows.

Theorem 3.1 (Positivity argument under volume constraint) *Let μ be a minimizer with respect to variations of finite volume and $\tilde{\mu}$ a measure on $\mathcal{G}$. Moreover, let $\Omega \subset N := \operatorname{supp} \mu$ and $\tilde{\Omega} \subset \tilde{N} := \operatorname{supp} \tilde{\mu}$ satisfying the volume constraint (3.1). Then*

$$0 \leq 2 \int_{\tilde{\Omega}} d\tilde{\mu}(\boldsymbol{x}) \int_{N\setminus\Omega} d\mu(\boldsymbol{y})\, \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}) - 2 \int_{\Omega} d\mu(\boldsymbol{x}) \int_{N\setminus\Omega} d\mu(\boldsymbol{y})\, \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}) \tag{3.2}$$

$$+ \int_{\tilde{\Omega}} d\tilde{\mu}(\boldsymbol{x}) \int_{\tilde{\Omega}} d\tilde{\mu}(\boldsymbol{y})\, \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}) - \int_{\Omega} d\mu(\boldsymbol{x}) \int_{\Omega} d\mu(\boldsymbol{y})\, \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}) . \tag{3.3}$$

The first summand in (3.2) coincides with the first summand in the nonlinear surface layer integral as introduced in [19]. Thus it is a nonlinear surface layer integral with a somewhat different structure. This nonlinear surface layer integral it is not conserved, but instead it satisfies an *in*equality. The second summand in (3.2) can be interpreted as the surface area of $\partial\Omega$. The two summands in (3.3), on the other hand, can be regarded as volume integrals over $\tilde{\Omega}$ and Ω.

It is useful to rewrite this inequality in a more geometric way.

Definition 3.2 (*Area and Volume*) Given a measurable subset $\Omega \subset N$, we define its area A and volume V by

$$A := \int_\Omega d\mu(\mathbf{x}) \int_{N\setminus\Omega} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$V := \mu(\Omega)\,.$$

For the measure $\tilde{\mu}$ we use the same notation with additional tildes.

We now let μ be a minimizer with respect to variations of finite volume. It satisfies the EL equations

$$\ell|_N \equiv \inf_N \ell = 0 \quad\text{with}\quad \ell(\mathbf{x}) := \int_N \mathcal{L}(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y}) - \mathfrak{s}\,.$$

Moreover, we assume that the measure $\tilde{\mu}$ has the property that

$$\tilde{\ell}|_{\tilde{N}} \equiv 0 \tag{3.4}$$

(this assumption will be discussed at the end of this section). Under these assumptions and using these notions, we can reformulate Theorem 3.1 as follows.

Theorem 3.3 *Let μ be a minimizer with respect to variations of finite volume and $\tilde{\mu}$ a measure on $\mathcal{G}$. Moreover, assume that $\tilde{\mu}$ satisfies (3.4). Then for any measurable subsets $\Omega \subset N$ and $\tilde{\Omega} \subset \tilde{N}$ which satisfy the volume constraint*

$$V = \tilde{V}\,, \tag{3.5}$$

the function $\mathfrak{M}(\tilde{\Omega}, \Omega)$ defined by

$$\mathfrak{M}(\tilde{\Omega}, \Omega) := 2 \int_{\tilde{\Omega}} d\tilde{\mu}(x) \int_{N\setminus\Omega} d\mu(y)\, \mathcal{L}(x, y) \;-\; \tilde{A} - A \tag{3.6}$$

is non-negative,

$$\mathfrak{M}(\tilde{\Omega}, \Omega) \geq 0\,. \tag{3.7}$$

In order to clarify the dependence on the measures μ and $\tilde{\mu}$, we sometimes also denote the function $\mathfrak{M}(\tilde{\Omega}, \Omega)$ in (3.6) by $\mathfrak{M}_{\tilde{\mu},\mu}(\tilde{\Omega}, \Omega)$ (as in (1.4) in the introduction).

Proof of Theorem 3.3 We rewrite the last integral in (3.3) as

$$\int_\Omega d\mu(\mathbf{x}) \int_\Omega d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$= \int_\Omega d\mu(\mathbf{x}) \int_N d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) - \int_\Omega d\mu(\mathbf{x}) \int_{N\setminus\Omega} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathfrak{s}V - A\,.$$

Rewriting the first integral in (3.3) similarly and using the volume constraint (3.5) gives the result. □

The assumption (3.4) can be understood as follows. We first point out that this condition follows from the restricted EL equations (2.9). Hence (3.4) holds whenever $\tilde{\mu}$ is a minimizing or critical measure. But (3.4) is much weaker than the restricted EL equations, because it is only the scalar component of these equations. Indeed, this condition can be satisfied for any given measure $\tilde{\mu}$ by changing its weight (i.e. by varying in the class of measures with fixed support supp $\tilde{\mu}$). We refer the interested reader to [29, 32].

We finally comment on the equality case in (3.7).

Remark 3.4 (Rigidity statements) It is a natural question whether the equality $\mathfrak{M}(\tilde{\Omega}, \Omega) = 0$ implies that $\tilde{\mu}|_{\tilde{\Omega}} = \mu|_{\Omega}$. The answer is yes, provided that the minimizer μ is unique. Such a uniqueness statement seems a sensible assumption. However, the complication arises that uniqueness will hold only up to symmetry transformations of the Lagrangian. For conceptual clarity, we decided to postpone such symmetry transformations until later (see Definition 6.1 in Sect. 6). Consequently, for simplicity we decided to leave uniqueness and rigidity statements out of the present paper. We merely remark at this point that, assuming that the measure μ is unique up to symmetry transformations, a rigidity statement should hold stating that if $\mathfrak{M}(\tilde{\Omega}, \Omega)$ vanishes, then the measures $\tilde{\mu}|_{\tilde{\Omega}}$ and $\mu|_{\Omega}$ coincide up to such symmetry transformations. Similarly, if the total mass is zero, then $\tilde{\mu}$ and μ are related to each other by symmetry transformations. An analogous statement should also hold for the quasilocal mass as will be introduced in Sect. 7. ◇

3.2 Generalizations of the positive nonlinear surface layer integral

In this section, we shall get rid of the volume constraint (3.5). Let μ and $\tilde{\mu}$ again be measures as in the statement of Theorem 3.3. Moreover, we now need to specialize our setting by assuming that the total volume of the minimizing measure μ is infinite,

$$\mu(N) = \infty \,. \tag{3.8}$$

We choose subsets $\Omega \subset N$ and $\tilde{\Omega} \subset \tilde{N}$ of finite (but not necessarily the same) volume. The idea for treating the case that the volumes V and $\tilde{V}$ are different is to compensate for this fact by adding to (3.7) the volume difference times the "action per volume." In order to compute this "action per volume" heuristically, we choose a subset $U \subset N$ of finite volume and consider the variation $(\mu_\tau)_{\tau \in [0,1]}$ with

$$\mu_\tau := \mu + \tau \chi_U \, \mu \,.$$

Then the action changes by

$$\frac{d}{d\tau} \big(\mathcal{S}(\mu_\tau) - \mathcal{S}(\mu) \big) \Big|_{\tau=0} = 2 \int_U d\mu(\mathbf{x}) \int_N d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) = 2\mathfrak{s} \, \mu(U) \,,$$

suggesting that the "action per volume" is given simply by $2\mathfrak{s}$. Taking into account the corresponding "volume contribution" gives the following theorem, whose proof also makes the above heuristic argument precise.

Theorem 3.5 *Let μ be a minimizer with respect to variations of finite volume which has infinite total volume (3.8). Moreover, let $\tilde{\mu}$ be a measure on $\mathcal{G}$ which satisfies (3.4). Let $\Omega \subset N$ and $\tilde{\Omega} \subset \tilde{N}$ be subsets of finite volume. Then the function $\mathfrak{M}(\tilde{\Omega}, \Omega)$ defined in (3.6) satisfies the inequality*

$$\mathfrak{M}(\tilde{\Omega}, \Omega) \geq \mathfrak{s} \big(\tilde{V} - V \big) \,.$$

Proof Given $n \in \mathbb{N}$ we choose a subset $U_n \subset N$ with $n < \mu(U_n) < \infty$. We define the measure

$$\hat{\mu} := \chi_{\tilde{\Omega}}\, \tilde{\mu} + \chi_{N \setminus \Omega}\, \mu - \frac{\tilde{V} - V}{\mu(U_n)}\, \chi_{U_n}\, \mu\,. \tag{3.9}$$

Obviously, the measures $\hat{\mu}$ and μ coincide outside the set of finite volume $\Omega \cup \tilde{\Omega} \cup U_n$. Moreover, with the last summand in (3.9) we arranged the volume constraint. Hence

$$0 \le \big(\mathcal{S}(\hat{\mu}) - \mathcal{S}(\mu)\big)$$
$$= 2 \int_{\mathcal{G}} d(\hat{\mu} - \mu)(\mathbf{x}) \int_N d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) + \int_{\mathcal{G}} d(\hat{\mu} - \mu)(\mathbf{x}) \int_N d(\hat{\mu} - \mu)(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y})\,.$$

Substituting (3.9) and multiplying out, we can use transformations similar as in the proof of Theorem 3.3 to obtain

$$0 \le \big(\mathcal{S}(\hat{\mu}) - \mathcal{S}(\mu)\big) = 2 \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) + \mathfrak{s}\tilde{V} - \tilde{A} - \mathfrak{s}V - A \tag{3.10}$$

$$- 2\, \frac{\tilde{V} - V}{\mu(U_n)} \int_{U_n} d\mu(\mathbf{x}) \int_N d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \tag{3.11}$$

$$+ \left(\frac{\tilde{V} - V}{\mu(U_n)}\right)^2 \int_{U_n} d\mu(\mathbf{x}) \int_{U_n} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \tag{3.12}$$

$$+ 2\, \frac{\tilde{V} - V}{\mu(U_n)} \int_{U_n} d\mu(\mathbf{x}) \int_{\Omega} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \tag{3.13}$$

$$- 2\, \frac{\tilde{V} - V}{\mu(U_n)} \int_{U_n} d\mu(\mathbf{x}) \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y})\,. \tag{3.14}$$

Using the EL equations, (3.11) simplifies to

$$(3.11) = -2\mathfrak{s}\left(\tilde{V} - V\right)\,.$$

The terms (3.12)–(3.14), on the other hand, tend to zero as $n \to \infty$, as one sees from the estimates

$$\left|(3.12)\right| \le \left(\frac{\tilde{V} - V}{\mu(U_n)}\right)^2 \int_{U_n} d\mu(\mathbf{x}) \int_N d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \le \mathfrak{s}\, \frac{|\tilde{V} - V|^2}{n}$$

$$\left|(3.13)\right| \le 2\, \frac{|\tilde{V} - V|}{\mu(U_n)} \int_{\Omega} d\mu(\mathbf{x}) \int_N d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \le 2\, \mathfrak{s}V\, \frac{|\tilde{V} - V|}{n}$$

$$\left|(3.14)\right| \le 2\, \frac{|\tilde{V} - V|}{\mu(U_n)} \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \int_N d\mu(\mathbf{x})\, \mathcal{L}(\mathbf{x}, \mathbf{y}) \le 2\, \frac{|\tilde{V} - V|}{n} \sup_{x \in \tilde{\Omega}} \left(\ell(\mathbf{x}) + \mathfrak{s}\right)\,.$$

According to our assumption (iv) on page 4, the function ℓ is bounded on $\mathcal{G}$. Therefore, all the expressions on the right tend to zero as n tends to infinity. □

We finally state a variant of the above theorem which does not require that the interacting measure $\tilde{\mu}$ has the property (3.4).

Proposition 3.6 *Let μ be a minimizer with respect to variations of finite volume which has infinite total volume (3.8). Then for any measure $\tilde{\mu}$ on $\mathcal{G}$ and all subsets $\Omega \subset N$ and $\tilde{\Omega} \subset \tilde{N}$*

of finite volume, the function $\mathfrak{M}(\tilde{\Omega}, \Omega)$ satisfies the lower bound

$$\mathfrak{M}(\tilde{\Omega}, \Omega) \geq 2\mathfrak{s}\tilde{V} - \mathfrak{s}V - \int_{\tilde{\Omega}} d\tilde{\mu}(x) \int_{\tilde{N}} d\tilde{\mu}(y) \, \mathcal{L}(x, y) \, . \tag{3.15}$$

Proof Dropping the relation (3.4), the first integral in (3.3) can be rewritten as

$$\int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) = \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{N}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) - \tilde{A} \, .$$

Hence the inequality (3.7) is modified to

$$2 \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \geq A + \mathfrak{s}V + \tilde{A} - \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{N}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \, .$$

Evaluating this inequality for $\tilde{\mu} = \hat{\mu}$ according to (3.9), we obtain

$$0 \leq \big(\mathcal{S}(\hat{\mu}) - \mathcal{S}(\mu) \big)$$
$$= 2 \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) - A - \mathfrak{s}V - \tilde{A} + \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{N}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$- 2 \frac{\tilde{V} - V}{\mu(U_n)} \int_{U_n} d\mu(\mathbf{x}) \int_{N} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) + \left(\frac{\tilde{V} - V}{\mu(U_n)} \right)^2 \int_{U_n} d\mu(\mathbf{x}) \int_{U_n} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$+ 2 \frac{\tilde{V} - V}{\mu(U_n)} \int_{U_n} d\mu(\mathbf{x}) \int_{\Omega} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) - 2 \frac{\tilde{V} - V}{\mu(U_n)} \int_{U_n} d\mu(\mathbf{x}) \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \, .$$

Now we can take the limit $n \to \infty$ exactly as in the proof of Theorem 3.5 to obtain the result.
□

4 A new proof of the positive mass theorem with volume constraint

We let $(\Omega_n)_{n \in \mathbb{N}}$ and $(\tilde{\Omega}_n)_{n \in \mathbb{N}}$ be exhaustions of N and $\tilde{N}$, respectively.

Theorem 4.1 *Let μ be a minimizer with respect to variations of finite volume which has infinite total volume (3.8). Moreover, let $\tilde{\mu}$ be a measure on $\mathcal{G}$ which has the property (3.4). Assume that $\tilde{\mu}$ is asymptotically flat (see Definition 2.5). Then the total mass (see Definition 2.4) is non-negative,*

$$\mathfrak{M}^{\text{tot}} \geq 0 \, .$$

Proof We let $\tilde{\Omega}_n \subset \tilde{N}$ and $\Omega_n \subset N$ be exhaustions by sets of finite volume. From Theorem 3.5 we know that for any n,

$$0 \leq \mathfrak{M}(\tilde{\Omega}_n, \Omega_n) = 2 \int_{\tilde{\Omega}_n} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega_n} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \; - \; \tilde{A}_n - A_n - \mathfrak{s} \big(\tilde{V}_n - V_n \big) \, .$$

Taking the limit $n \to \infty$, we can use (2.14) and (2.15) to obtain

$$0 \leq \liminf_{n \to \infty} \bigg(2 \int_{\tilde{\Omega}_n} d\mu(\mathbf{x}) \int_{N \setminus \Omega_n} d\mu(\mathbf{y}) \, \nabla_{1, \mathfrak{v}} \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$- \int_{\Omega_n} d\mu(\mathbf{x}) \int_{N \setminus \Omega_n} d\mu(\mathbf{y}) \, \big(\nabla_{1, \mathfrak{v}} + \nabla_{2, \mathfrak{v}} \big) \mathcal{L}(\mathbf{x}, \mathbf{y}) \; - \; \mathfrak{s} \big(\tilde{V}_n - V_n \big) \bigg)$$

$$= \liminf_{n \to \infty} \left(\int_{\tilde{\Omega}_n} d\mu(\mathbf{x}) \int_{N \setminus \Omega_n} d\mu(\mathbf{y}) \left(\nabla_{1,\mathfrak{v}} - \nabla_{2,\mathfrak{v}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) - \mathfrak{s} \left(\tilde{V}_n - V_n \right) \right)$$

$$= \liminf_{n \to \infty} \left(\int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \right.$$

$$\left. - \int_{\Omega} d\mu(\mathbf{x}) \int_{\tilde{N} \setminus \tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) - \mathfrak{s} \left(\tilde{V}_n - V_n \right) \right) = \mathfrak{M}^{\text{tot}} \,.$$

This gives the result.   □

5 A positive mass theorem without volume constraint

5.1 Rewriting surface layer integrals as surface integrals

For the proof of the positive mass theorem, it is essential that the nonlinear surface layer integrals in (1.2) and (1.4) can be linearized near infinity, giving a surface layer integral of the form (2.17) (for details see the proof of Theorem 4.1). From the computational point of view, it is desirable to rewrite such linear surface layer integrals as standard surface integrals over the boundary of Ω. We now give a general procedure for doing so. This procedure will also leads us to a method for treating the volume constraint.

We again assume that μ describes the vacuum. Moreover, from now on we assume that the vacuum is smooth and three-dimensional and that the measure is translation invariant, i.e.

$$N \simeq \mathbb{R}^3 \quad \text{and} \quad d\mu(\mathbf{x}) = d^3\mathbf{x} \,. \tag{5.1}$$

Note that, if we assume again that the Lagrangian is of short range (2.18), in the surface layer integral (2.17), the jet $\mathfrak{v}$ enters only in a δ-neighborhood of $\partial\Omega$. For the following construction, however, we need to extend $\mathfrak{v}$ smoothly to all of N. Then, making use of the fact that the integrand is anti-symmetric, we can rewrite (2.17) as

$$\mathfrak{M}(\Omega, \mathfrak{v}) = \int_{\Omega} d\mu(\mathbf{x}) \int_N d\mu(\mathbf{y}) \left(\nabla_{1,\mathfrak{v}} - \nabla_{2,\mathfrak{v}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \,. \tag{5.2}$$

Given a parameter $s \in [0, 1]$ we introduce the new variables $(\boldsymbol{\zeta}, \boldsymbol{\xi})$ by

$$\boldsymbol{\zeta} = s\mathbf{x} + (1 - s)\,\mathbf{y} \,, \qquad\qquad \boldsymbol{\xi} = \mathbf{y} - \mathbf{x} \,.$$
$$\mathbf{x} = \boldsymbol{\zeta} - (1 - s)\boldsymbol{\xi} \,, \qquad\qquad \mathbf{y} = \boldsymbol{\zeta} + s\boldsymbol{\xi} \,.$$

Note that $\boldsymbol{\zeta}$ is a convex combination of $\mathbf{x}$ and $\mathbf{y}$. In this way, increasing the parameter s from zero to one, one can continuously deform these variables and interchange the roles of $\mathbf{x}$ and $\mathbf{y}$. We now consider the integral

$$\varphi(s, \boldsymbol{\zeta}) := \int_N \nabla_{1,\mathfrak{v}} \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d^3\boldsymbol{\xi} = \int_N \nabla_{1,\mathfrak{v}} \mathcal{L}\big(\boldsymbol{\zeta} - (1 - s)\boldsymbol{\xi}, \boldsymbol{\zeta} + s\boldsymbol{\xi}\big) \, d^3\boldsymbol{\xi} \,.$$

Then

$$\varphi(1, \boldsymbol{\zeta}) = \int_N \nabla_{1,\mathfrak{v}} \mathcal{L}(\boldsymbol{\zeta}, \boldsymbol{\zeta} + \boldsymbol{\xi}) \, d^3\boldsymbol{\xi} = \int_N \nabla_{1,\mathfrak{v}} \mathcal{L}(\boldsymbol{\zeta}, \mathbf{y}) \, d^3\mathbf{y}$$

$$\varphi(0, \boldsymbol{\zeta}) = \int_N \nabla_{1,\mathfrak{v}} \mathcal{L}(\boldsymbol{\zeta} - \boldsymbol{\xi}, \boldsymbol{\zeta}) \, d^3\boldsymbol{\xi} = \int_N \nabla_{1,\mathfrak{v}} \mathcal{L}(\mathbf{x}, \boldsymbol{\zeta}) \, d^3\mathbf{x} = \int_N \nabla_{2,\mathfrak{v}} \mathcal{L}(\boldsymbol{\zeta}, \mathbf{y}) \, d^3\mathbf{y} \,,$$

making it possible to write the inner integral of the surface layer integral (2.17) as the difference

$$\int_N \left(\nabla_{1,\mathfrak{v}} - \nabla_{2,\mathfrak{v}}\right)\mathcal{L}(\zeta, \mathbf{y})\, d^3\mathbf{y} = \varphi(1, \zeta) - \varphi(0, \zeta)\,. \tag{5.3}$$

We next expand the right side in a Taylor series about $s = \frac{1}{2}$. It is a remarkable fact that each term of this Taylor series is a total divergence:

Theorem 5.1 *The inner integral of the surface layer integral (5.2) can be written as a divergence,*

$$\int_N \left(\nabla_{1,\mathfrak{v}} - \nabla_{2,\mathfrak{v}}\right)\mathcal{L}(\zeta, y)\, d^3y = \sum_{\alpha=1}^{3} \frac{\partial}{\partial \zeta_\alpha} A^\alpha(\zeta)\,, \tag{5.4}$$

where the vector field $A(\zeta)$ has the formal power expansion

$$A_\alpha(\zeta) := \sum_{k=0}^{\infty} A_\alpha^{(k)}(\zeta) \quad and \tag{5.5}$$

$$A_\alpha^{(k)}(\zeta) := \frac{1}{4^k\,(2k+1)!} \int_N \xi_\alpha \left(\xi\, \frac{\partial}{\partial \zeta}\right)^{2k} \nabla_{1,\mathfrak{v}}\mathcal{L}\!\left(\zeta - \frac{\xi}{2}, \zeta + \frac{\xi}{2}\right) d^3\xi\,. \tag{5.6}$$

The vector field $A_\alpha(\zeta)$ will be referred to as the *alignment vector field*.

Before coming to the proof of this theorem, we make a two short remarks. We first point out that, similar to the Taylor series of a smooth function, the formal power series (5.5) will in general *not* converge. Nevertheless, the summands for larger k will become small in the sense that they involve higher scaling factors $\delta/\ell_{\mathrm{macro}}$. This will be explained in more detail in the paragraph after Corollary 5.2.

The second remark is to note that the method of this theorem works immediately extends to other general surface layer integrals. More precisely, one may replace the integrand $(\nabla_{1,\mathfrak{v}} - \nabla_{2,\mathfrak{v}})\mathcal{L}(\zeta, y)$ in (5.4) by any other function which is smooth and *anti-symmetric* in its two arguments ζ and y.

Proof of Theorem 5.1 We begin by computing the s-derivatives of the function $\varphi(s, \zeta)$. The first derivative can be calculated by

$$\frac{d}{ds}\varphi(s, \zeta) = \frac{d}{ds} \int_N \nabla_{1,\mathfrak{v}}\mathcal{L}\big(\zeta - (1-s)\xi, \zeta + s\xi\big)\, d^3\xi$$

$$= \int_N \left(D_{1,\xi} + D_{2,\xi}\right)\nabla_{1,\mathfrak{v}}\mathcal{L}\big(\zeta - (1-s)\xi, \zeta + s\xi\big)\, d^3\xi$$

$$= \int_N \sum_{\alpha=1}^{3} \xi^\alpha\, \frac{\partial}{\partial \zeta^\alpha}\nabla_{1,\mathfrak{v}}\mathcal{L}\big(\zeta - (1-s)\xi, \zeta + s\xi\big)\, d^3\xi\,.$$

Computing the higher derivatives iteratively gives

$$\frac{d^p}{ds^p}\varphi(s, \zeta) = \int_N \left(\xi\, \frac{\partial}{\partial \zeta}\right)^p \nabla_{1,\mathfrak{v}}\mathcal{L}\big(\zeta - (1-s)\xi, \zeta + s\xi\big)\, d^3\xi\,.$$

Using these formulas in (5.3) gives

$$\int_N \left(\nabla_{1,\mathfrak{v}} - \nabla_{2,\mathfrak{v}}\right)\mathcal{L}(\zeta, \mathbf{y})\, d^3\mathbf{y} = \sum_{\ell=0}^{\infty} \frac{1}{\ell!}\, \partial_s^\ell \varphi(s, \zeta)\Big|_{s=\frac{1}{2}} \left(\left(\frac{1}{2}\right)^\ell - \left(-\frac{1}{2}\right)^\ell\right)$$

$$= \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} \, \partial_s^{2k+1} \varphi(s, \boldsymbol{\zeta}) \Big|_{s=\frac{1}{2}} \, \frac{2}{2^{2k+1}} = \sum_{k=0}^{\infty} \frac{1}{4^k \, (2k+1)!} \, \partial_s^{2k+1} \varphi(s, \boldsymbol{\zeta}) \Big|_{s=\frac{1}{2}}$$

$$= \sum_{k=0}^{\infty} \frac{1}{4^k \, (2k+1)!} \int_N \left(\boldsymbol{\xi} \, \frac{\partial}{\partial \boldsymbol{\zeta}} \right)^{2k+1} \nabla_{1,\mathfrak{v}} \mathcal{L} \left(\boldsymbol{\zeta} - \frac{\boldsymbol{\xi}}{2}, \boldsymbol{\zeta} + \frac{\boldsymbol{\xi}}{2} \right) d^3 \boldsymbol{\xi} \, .$$

This concludes the proof.   □

Using this expansion in (5.2) we can express the linearized quasilocal mass in as a standard surface integral.

Corollary 5.2 *Assume that $\Omega \subset N \simeq \mathbb{R}^3$ has a smooth boundary with outer normal ν. Then the linearized quasilocal mass (2.17) can be written as*

$$\mathfrak{M}(\Omega, \mathfrak{v}) = \int_{\partial \Omega} \sum_{\alpha=1}^{3} A^{\alpha}(\boldsymbol{x}) \, \nu_{\alpha}(\boldsymbol{x}) \, d\mu_{\partial \Omega}(\boldsymbol{x}) \, .$$

Proof We integrate (5.4) over Ω and use (5.2),

$$\mathfrak{M}(\Omega, \mathfrak{v}) = \int_{\Omega} \sum_{\alpha=1}^{3} \frac{\partial}{\partial \boldsymbol{\zeta}_{\alpha}} A^{\alpha}(\boldsymbol{\zeta}) \, d^3 \boldsymbol{\zeta} \, .$$

Applying the Gauss divergence theorem gives the result.   □

We finally explain in which sense the higher orders of the above expansion are small. Each factor $\boldsymbol{\xi}$ in (5.6) gives a scaling factor δ (where δ denotes again the range of the Lagrangian as introduced after (2.18)). Each such factor comes with a $\boldsymbol{\zeta}$-derivative acting on the jet $\mathfrak{v}$. We assume that each $\boldsymbol{\zeta}$-derivative gives a scaling factor ℓ_{macro}, where ℓ_{macro} can be thought of as the length scale on which the macroscopic physical objects (like currents, fields, curvature) change. Then

$$A^{(k)} \text{involves the scaling factor} \quad \left(\frac{\delta}{\ell_{\text{macro}}} \right)^{2k} . \tag{5.7}$$

Working with corresponding error terms, we may truncate the formal power series (5.5). Indeed, in this paper, it will be sufficient to consider the vector fields $A^{(0)}$ and $A^{(1)}$. We note that the reader who prefers to avoid formal power expansions may replace (5.5) by a Taylor polynomial and estimate the remainder term using the scaling (5.7).

We finally remark that the scalings could be avoided by imposing strict inequalities of the form

$$\sup_{\mathbf{x} \in N} \| \partial^{\kappa} \mathfrak{v}^i(\mathbf{x}) \| \leq \frac{c}{|\ell_{\text{macro}}^{|\kappa|}|} \, ,$$

to be satisfied for all multi-indices κ with $|\kappa| \leq p$ and some maximal order of differentiability p ($p = 2$ seems sufficient for our purposes). Here the index i refers to the components in a chart of $\mathcal{G}$, for example in symmetric wave charts for causal fermion systems as constructed in [21, 28]. Since such inequalities are straightforward but lengthy to state, we do not enter the details but prefer instead to work with scaling factors of the form (5.7).

5.2 Asymptotic alignment

The main difficulty in deriving a positive mass theorem without volume constraint is to control the volume difference $\tilde{V} - V$ in the positivity statement of Theorem 3.5. Here the basic problem is that, at the present stage, for given $\tilde{\Omega}$ the set Ω can be chosen arbitrarily. In order to remove this freedom, one needs to canonically identify $\tilde{N}$ with N near infinity. We now explain how this can be done in the linearized setting (the nonlinear setting will be treated in Sect. 5.3). In the linearized description, the gravitating system is described by a jet $\mathfrak{v}$. The freedom in identifying $\tilde{N}$ with N corresponds to the fact that $\mathfrak{v}$ is determined only up to inner solutions. Thus we have the freedom to transform $\mathfrak{v}$ according to

$$\mathfrak{v} \to \mathfrak{v} + \mathfrak{u}, \tag{5.8}$$

where $\mathfrak{u} = (\operatorname{div} \mathbf{u}, \mathbf{u})$ is an inner solution. This inner solution changes the linearized quasilocal mass, as is worked out in the next lemma (a similar statement in the four-dimensional case is given in [19, Proposition 3.5]).

Lemma 5.3 *For an inner solution* $\mathfrak{u} = (\operatorname{div} \mathbf{u}, \mathbf{u})$*, the linearized quasilocal mass reduces to the flux integral*

$$\mathfrak{M}(\Omega, \mathfrak{u}) = \mathfrak{s} \int_{\partial\Omega} \sum_{\alpha=1}^{3} \mathbf{u}^\alpha(\mathbf{x}) \, \nu_\alpha(\mathbf{x}) \, d\mu_{\partial\Omega}(\mathbf{x}) \, .$$

Proof We compute the linearized quasilocal mass (2.17) for the inner solution $\mathfrak{v} = \mathfrak{u}$, and apply the Gauss divergence theorem,

$$\mathfrak{M}(\Omega, \mathfrak{u}) = \int_\Omega \mu(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \sum_{\alpha=1}^{3} \left(\frac{\partial}{\partial \mathbf{x}_\alpha} \Big(\mathbf{u}^\alpha(\mathbf{x}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \Big) - \frac{\partial}{\partial \mathbf{y}_\alpha} \Big(\mathbf{u}^\alpha(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \Big) \right)$$

$$= \int_{\partial\Omega} d\mu_{\partial\Omega}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \sum_{\alpha=1}^{3} \nu_\alpha(\mathbf{x}) \, \mathbf{u}^\alpha(\mathbf{x}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$+ \int_\Omega d\mu(\mathbf{x}) \int_{\partial\Omega} d\mu_{\partial\Omega}(\mathbf{y}) \sum_{\alpha=1}^{3} \nu_\alpha(\mathbf{y}) \, \mathbf{u}^\alpha(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= \int_{\partial\Omega} d\mu_{\partial\Omega}(\mathbf{x}) \sum_{\alpha=1}^{3} \nu_\alpha(\mathbf{x}) \, \mathbf{u}^\alpha(\mathbf{x}) \int_N d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \, .$$

Carrying out the last integral with the help of (2.5) and (2.6) gives the result. □

This lemma shows that, in order to get a well-defined linearized quasilocal mass, we must fix the freedom to add inner solutions (5.8). To this end, it turns out to be useful to demand that the vector field $A^{(0)}(\zeta)$ defined in (5.6) must vanish.

Definition 5.4 The jet $\mathfrak{v} \in \mathfrak{J}^{\text{lin}}$ is **asymptotically aligned** if there is a compact set $K \subset N$ such that

$$0 = A_\alpha^{(0)}(\zeta) := \int_N \xi_\alpha \, \nabla_{1,\mathfrak{v}} \mathcal{L}\Big(\zeta - \frac{\xi}{2}, \zeta + \frac{\xi}{2}\Big) \, d^3\xi \tag{5.9}$$

for all $\alpha \in \{1, 2, 3\}$ and $\zeta \in N \setminus K$. The corresponding total mass (2.16) is referred to as the **aligned total mass** $\mathfrak{M}^{\text{aligned}}$.

Clearly, we need to verify that this condition can be satisfied and that it determines the inner solution uniquely. In preparation, we compute the alignment vector field for an inner solution $\mathfrak{v} = \mathfrak{u}$. In the computations, we always assume that the Lagrangian on N is *translation invariant*, meaning that $\mathcal{L}(\mathbf{x}, \mathbf{y})$ depends only on the difference vector $\mathbf{y} - \mathbf{x}$. We use the short notation with square brackets,

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathcal{L}[\boldsymbol{\xi}] \qquad \text{for all } \mathbf{x}, \mathbf{y} \in N \simeq \mathbb{R}^3 \text{ and } \boldsymbol{\xi} := \mathbf{y} - \mathbf{x}. \tag{5.10}$$

Moreover, we assume that the Lagrangian in the vacuum is spherically symmetric, i.e.

$$\mathcal{L}[\boldsymbol{\xi}] = \mathcal{L}[\boldsymbol{\xi}'] \qquad \text{for all } \boldsymbol{\xi}, \boldsymbol{\xi}' \in \mathbb{R}^3 \text{ with } |\boldsymbol{\xi}| = |\boldsymbol{\xi}|. \tag{5.11}$$

Lemma 5.5 *Assume that the Lagrangian in the vacuum is translation invariant (5.10) and spherically symmetric (5.11). Let $\boldsymbol{u} = (a := \operatorname{div} \boldsymbol{u}, \boldsymbol{u})$ be an inner solution. Then the corresponding alignment vector field is given by*

$$A_\alpha^{(0)}(\boldsymbol{\zeta}) = \mathfrak{s}\, u_\alpha(\mathbf{x}) + \frac{1}{24}\, \delta^2\, \mathfrak{s}_2\, \Delta_{\mathbb{R}^3} u_\alpha(\mathbf{x}) - \frac{1}{12}\, \delta^2\, \mathfrak{s}_2\, \partial_\alpha a(\mathbf{x}) + \mathcal{O}\left(\frac{\delta^3}{\ell_{\mathrm{macro}}^3}\right), \tag{5.12}$$

where $\Delta_{\mathbb{R}^3}$ denotes the Laplacian in $\mathbb{R}^3$, δ is again the range of the Lagrangian (see below (2.18)) and

$$\mathfrak{s}_2 := \frac{1}{\delta^2} \int_N |\boldsymbol{\xi}|^2 \mathcal{L}[\boldsymbol{\xi}]\, d^3\boldsymbol{\xi}. \tag{5.13}$$

Proof We denote the two arguments of the Lagrangian as usual by $\mathbf{x}$ and $\mathbf{y}$. Thus

$$\boldsymbol{\zeta} = \frac{1}{2}\, (\mathbf{y} + \mathbf{x}), \quad \boldsymbol{\xi} = \mathbf{y} - \mathbf{x}$$

$$\mathbf{x} = \boldsymbol{\zeta} - \frac{\boldsymbol{\xi}}{2}, \qquad \mathbf{y} = \boldsymbol{\zeta} + \frac{\boldsymbol{\xi}}{2}.$$

Then the partial derivatives transform according to

$$\frac{\partial}{\partial \boldsymbol{\xi}_\alpha} = \frac{1}{2}\left(\frac{\partial}{\partial \mathbf{y}_\alpha} - \frac{\partial}{\partial \mathbf{x}_\alpha}\right), \quad \frac{\partial}{\partial \boldsymbol{\zeta}_\alpha} = \frac{\partial}{\partial \mathbf{y}_\alpha} + \frac{\partial}{\partial \mathbf{x}_\alpha}.$$

We also make use of the fact that the vacuum Lagrangian depends only on the difference vector $\boldsymbol{\xi}$; we write

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathcal{L}[\boldsymbol{\xi}].$$

This implies that

$$\frac{\partial}{\partial \mathbf{x}_\alpha} \mathcal{L}(\mathbf{x}, \mathbf{y}) = -\frac{\partial}{\partial \boldsymbol{\xi}_\alpha} \mathcal{L}[\xi] \qquad \text{and} \qquad \frac{\partial}{\partial \mathbf{y}_\alpha} \mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{\partial}{\partial \boldsymbol{\xi}_\alpha} \mathcal{L}[\xi].$$

Using these relations, we obtain

$$A^{(0)}\alpha(\boldsymbol{\zeta}) = \int_N \boldsymbol{\xi}_\alpha\, a(\mathbf{x})\, \mathcal{L}[\xi]\, d^3\boldsymbol{\xi} + \int_N \boldsymbol{\xi}_\alpha\, \mathbf{u}_\beta(\mathbf{x})\, \frac{\partial}{\partial \mathbf{x}_\beta} \mathcal{L}[\xi]\, d^3\boldsymbol{\xi}$$

$$= \int_N \boldsymbol{\xi}_\alpha\, a(\mathbf{x})\, \mathcal{L}[\xi]\, d^3\boldsymbol{\xi} - \int_N \boldsymbol{\xi}_\alpha\, \mathbf{u}_\beta(\mathbf{x})\, \frac{\partial}{\partial \boldsymbol{\xi}_\beta} \mathcal{L}[\xi]\, d^3\boldsymbol{\xi}.$$

Integrating by parts gives

$$A^{(0)}\alpha(\boldsymbol{\zeta}) = \int_N \left(\boldsymbol{\xi}_\alpha\, a(\mathbf{x}) + \mathbf{u}_\alpha(\mathbf{x})\right) \mathcal{L}[\xi]\, d^3\boldsymbol{\xi} + \int_N \boldsymbol{\xi}_\alpha\, \frac{\partial \mathbf{u}_\beta(\mathbf{x})}{\partial \boldsymbol{\xi}_\beta}\, \mathcal{L}[\xi]\, d^3\boldsymbol{\xi}$$

$$= \int_N \Big(\boldsymbol{\xi}_\alpha \, a(\mathbf{x}) + \mathbf{u}_\alpha(\mathbf{x}) \Big) \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\boldsymbol{\xi} - \frac{1}{2} \int_N \boldsymbol{\xi}_\alpha \, \mathrm{div} \, \mathbf{u}(\mathbf{x}) \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\boldsymbol{\xi}$$

$$= \int_N \Big(\mathbf{u}_\alpha(\mathbf{x}) + \boldsymbol{\xi}_\alpha \, \frac{a(\mathbf{x})}{2} \Big) \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\boldsymbol{\xi} \; .$$

We next expand a and $\mathfrak{u}_\alpha$ in a Taylor series about $\boldsymbol{\zeta}$. This gives

$$A^{(0)}{}_\alpha(\boldsymbol{\zeta}) = \int_N \Big(\mathbf{u}_\alpha(\boldsymbol{\zeta}) - \frac{1}{2} \, \boldsymbol{\xi}^\beta \partial_\beta \mathbf{u}_\alpha(\mathbf{x}) + \frac{1}{8} \, \boldsymbol{\xi}^\beta \boldsymbol{\xi}^\gamma \, \partial_{\beta\gamma} \mathbf{u}_\alpha(\mathbf{x}) \Big) \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\boldsymbol{\xi}$$

$$+ \int_N \Big(\frac{1}{2} \, \boldsymbol{\xi}_\alpha \, a(\boldsymbol{\zeta}) - \frac{1}{4} \, \boldsymbol{\xi}_\alpha \boldsymbol{\xi}^\beta \, \partial_\beta a(\boldsymbol{\zeta}) \Big) \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\boldsymbol{\xi} + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\mathrm{macro}}^3} \Big) \, . \qquad (5.14)$$

Now we can use spherical symmetry of the unperturbed Lagrangian to obtain the formulas

$$\int_N \boldsymbol{\xi}_\alpha \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\mathbf{y} = 0$$

$$\int_N \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\beta \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\mathbf{y} = \frac{1}{3} \, \delta_{\alpha\beta} \int_N |\boldsymbol{\xi}|^2 \mathcal{L}[\boldsymbol{\xi}] \, d^3\boldsymbol{\xi} \stackrel{(5.13)}{=} \frac{1}{3} \, \delta^2 \, \mathfrak{s}_2 \, \delta_{\alpha\beta} \, . \qquad (5.15)$$

Using these relations in (5.14) gives the result. $\qquad\qquad\square$

This lemma immediately shows that asymptotic alignment can be arranged by a unique inner solution: As explained after (5.7), the first summand in (5.12) dominates all the other terms. Then the inhomogeneous equation $A^{(0)}[\mathfrak{u}] = -A^{(0)}[\mathfrak{v}]$ reduces to the local equation

$$\mathfrak{u}(x) = -\frac{1}{s} \, A^{(0)}[\mathfrak{v}] \, ,$$

which clearly has a unique solution. The errors can be treated with an iteration argument.

We remark that, in the present asymptotic expansion, the errors could be quantified in terms of decay properties of the derivatives of $\mathfrak{u}$ like $\|D^\kappa \mathfrak{u}(\mathbf{x})\| = \mathcal{O}(|\mathbf{x}|^{-|\kappa|})$. Then the errors could be made arbitrarily small by choosing the compact set K sufficiently large. The iteration argument could be carried out with the help of Banach's fixed point theorem. For brevity, we do not enter the details of these standard constructions.

5.3 Local alignments and the local volume condition

We now return to the nonlinear setting with two measures $\tilde{\mu}$ and μ. Linearizing near infinity and choosing asymptotic alignment, we get a canonical identification of $\tilde{N}$ and N in the asymptotic end. This also gives a unique identification of $\tilde{\Omega}$ with Ω. The next question is how to control the difference of inner volumes $\tilde{V} - V$. In order to analyze this question, we assume that $\tilde{\mu}$ can be obtained from μ by a *smooth deformation* (f_τ, F_τ) with $\tau \in [0, 1]$ and $f_\tau \in C^\infty(N, \mathbb{R})$ and $F_\tau \in C^\infty(N, \mathcal{G})$, i.e.

$$(f_0, F_0) = \big(1_N, \mathrm{id}_N\big) \quad \text{and} \quad \tilde{\mu} = (F_1)_* \big(f_1 \, \mu\big) \qquad (5.16)$$

(where 1_N is the constant function one and id_N is the identity map on N). Before entering the construction, we point out that the smoothness of F_τ poses restrictions for the topology of $\tilde{N}$. Indeed, in the typical situation that F_τ is a diffeomorphism to its image, the topology of $\tilde{N}$ must be trivial (interesting physical situations like a topology change or an event horizon will be considered in Sect. 8).

Given $\tau \in [0, 1]$, we denote the corresponding space by $\tilde{N}_\tau := \operatorname{supp} \tilde{\mu}_\tau$. The infinitesimal change of the measure is described by the jet

$$\mathfrak{v}_\tau := \frac{d}{d\tau}\big(f_\tau, F_\tau\big) \in \mathfrak{J}_\tau^{\text{vary}} \,.$$

Given $\boldsymbol{\zeta} \in \tilde{N}_\tau$, we assume that there is a canonical family of coordinate systems around $\boldsymbol{\zeta}$, with are centered in the sense that $\mathbf{x}^\alpha(\boldsymbol{\zeta}) = 0$ and are unique up to linear transformations

$$\mathbf{x}^\alpha \to A_\beta^\alpha \, \mathbf{x}^\beta \tag{5.17}$$

(with a real 3×3-matrix A). Clearly, in the vacuum $\tau = 0$, one takes the linear coordinates of $N \simeq \mathbb{R}^3$. If $\tilde{N}_\tau$ comes from a causal fermion system constructed in a Riemannian or Lorentzian spacetime (as in the examples in Sect. 10 and Appendix A), then one can choose the coordinates obtained from the exponential map. Alternatively, for causal fermion systems a distinguished family of coordinates is obtained by restricting the so-called symmetric wave charts [21, 28], being charts on $\mathcal{G}$, to N. More generally, for causal variational principles one can work with the Gaussian-type coordinates as constructed in [6]. Here we do not need to specify how precisely the coordinate systems are chosen. Instead, we simply assume that we are given canonical centered charts around $\boldsymbol{\zeta}$, which are unique up to linear transformations (5.17).

Working in these coordinates, similar to (5.9) we impose the *local alignment condition*

$$\int_N \boldsymbol{\xi}_\alpha \, \nabla_{1,\mathfrak{v}_\tau} \mathcal{L}\Big(\boldsymbol{\zeta} - \frac{\boldsymbol{\xi}}{2}, \boldsymbol{\zeta} + \frac{\boldsymbol{\xi}}{2}\Big) \, d\tilde{\mu}_\tau(\boldsymbol{\xi}) = 0 \,. \tag{5.18}$$

Exactly as explained for asymptotic alignment, this condition can be satisfied by adding to $\mathfrak{v}_\tau$ a unique inner solution $\mathfrak{u}_\tau$. For clarity, we denote the jet constructed in this way from coordinates around a point $\mathbf{x} \in \tilde{N}_\tau$ by $\mathfrak{v}_{\tau,\mathbf{x}}$.

Having carried out this construction for every $\mathbf{x} \in \tilde{N}_\tau$, we obtain a unique global vector field $\mathbf{v}_\tau$ by setting

$$\hat{\mathbf{v}}_\tau(\mathbf{x}) = \mathbf{v}_{\tau,\mathbf{x}}(\mathbf{x}) \,.$$

We assume that this vector field is again smooth. We choose the scalar component $\hat{a}_\tau$ such that the corresponding jet $\hat{\mathfrak{v}}_\tau = (\hat{a}_\tau, \hat{\mathbf{v}}_\tau)$ describes the variation of measures. More precisely, we choose

$$\hat{a}_\tau(\mathbf{x}) = a_{\tau,\mathbf{y}}(\mathbf{x}) + \operatorname{div} \hat{\mathbf{v}}_\tau(\mathbf{x}) - \operatorname{div} \mathbf{v}_{\tau,\mathbf{y}}(\mathbf{x})$$

for an arbitrary base point $\mathbf{y} \in \tilde{N}_\tau$. We refer to $\hat{\mathfrak{v}}_\tau$ as the *locally aligned jet*. We note for clarity that the locally aligned jet will in general *not* satisfy the local alignment condition (i.e., the equation obtained by replacing $\mathfrak{v}_\tau$ in (5.18) by $\hat{\mathfrak{v}}_\tau$ will be violated by an error term). But it is asymptotically aligned as in Definition 5.4 (because the error term decays at infinity).

Performing this construction for every $\tau \in [0, 1]$, we obtain a family of jets $(\hat{\mathfrak{v}}_\tau)_{\tau \in [0,1]}$ which define a unique flow $(\hat{f}_\tau, \hat{F}_\tau)_{\tau \in [0,1]}$ via

$$\frac{d}{d\tau}\big(\hat{f}_\tau, \hat{F}_\tau\big) = \hat{\mathfrak{v}}_\tau \,.$$

By construction, the resulting flow of measures coincides with the original flow, i.e.

$$(\hat{F}_\tau)_*\big(\hat{f}_\tau \, \mu\big) = (F_\tau)_*\big(f_\tau \, \mu\big) \qquad \text{for all } \tau \in [0, 1]$$

(but the flows differ by an inner solution for every τ).

We now demand that this flow increases the volume:

Definition 5.6 The smooth deformation (f_τ, F_τ) satisfies the **local volume condition** if the locally aligned jets $\hat{\mathfrak{v}}_\tau = (\hat{a}_\tau, \hat{\mathbf{v}}_\tau)$ have a non-negative scalar component, i.e.

$$\hat{a}_\tau(\mathbf{x}) \geq 0 \qquad for\ all\ \tau \in [0, 1]\ and\ \mathbf{x} \in \tilde{N}_\tau\ . \tag{5.19}$$

We now explain what this condition means and how it can be verified in the applications. The smooth deformation is obtained typically by continuously bringing matter fields into the system. For example, in the setting of causal fermion systems one can introduce an additional physical wave function ψ and can continuously increase its amplitude. The local volume condition states that the aligned jets only increase the volume. This condition can be understood similar to a local energy condition. Instead of stating it for the energy-momentum tensor (which does not exist at the present level of generality), it is a statement on the volume change when introducing matter. We point out that the local volume condition can be verified at a point $\mathbf{x}_0$ by performing constructions in a neighborhood U of this point, provided that this neighborhood is much larger than the range δ of the Lagrangian (here we assume again that the Lagrangian is of short range as introduced before (2.18)). More precisely, choosing Gaussian-type coordinates around every $\mathbf{x} \in U$ and satisfying the local alignment condition (5.18) at $\mathbf{x}$, we obtain a vector field $\hat{\mathbf{v}}$, and computing its divergence at $\mathbf{x}_0$ allows us to compute $\hat{a}(\mathbf{x}_0)$ in (5.19). In this way, similar to the usual local energy conditions, the local volume condition can be verified locally.

We finally remark that, under additional assumptions on the smooth deformation (f_τ, F_τ), the local volume condition can be stated in an alternative form which may be of advantage for computations. We first state this alternative formulation and explain if afterward.

Proposition 5.7 *Assume that all the measures $\tilde{\mu}_\tau := (F_\tau)_*(f_\tau \mu)$ are critical in the sense that they satisfy the restricted EL equations (2.7). Then the local volume condition in Definition 5.6 can be stated equivalently by demanding that the divergence of the vector field $A_\tau(\boldsymbol{\zeta})$ in (5.5) is non-negative,*

$$\operatorname{div} A_\tau(\boldsymbol{\zeta}) \geq 0 \qquad for\ all\ \tau \in [0, 1]\ and\ \boldsymbol{\zeta} \in \tilde{N}_\tau\ .$$

Proof For any $\tau \in [0, 1]$ and $\mathbf{x}_\tau \in \tilde{N}_\tau$, the restricted EL equations imply that

$$0 = \nabla_{\hat{\mathfrak{v}}_\tau} \ell(\mathbf{x}_\tau) = \int_{\tilde{N}_\tau} \nabla_{1,\hat{\mathfrak{v}}} \mathcal{L}(\mathbf{x}_\tau, \mathbf{y})\, d\tilde{\mu}_\tau(\mathbf{y}) - (\nabla_{\hat{\mathfrak{v}}} \mathfrak{s})(\mathbf{x})$$

with $\mathbf{x}_\tau := F_\tau(\mathbf{x})$. On the other hand, differentiating the scalar component of the restricted EL equations with respect to τ gives

$$0 = \frac{d}{d\tau} \ell(F_\tau(x)) = \int_{\tilde{N}_\tau} (\nabla_{1,\hat{\mathfrak{v}}} + \nabla_{2,\hat{\mathfrak{v}}}) \mathcal{L}(\mathbf{x}_\tau, \mathbf{y})\, d\tilde{\mu}_\tau(\mathbf{y}) - (\nabla_{\hat{\mathfrak{v}}} \mathfrak{s})(\mathbf{x})\ .$$

Using these equations, the infinitesimal volume change (5.19) can be written as

$$\hat{a}_\tau(\mathbf{x}) = (\nabla_{\hat{\mathfrak{v}}} \mathfrak{s})(\mathbf{x}) = \int_{\tilde{N}_\tau} (\nabla_{1,\hat{\mathfrak{v}}_\tau} - \nabla_{2,\hat{\mathfrak{v}}_\tau}) \mathcal{L}(\mathbf{x}, \mathbf{y})\, d\tilde{\mu}_\tau(\mathbf{y}) = \operatorname{div} A_\tau(\mathbf{x})\ ,$$

where in the last step we applied Theorem 5.1. $\qquad\square$

This result can be used in computations as follows. First, in order to arrange that the measures $\tilde{\mu}_\tau$ are all critical, one can solve the inhomogeneous linearized field equations for any τ for example using the methods developed in [29]. Once this has been accomplished, one can compute the vector field $A_\tau(\boldsymbol{\zeta})$ in (5.5) term by term by evaluating the integrals (5.6). As explained after (5.7), this gives an expansion in powers of the range δ of the Lagrangian,

making it possible to truncate the series. We now illustrate this procedure for the contribution $\sim \delta^2$ (as we shall see in the examples of Sect. 10 and Appendix A, this is indeed the contribution which gives the correspondence to the ADM mass and the Brown–York mass). To this end, we expand according to

$$\text{div } A_\tau(\zeta) = \text{div } A_\tau^{(0)}(\zeta) + \text{div } A_\tau^{(1)}(\zeta) + \mathcal{O}(\delta^4). \tag{5.20}$$

In order to get rid of the first summand, we arrange that the local alignment condition (5.18) holds, i.e.

$$0 = A_\tau^{(0)}(\zeta) := \int_N \xi_\alpha \, \nabla_{1, \hat{\mathfrak{v}}_\tau} \mathcal{L}\left(\zeta - \frac{\xi}{2}, \zeta + \frac{\xi}{2}\right) d\tilde{\mu}_\tau(\xi) = 0. \tag{5.21}$$

Then the local volume condition can be stated as

$$\text{div } A_\tau^{(1)}(\zeta) \geq 0 + \mathcal{O}(\delta^4) \qquad \text{for all } \tau \in [0, 1] \text{ and } \zeta \in \hat{N}_\tau. \tag{5.22}$$

In the above computation it is very convenient that in (5.21) we may allow for an error term of the order

$$A_\tau^{(0)}(\zeta) = \mathcal{O}(\delta^2). \tag{5.23}$$

Indeed, this error determines the inner solutions up to error terms of order $\mathcal{O}(\delta^2)$. This error of the inner solutions gives rise to an error in $A^{(1)}$ which is of the order $\mathcal{O}(\delta^4)$ and can thus be absorbed into the error term in (5.22). Therefore, the condition (5.22) is unaffected by the error in (5.23) (note that this is not apparent when analyzing the error terms in the expansion (5.20)).

5.4 Statement and proof of the positive mass theorem

We are now in the position for stating and proving the positive mass theorem without volume constraint.

Theorem 5.8 (Positive mass theorem for aligned total mass) *Let μ be a minimizer with respect to variations of finite volume which has infinite total volume (3.8). Moreover, let $\tilde{\mu}$ be a measure on $\mathcal{G}$ which has the property (3.4) and is asymptotically flat (see Definition 2.5). Moreover, assume that $\tilde{\mu}$ can be obtained from μ by a smooth deformation (f_τ, F_τ) which satisfies the local volume condition (see Definition 5.6). Then the aligned total mass is non-negative,*

$$\mathfrak{M}^{\text{aligned}} \geq 0.$$

Proof The resulting variation $(\hat{f}_\tau, \hat{F}_\tau)$ increases the volume and is asymptotically aligned. Therefore, applying Theorem 3.5 choosing $\tilde{\Omega} = \hat{F}_1(\Omega)$, we obtain

$$\mathfrak{M} = \lim_{\Omega \nearrow N} \mathfrak{M}(\tilde{\Omega}, \Omega) \geq \mathfrak{s}(\tilde{V} - V) \geq 0.$$

This concludes the proof. □

Computing this aligned total mass in the Schwarzschild geometry, one finds that this mass coincides, up to a prefactor, with the ADM mass M.

Theorem 5.9 *Identifying the asymptotic ends with linear alignment, the total mass in the Schwarzschild geometry is given by*

$$\mathfrak{M} = \frac{\pi}{9}\, \delta^2 \, \mathfrak{s}_2 \, M \;. \tag{5.24}$$

The proof is given in Appendix A (see page 46).

6 An equivariant positive mass theorem

In the context of causal fermion systems, the two causal fermion systems describing the vacuum and the gravitating system are defined on two different Hilbert spaces. Before defining the mass, one needs to unitarily identify these Hilbert spaces. The fact that this identification is not canonical, gives rise to the freedom to unitarily transform these measures (this point, which is of central importance in getting the connection to quantum field theory, is described in detail in [20, Section 3]). We now generalize this unitary freedom to the setting of causal variational principles, where it corresponds to the freedom in performing isometries of the Lagrangian. This freedom will also be important in our definitions of the quasilocal mass and synthetic scalar curvature in Sects. 7 and 9.

The following notion has been introduced similarly in [24, Section 3.1].

Definition 6.1 A diffeomorphism $\Phi \in C^\infty(\mathcal{G}, \mathcal{G})$ describes a **symmetry of the Lagrangian** if

$$\mathcal{L}\big(\Phi(\mathbf{x}), \Phi(\mathbf{y})\big) = \mathcal{L}(\mathbf{x}, \mathbf{y}) \qquad for\ all\ \mathbf{x}, \mathbf{y} \in \mathcal{G} \;.$$

Such diffeomorphisms clearly form a group, denoted by $\mathcal{G}$, the *group of symmetries of the Lagrangian*. Clearly, transforming the vacuum measure by a diffeomorphism $\Phi \in \mathcal{G}$ does not leaves the causal action and the Euler–Lagrange equations unchanged. In particular,

$$\int_\mathcal{G} \mathcal{L}\big(\Phi(\mathbf{x}), \mathbf{y}\big)\, d\big(\Phi_*\mu\big)(\mathbf{y}) = \int_\mathcal{G} \mathcal{L}(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y}) \qquad \text{for all } \mathbf{x} \in \mathcal{G} \;.$$

In order to clarify the dependence on the measures, following the notation (1.4) in the introduction we also denote the aligned total mass (see Definition 5.4) by $\mathfrak{M}^{\text{aligned}}_{\tilde{\mu}, \mu}$.

Definition 6.2 The measure $\tilde{\mu}$ is **asymptotically deformation-admissible** w.r.t. μ if it is asymptotically flat w.r.t. μ and if $\tilde{\mu}$ can be obtained from μ by a smooth deformation (f_τ, F_τ) which satisfies the local volume condition (see Definition 5.6).

The **equivariant aligned total mass** is defined by

$$\mathfrak{M}^{\text{equi}} := \inf \left\{ \mathfrak{M}^{\text{aligned}}_{\tilde{\mu}, \Phi_*\mu} \,\bigg|\, \Phi \in \mathcal{G}\ with \right.$$

$$\left. \tilde{\mu}\ asymptotically\ deformation - admissible\ w.r.t.\ \Phi_*\mu \right\} \;.$$

Theorem 6.3 *Under the assumptions of Theorem 5.8, also the equivariant aligned total mass is non-negative,*

$$\mathfrak{M}^{\text{equi}} \geq 0 \;.$$

Proof Theorem 5.8 implies that $\mathfrak{M}^{\text{aligned}}_{\tilde{\mu}, \Phi_*\mu} \geq 0$ for all $\tilde{\mu}$ which are asymptotically deformation-admissible w.r.t. $\Phi_*\mu$. Therefore, the infimum of all these aligned masses is again non-negative. $\qquad \square$

7 A positive quasilocal mass

Given an open subset $\tilde{\Omega} \subset \tilde{N}$ of the gravitating system, the goal of this section is to introduce its quasilocal mass $\mathfrak{M}(\tilde{\Omega})$. This quasilocal mass will be non-negative, and it will bound the total mass from below. In preparation, we need to think about how to "localize" the above notions to the subset $\tilde{\Omega}$. Ideally, the quasilocal mass should depend only on $\tilde{\mu}$ restricted to $\tilde{\Omega}$, but not on the "ambient space" $\tilde{N} \backslash \tilde{\Omega}$. This requirement is indeed satisfied for the functional $\mathfrak{M}(\tilde{\Omega}, \Omega)$ in (3.6), as becomes obvious when rewriting it as

$$\mathfrak{M}(\tilde{\Omega}, \Omega) = 2 \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \backslash \Omega} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$+ \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) - A - \mathfrak{s}\tilde{V} \tag{7.1}$$

Likewise, this requirement is also satisfied for the functional in Proposition 3.6, which we now denote and rewrite as follows,

$$\mathfrak{N}(\tilde{\Omega}, \Omega) := \mathfrak{M}(\tilde{\Omega}, \Omega) \, - \, 2\mathfrak{s}\tilde{V} + \mathfrak{s}V + \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{N}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \tag{7.2}$$

$$= 2 \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \backslash \Omega} d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$+ \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \, - \, A - 2\mathfrak{s}\tilde{V} + \mathfrak{s}V \tag{7.3}$$

(note that in the derivation of (7.1) we used (3.4), whereas in (7.3) we did not). However, in order to satisfy the local alignment condition (5.18) for $\boldsymbol{\zeta} \in \tilde{\Omega}$, we need to integrate over all $\boldsymbol{\xi}$ for which the integrand is non-zero. This will in general make it necessary to evaluate the Lagrangian for arguments $\boldsymbol{\zeta} \pm \boldsymbol{\xi}/2$ which lie outside $\tilde{\Omega}$. This is the reason why our notion of quasilocal mass will not depend only on $\tilde{\Omega}$, but on a neighborhood $\tilde{K}$ of $\tilde{\Omega}$. If we assumed that the Lagrangian were of short range (2.18), then this neighborhood could be chosen as the δ-neighborhood $\tilde{K} = \overline{B_\delta(\tilde{\Omega})}$. More generally, the "width of the boundary layer $\tilde{K} \setminus \tilde{\Omega}$ should be large on the scale δ." In order to make this statement mathematically precise, we adapt the notion of *compact range* introduced in [27, Def. 3.3] to our setting.

Definition 7.1 The set $\tilde{\Omega}$ can be realized by a **local smooth deformation of** Ω **inside** K if there is a pair of mappings $(f, F) : [0, 1] \times N \to \mathbb{R}_0^+ \times \mathcal{G}$ with the following properties:

(i) Both mappings f and F are measurable. Moreover, their restrictions to K are smooth, i.e.

$$f|_{[0,1] \times K} \in C^\infty([0, 1] \times K, \mathbb{R} +_0), \qquad F|_{[0,1] \times K} \in C^\infty([0, 1] \times K, \mathcal{G}).$$

(ii) The mappings f and F have the correct boundary values at $\tau = 0$ and $\tau = 1$ inside K; i.e., weakening the conditions (5.16),

$$f_0 = 1_N, \quad F_0 = \mathrm{id}_N$$
$$F_1(N) \subset \tilde{N} \quad and \quad (F_1)_*(f_1 \, \mu|_K) = \tilde{\mu}|_{F_1(K)}.$$

(iii) Setting $N_\tau = F_\tau(N)$, $K_\tau := F_\tau(K)$ and $\Omega_\tau := F_\tau(\Omega)$, for any $\tau \in [0, 1]$ and $\boldsymbol{\zeta} \in \Omega_\tau$ we again choose the distinguished coordinates centered at $\boldsymbol{\zeta}$ unique up to linear transformations (5.17). We assume that, in these coordinates, for every $\mathbf{y} \in N_\tau \setminus \Omega_\tau$ and every $\mathbf{x}_0 \in N$ with $F_\tau(\mathbf{x}_0) = 2\boldsymbol{\zeta} - \mathbf{y}$, there is $\varepsilon > 0$ such that

$$\mathcal{L}(F_{\tau'}(\mathbf{x}_0), \mathbf{y}) = 0 \qquad \text{for all } \tau' \in [0, 1] \text{ with } |\tau' - \tau| < \varepsilon.$$

The last condition ensures that, for any $\zeta \in \Omega_\tau$, the integrand in (5.18) is well-defined and vanishes whenever one of the arguments $\zeta \pm \xi/2$ lies outside K_τ. Therefore, the local alignment condition (5.18) is well-defined for any $\zeta \in \Omega_\tau$. We again consider the resulting flow $(\hat{f}_\tau, \hat{F}_\tau)$ and choose $\Omega \subset N$ such that

$$\tilde{\Omega} = \hat{F}_1(\Omega) \,.$$

We assume that the local volume condition (see Definition 5.6) is satisfied on Ω, i.e.

$$\hat{a}_\tau(\mathbf{x}) \geq 0 \qquad \text{for all } \tau \in [0, 1] \text{ and } \mathbf{x} \in \hat{F}_\tau(\Omega) \,. \tag{7.4}$$

Then the volume difference $\tilde{V} - V$ is positive. Therefore, Theorem 3.5 yields that

$$\mathfrak{M}(\tilde{\Omega}, \Omega) \geq 0 \,.$$

Following the procedure for the equivariant mass, we define the quasilocal mass by taking the infimum over the isometries of the Lagrangian.

Definition 7.2 (*Quasilocal mass*) Let μ be a minimizer with respect to variations of finite volume which has infinite total volume (3.8). Moreover, let $\tilde{\mu}$ be a measure on $\mathcal{G}$ which has the property (3.4). The subset $\tilde{\Omega} \subset \tilde{N}$ is **deformation-admissible** w.r.t. μ for Ω if $\tilde{\Omega}$ can be realized by a local deformation of Ω inside K (see Definition 7.1), and if the resulting deformation $(\hat{f}, \hat{F})$ satisfies the local volume condition (7.4). Given a subset $\tilde{\Omega} \subset \tilde{N}$, we define the **quasilocal mass** by

$$\mathfrak{M}(\tilde{\Omega}) := \inf \Big\{ \mathfrak{M}_{\tilde{\mu}, \Phi_* \mu}(\tilde{\Omega}, \Omega) \,\Big|\, \Phi \in \mathcal{G}, \Omega \subset \Phi(N) \text{ with } \tilde{\Omega} \text{ is}$$
$$\text{deformation} - \text{admissible w.r.t. } \Phi_* \mu \text{ for} \Omega \Big\} \tag{7.5}$$

(where for the function $\mathfrak{M}(.,.)$ we used again the notation with subscripts (1.4)).

Theorem 7.3 *The quasilocal mass is non-negative,*

$$\mathfrak{M}(\tilde{\Omega}) \geq 0 \,.$$

Proof Follows exactly as the proofs of Theorems 5.8 and 6.3. $\qquad\square$

8 Bounding the total mass by the quasilocal mass

In this section we prove that the total mass is at least as large as the quasilocal mass.

Theorem 8.1 *Let μ be a minimizer with respect to variations of finite volume which has infinite total volume (3.8). Moreover, let $\tilde{\mu}$ be a measure on $\mathcal{G}$ which has the property (3.4) and is asymptotically flat (see Definition 2.5). Given sets $U \subset N$ and $\tilde{U} \subset \tilde{N}$ we assume that $\tilde{N} \backslash \tilde{\Omega}$ can be realized by a local smooth deformation of $N \setminus \Omega$ inside a set $N \setminus K$ (see Definition 7.1) which satisfies the local volume condition (see Definition 5.6). Then the* **quasilocal mass** *of the subset $\tilde{U} \subset \tilde{N}$ bounds the equivariant total mass by*

$$\mathfrak{M}^{\text{equi}} \geq \mathfrak{M}(\tilde{U}) \,.$$

Proof We choose a large relatively compact set $\tilde{K} \subset \tilde{N}$ and apply Theorem 3.5 for $\tilde{\Omega}$ replaced by $\tilde{K} \setminus \tilde{\Omega}$. (see Fig. 2). $\qquad\square$
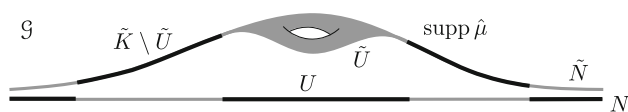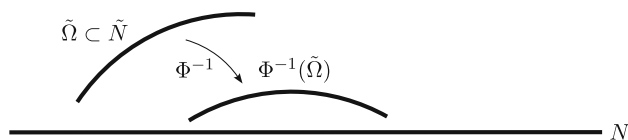
Fig. 2 The measure $\hat{\mu}$

Fig. 3 Isometric transformation of $\tilde{\Omega}$

We point out that this theorem also applies in cases when $\tilde{N}$ has a *non-trivial topology*. All we need is that the exterior region $\tilde{N} \setminus \tilde{U}$ is diffeomorphic to $\mathbb{R}^3$ minus a ball. In other words, the non-trivial topology must be included in the inner region $\tilde{U}$, as is indicated by the handle in Fig. 2. Our theorem also applies in the case that the inner region $\tilde{U}$ contains an event horizon. In this case, the resulting inequalities are reminiscent and should be closely related to the Riemannian Penrose inequality [3, 34]. We plan to investigate this connection in a separate publication.

9 A synthetic definition of scalar curvature

In this section we explore the inequality in Theorem 3.5 in the case that $\tilde{\Omega}$ and Ω are small neighborhoods of given points $\tilde{\mathbf{x}} \in \tilde{N}$ and $\mathbf{x} \in N$. In this case, the linearized description applies not only near the boundary of $\tilde{\Omega}$ but in all of $\tilde{\Omega}$. Therefore, the quasilocal mass inside the curly brackets in (7.5) simplifies to

$$
\mathfrak{M}_{\tilde{\mu}, \Phi_*\mu}\big(\tilde{\Omega}, \Phi\Omega\big) = \int_\Omega d\mu(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, (\nabla_{1,\mathfrak{v}} - \nabla_{2,\mathfrak{v}}) \mathcal{L}(\mathbf{x}, \mathbf{y})
$$
$$
= \int_\Omega d\mu(\mathbf{x}) \int_N d\mu(\mathbf{y}) \, (\nabla_{1,\mathfrak{v}} - \nabla_{2,\mathfrak{v}}) \mathcal{L}(\mathbf{x}, \mathbf{y}) = \int_\Omega \big(\nabla_{\mathfrak{v}}\mathfrak{s} - \Delta\mathfrak{v}\big)(\mathbf{x}) \, d\mu(\mathbf{x}) , \qquad (9.1)
$$

where the jet $\mathfrak{v}$ depends on the choice of Φ (see Fig. 3).

In order to define scalar curvature, we take the integrand in (9.1) and minimize it under variations of Φ. Following the procedure for the quasilocal mass, we again assume linear alignment. This leads us to the following definition.

Definition 9.1 We say that Φ and the identification of $\tilde{\Omega}$ with Ω is **deformation-admissible** if the quasilocal mass can be described linearly by (9.1) and if the local alignment condition

$$
\int_N \boldsymbol{\xi}_\alpha \, \nabla_{1,\mathfrak{v}} \mathcal{L}\Big(\zeta - \frac{\boldsymbol{\xi}}{2}, \zeta + \frac{\boldsymbol{\xi}}{2}\Big) \, d^3\boldsymbol{\xi} = 0 \qquad (9.2)
$$

holds. The **synthetic scalar curvature** scal at $\mathbf{x} \in \Omega$ is defined by

$$
\mathrm{scal}(\mathbf{x}) := \inf \big\{ \big(\nabla_{\mathfrak{v}}\mathfrak{s} - \Delta\mathfrak{v}\big)(\mathbf{x}) \,\big|\, \Phi \text{ and } \Omega \text{ are deformation} - \text{admissible} \big\} . \qquad (9.3)
$$

Comparing this definition with that of the quasilocal mass (see Definition 7.2), it is obvious in view of (9.1) that

$$\mathfrak{M}(\tilde{\Omega}) \geq \int_{\Omega} \text{scal}(\mathbf{x}) \, d\mu(\mathbf{x}) \, .$$

In this sense, our synthetic definition of scalar curvature fits together with our notion of quasilocal mass.

10 Example: ultrastatic spacetimes

We now consider the example of ultrastatic spacetimes. In this case, the geometry is determined by the Riemannian metric on the Cauchy surfaces. Our main goal is to show that the synthetic definition of scalar curvature introduced in Sect. 9 agrees in a well-defined limiting case with the scalar curvature of the Riemannian metric. Moreover, we will work out similarities of the quasilocal mass with the Brown–York mass.

10.1 Construction of the Lagrangian

In order to make the example as simple as possible, we keep the background on causal fermion systems to a minimum. In particular, we will only introduce those objects which are essential for the computations, namely the Lagrangian and its jet derivatives $\nabla_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x},\mathbf{y})$ and $\nabla_{2,\mathfrak{v}}\mathcal{L}(\mathbf{x},\mathbf{y})$. However, we do not explain what the space $\mathcal{G}$ is, because this would make it necessary to introduce the so-called local correlation operators. The reader who wants to get a deeper understanding of the connection to causal fermion systems may find it helpful to consult the introduction to static causal fermion systems in [30, Section 3], the survey article [12] or the textbooks [11, 22].

We let $(\mathcal{N}, g)$ be a three-dimensional orientable, complete and asymptotically flat Riemannian manifold. We let $\mathcal{M} := \mathbb{R} \times \mathcal{N}$ be the corresponding ultrastatic spacetime with the line element

$$ds^2 = dt^2 - g_{\alpha\beta}(\mathbf{x}) \, d\mathbf{x}^\alpha \, d\mathbf{x}^\beta \, . \tag{10.1}$$

The completeness of $\mathcal{N}$ implies that this spacetime is globally hyperbolic. Next, being three-dimensional, the manifold $\mathcal{N}$ is spin. Let $\mathcal{D}_{\mathcal{N}}$ denote the intrinsic Dirac operator on $\mathcal{N}$. Using standard elliptic theory (see [40, Proposition 8.2.7] and [31]), the operator $\mathcal{D}_{\mathcal{N}}$ with domain $C_0^\infty(\mathcal{N}, S\mathcal{N})$ is essentially self-adjoint on the Hilbert space $L^2(\mathcal{N}, S\mathcal{N})$. Thus its closure, which we again denote by $\mathcal{D}_{\mathcal{N}}$, is a self-adjoint operator with domain $\mathcal{D}(\mathcal{D}_{\mathcal{N}})$. The spectral theorem yields

$$\mathcal{D}_{\mathcal{N}} = \int_{\mathbb{R}} \lambda \, dF_\lambda \, ,$$

where F_λ denotes the spectral-measure of $\mathcal{D}_{\mathcal{N}}$.

The Dirac operator in spacetime can be written in block matrix notation as

$$\mathcal{D} = \begin{pmatrix} i\partial_t & -\mathcal{D}_{\mathcal{N}} \\ \mathcal{D}_{\mathcal{N}} & -i\partial_t \end{pmatrix} .$$

Since the Dirac operator is time independent, we can separate the time dependence with a plane wave ansatz,

$$\psi(t, x) = e^{-i\omega t} \chi(x) .$$

The sign of ω gives a natural decomposition of the solution space into two subspaces. This is often referred to as "frequency splitting," and the subspaces are called the solutions of *positive* and *negative frequency*, respectively.

We now construct corresponding causal fermion systems. To this end, we choose a parameter $\varepsilon > 0$, referred to as the *regularization length* (this length should be thought of as being very small; it can be identified with the Planck length). Given a parameter m (the *rest mass*) and setting

$$\omega(\lambda) := -\sqrt{k^2 + \lambda^2} , \qquad (10.2)$$

the *regularized kernel of the fermionic projector* $P^\varepsilon(x, y)$ is defined by

$$P^\varepsilon\big((t, \mathbf{x}), (t', \mathbf{y})\big) := -\int_{\mathbb{R}} \begin{pmatrix} \omega(\lambda) + m & -\lambda \\ \lambda & -\omega(\lambda) + m \end{pmatrix} e^{-i\omega(k)(t-t')} e^{\varepsilon\omega} \, dF_\lambda(\mathbf{x}, \mathbf{y}) .$$

By direct computation, one verifies that this kernel satisfies the Dirac equation

$$(\mathcal{D}_x - m) P(x, y) = 0 .$$

Due to the minus sign in (10.2), this kernel is composed of all negative-frequency solutions of the Dirac equation. The factor $\exp(\varepsilon\omega)$ in the integrand can be understood as a convergence-generating factor. In position space, it gives rise to a smoothing of the kernel on the regularization scale.

We next compute the Lagrangian of the causal action principle (for details see for example [11, Section 1.1] or [30, Section 2.2]). To this end, we form the *closed chain*

$$A_{xy} := P(x, y) P(y, x) .$$

Denoting its eigenvalues by $\lambda_1^{xy}, \cdots, \lambda_4^{xy} \in \mathbb{C}$, the Lagrangian in spacetime is defined by

$$\mathcal{L}(x, y) = \frac{1}{8} \sum_{i, j=1}^{4} \left(|\lambda_i^{xy}| - |\lambda_j^{xy}| \right)^2 .$$

The *static Lagrangian* is obtained by integrating over one time variable (for more details on this construction see [30, Section 3])

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) := \int_{-\infty}^{\infty} \mathcal{L}\big((0, \mathbf{x}), (t, \mathbf{y})\big) \, dt . \qquad (10.3)$$

We remark that this Lagrangian decays on the Compton scale m^{-1}. Therefore, choosing $\delta \gg m^{-1}$, the assumption of short range as introduced before (2.18) holds approximately, meaning that $\mathcal{L}(\mathbf{x}, \mathbf{y})$ and its derivative is very small if $d(\mathbf{x}, \mathbf{y}) > \delta$. Nevertheless, one should keep in mind that the strict assumption (2.18) is a mathematical idealization.

Due to the microlocal nature of the construction, one may expect that the static Lagrangian should depend only the geodesic distance $d(\mathbf{x}, \mathbf{y})$. This is indeed the case, up to errors of the following form,

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathcal{L}\big[d(\mathbf{x}, \mathbf{y})\big] \left(1 + \mathcal{O}\Big(\frac{\varepsilon}{\ell_{\text{macro}}}\Big) + \mathcal{O}\Big(\frac{1}{m\,\ell_{\text{macro}}}\Big) \right) . \qquad (10.4)$$

In what follows, we shall not use this formula (details can be found in [5, Appendix A], also based on [17, Section 5]). The reason is that, for getting the connection to the Brown–York mass and scalar curvature, it will suffice to compute the surface layer integrals in regions of spacetime which can be described by linearized gravity. This simplifies the computations considerably, as we now explain in detail.

10.2 Linearized gravity

In linearized gravity, one assumes that, in a suitable coordinate chart, the metric $g_{\alpha\beta}$ differs from the Euclidean metric $\delta_{\alpha\beta}$ by a small tensor denoted by $h_{\alpha\beta}$,

$$g_{\alpha\beta}(\mathbf{x}) = \delta_{\alpha\beta} + h_{\alpha\beta}(\mathbf{x}) \,. \tag{10.5}$$

Here "small" means that we only take into account the tensor $h_{\alpha\beta}$ linearly (in other words, we work with error terms $\mathcal{O}((h_{\alpha\beta})^2)$ throughout). In this formalism, tensor indices are raised and lowered with respect to the Euclidean metric (for details in the more general Lorentzian setting see for example [35, §105]). Therefore, we do not need to distinguish between upper and lower indices. For notational simplicity, we write all indices as lower indices (this convention is useful mainly in view of keeping track of the signs in the computations in Lorentzian signature in Appendix A). Using the Einstein summation convention, we sum over all double indices (this is unproblematic because we only consider Euclidean coordinate transformations).

In the resulting formalism, one has the freedom to perform infinitesimal coordinate transformations

$$\mathbf{x}'^{\alpha} = \mathbf{x}^{\alpha} + \boldsymbol{\zeta}^{\alpha}(\mathbf{x}) \,. \tag{10.6}$$

This transforms the linearized metric according to

$$h'_{\alpha\beta} = h_{\alpha\beta} - \partial_{\alpha}\boldsymbol{\zeta}_{\beta} - \partial_{\beta}\boldsymbol{\zeta}_{\alpha} \,. \tag{10.7}$$

The trace of the metric denoted by

$$h := h_{\alpha\alpha} = \delta^{\alpha\beta} h_{\alpha\beta}$$

transforms according to

$$h' = h - 2\partial_{\alpha}\boldsymbol{\zeta}_{\alpha} \,.$$

It follows that the volume form behaves as follows,

$$\sqrt{\det g'} = \sqrt{1 + h'} = 1 + \frac{h'}{2} = 1 + \frac{h}{2} - \partial_{\alpha}\boldsymbol{\zeta}_{\alpha} = \sqrt{\det g} - \partial_{\alpha}\boldsymbol{\zeta}_{\alpha} \,. \tag{10.8}$$

Moreover, the Lagrangian transforms according to

$$\delta\mathcal{L}(\mathbf{x}, \mathbf{y}) := \mathcal{L}(\mathbf{x}', \mathbf{y}') - \mathcal{L}(\mathbf{x}, \mathbf{y}) = -\left(\boldsymbol{\zeta}_{\alpha}(\mathbf{x}) \frac{\partial}{\partial \mathbf{x}_{\alpha}} + \boldsymbol{\zeta}_{\alpha}(\mathbf{y}) \frac{\partial}{\partial \mathbf{y}_{\alpha}}\right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \,. \tag{10.9}$$

We remark that the relative sign of the two terms can be verified for example by integrating over y and using that

$$\int_N \left(\partial_{\alpha}\boldsymbol{\zeta}_{\alpha}(\mathbf{y}) + \boldsymbol{\zeta}_{\alpha}(\mathbf{y}) \frac{\partial}{\partial \mathbf{y}_{\alpha}}\right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d^3\mathbf{y} = \int_N \frac{\partial}{\partial \mathbf{y}_{\alpha}} \left(\boldsymbol{\zeta}_{\alpha}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})\right) d^3\mathbf{y} = 0 \,,$$

being consistent with the fact that the spacetime integral is diffeomorphism invariant.

10.3 Description with the diffeomorphism jet and inner solutions

By translational symmetry, the unperturbed static Lagrangian in Minkowski space depends only on the difference vector $\boldsymbol{\xi} := \mathbf{y} - \mathbf{x}$. We again use the short notation $\mathcal{L}[\boldsymbol{\xi}]$ (due to rotational symmetry, it actually depends only on the norm of the vector $\boldsymbol{\xi}$, but we prefer not to make this explicit in our notation). We denote the linear perturbation of the Lagrangian by $\delta\mathcal{L}(\mathbf{x}, \mathbf{y})$, i.e.

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathcal{L}[\boldsymbol{\xi}] + \delta\mathcal{L}(\mathbf{x}, \mathbf{y}) + \mathcal{O}\big((h_{\alpha\beta})^2\big) \,.$$

The first question is how to compute $\delta\mathcal{L}(\mathbf{x}, \mathbf{y})$. In view of (10.4), the first idea is to compute the linear perturbation of the geodesic distance. A direct calculation gives (for details see for example [30, eq. B.6]))

$$\delta\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \int_0^1 h_{\alpha\beta}|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, d\tau \, \boldsymbol{\xi}_\alpha \frac{\partial}{\partial\boldsymbol{\xi}_\beta} \mathcal{L}[\boldsymbol{\xi}] \,. \tag{10.10}$$

This formula indeed holds, up to error terms which will be specified below. Before doing so, we make a few explanatory remarks on this formula. The integral can be understood as an integration of the perturbation along the unperturbed geodesic joining $\mathbf{x}$ and $\mathbf{y}$ (which here simply is the straight line segment). A simple way of understanding the above formula for $\delta\mathcal{L}(\mathbf{x}, \mathbf{y})$ is to verify that it describes the correct behavior under infinitesimal coordinate transformations. Indeed, for the perturbed metric $h_{\alpha\beta} = -\partial_\alpha\boldsymbol{\zeta}_\beta - \partial_\beta\boldsymbol{\zeta}_\alpha$ (see (10.7)) one gets

$$\delta\mathcal{L}(\mathbf{x}, \mathbf{y}) = -\frac{1}{2} \int_0^1 \big(\partial_\alpha\boldsymbol{\zeta}_\beta + \partial_\beta\boldsymbol{\zeta}_\alpha\big)\big|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, d\tau \, \boldsymbol{\xi}_\alpha \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] \tag{10.11}$$

$$= -\int_0^1 \boldsymbol{\xi}_\alpha \partial_\alpha\boldsymbol{\zeta}_\beta|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, d\tau \, \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] \tag{10.12}$$

$$= -\int_0^1 \frac{d}{d\tau}\boldsymbol{\zeta}_\beta|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, d\tau \, \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] \tag{10.13}$$

$$= -\big(\boldsymbol{\zeta}_\beta(\mathbf{y}) - \boldsymbol{\zeta}_\beta(\mathbf{x})\big) \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] = -\Big(\boldsymbol{\zeta}_\alpha(\mathbf{x}) \frac{\partial}{\partial\mathbf{x}_\alpha} + \boldsymbol{\zeta}_\alpha(\mathbf{y}) \frac{\partial}{\partial\mathbf{y}_\alpha}\Big)\mathcal{L}[\boldsymbol{\xi}] \,, \tag{10.14}$$

giving agreement with (10.9). For clarity, we point out that in (10.12) we used that the function $\mathcal{L}[\boldsymbol{\xi}]$ depends only on the norm $|\boldsymbol{\xi}|$. Therefore, its gradient points in the direction ξ, implying that

$$\boldsymbol{\xi}_\alpha \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] = \boldsymbol{\xi}_\beta \frac{\partial}{\partial\boldsymbol{\xi}_\alpha}\mathcal{L}[\boldsymbol{\xi}] \,.$$

In (10.13), on the other hand, we rewrote the derivative in the direction $\boldsymbol{\xi}$ as a τ-derivatives (this formula is immediately verified by carrying out the τ-derivative with the chain rule and comparing with (10.12)). Finally, in (10.14) we integrated the τ-derivative by parts, giving boundary terms at $\mathbf{y}$ and $\mathbf{x}$.

The formula (10.10) can be derived rigorously using the method of integration along characteristics (for the general context see for example [33]) or, more specifically for the Dirac equation in Minkowski space, the so-called light-cone expansion of the kernel of the fermionic projector (see [7, Appendix B] and [11, Section 2.3 and §4.5.2] as well as [11, Appendix F] and [26] for the issues related to the regularization). This procedure also gives rise to correction terms involving the curvature tensor. All these correction terms will be

negligible in our computations, because, following the considerations after Proposition 5.7, they can be absorbed into the error term (5.23).

Our next step is to write the formula (10.10) in the jet formalism. Since both arguments $\mathbf{x}$ and $\mathbf{y}$ are perturbed, we have

$$\delta\mathcal{L}(\mathbf{x}, \mathbf{y}) = \big(D_{1,\mathbf{v}} + D_{2,\mathbf{v}}\big)\mathcal{L}(\mathbf{x}, \mathbf{y}) \,,$$

where $\mathbf{v}$ is a vector field on $\mathcal{G}$ along N. In the jet formalism, the change of the volume form is described by the scalar component of the jet. According to (10.8), we need to choose the scalar component as $h/2$, leading us to the formula

$$\begin{aligned}
\big(\nabla_{1,\mathfrak{v}} + \nabla_{2,\mathfrak{v}}\big)\mathcal{L}(\mathbf{x}, \mathbf{y}) = {} & \frac{1}{2}\big(h(\mathbf{x}) + h(\mathbf{y})\big)\,\mathcal{L}[\boldsymbol{\xi}] \\
& + \frac{1}{2}\int_0^1 h_{\alpha\beta}|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,d\tau\,\boldsymbol{\xi}_\alpha\,\frac{\partial}{\partial\boldsymbol{\xi}_\beta}\,\mathcal{L}[\boldsymbol{\xi}]\,.
\end{aligned} \tag{10.15}$$

The shortcoming of the considerations so far is that we can only compute the combination of derivatives $(\nabla_{1,\mathfrak{v}} + \nabla_{2,\mathfrak{v}})\mathcal{L}(\mathbf{x}, \mathbf{y})$, but it remains unclear what a single derivative like $\nabla_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x}, \mathbf{y})$ should be. Such derivative terms, however, are essential for computing anti-symmetric derivatives as in (2.17) or single derivatives as in (5.6). At this point, we need to use the detailed form of the light-cone expansion involving unbounded integrals as developed in [9, Appendix F] and, more generally and more systematically, in [16, Appendix C]. For simplicity, we shall not derive these formulas, but merely state and briefly explain them. The resulting formulas for the jet derivatives are

$$\nabla_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{h(\mathbf{x})}{2}\,\mathcal{L}[\boldsymbol{\xi}] - \frac{1}{4}\int_{-\infty}^{\infty}\epsilon(\tau)\,d\tau\,h_{\alpha\beta}|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,\boldsymbol{\xi}_\alpha\,\frac{\partial}{\partial x_\beta}\mathcal{L}[\boldsymbol{\xi}] \tag{10.16}$$

$$\nabla_{2,\mathfrak{v}}\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{h(\mathbf{y})}{2}\,\mathcal{L}[\boldsymbol{\xi}] + \frac{1}{4}\int_{-\infty}^{\infty}\epsilon(1-\tau)\,d\tau\,h_{\alpha\beta}|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,\boldsymbol{\xi}_\alpha\,\frac{\partial}{\partial y_\beta}\mathcal{L}[\boldsymbol{\xi}] \tag{10.17}$$

(here ϵ is the *sign function* $\epsilon(\tau) = 1$ if $\tau > 0$ and $\epsilon(\tau) = -1$ otherwise). Similar to (10.15), these formulas again involve line integrals, but this time along the whole straight line through the points $\mathbf{x}$ and $\mathbf{y}$. These unbounded integrals are nonlocal in the sense that, even if $\mathbf{x}$ and $\mathbf{y}$ are close together, the integral extends all the way to infinity. If we add the formulas (10.16) and (10.17), the unbounded integrals combine to a bounded integral, giving back the previous formula (10.15). A simple way of understanding the form of (10.16) (and similarly (10.17)) is by considering an infinitesimal coordinate transformation. In this case, we obtain similar to (10.11)

$$\begin{aligned}
\nabla_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x}, \mathbf{y}) = {} & -\mathrm{div}\,\boldsymbol{\zeta}(\mathbf{x})\,\mathcal{L}[\boldsymbol{\xi}] - \frac{1}{2}\int_{-\infty}^{\infty}\epsilon(\tau)\,\boldsymbol{\xi}_\alpha\partial_\alpha\boldsymbol{\zeta}_\beta|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,d\tau\,\frac{\partial}{\partial\boldsymbol{\xi}_\beta}\,\mathcal{L}[\boldsymbol{\xi}] \\
= {} & -\mathrm{div}\,\boldsymbol{\zeta}(\mathbf{x})\,\mathcal{L}[\boldsymbol{\xi}] - \frac{1}{2}\int_{-\infty}^{\infty}\epsilon(\tau)\frac{d}{d\tau}\boldsymbol{\zeta}_\beta|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,d\tau\,\frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] \\
= {} & -\mathrm{div}\,\boldsymbol{\zeta}(\mathbf{x})\,\mathcal{L}[\boldsymbol{\xi}] - \boldsymbol{\zeta}_\alpha(\mathbf{x})\,\frac{\partial}{\partial\mathbf{x}_\alpha}\mathcal{L}[\boldsymbol{\xi}]\,.
\end{aligned} \tag{10.18}$$

The first summand describes the infinitesimal change of the volume form at $\mathbf{x}$. Comparing with (10.9), one sees that the second summand describes the variation of the Lagrangian by the infinitesimal coordinate change, again at the point $\mathbf{x}$. The fact that the jet (10.16) and (10.17) describes the behavior under infinitesimal coordinate transformations or, in more mathematical language, under infinitesimal diffeomorphisms, is the reason why we refer to $\mathfrak{v}$

as the *diffeomorphism jet*. We remark that the formula (10.18) can be expressed by saying that, for infinitesimal coordinate transformations, the jet $\mathfrak{v}$ reduces to an inner solution $\mathfrak{u}$ with

$$\mathfrak{u} = -\left(\operatorname{div}\boldsymbol{\zeta}, \boldsymbol{\zeta}\right). \tag{10.19}$$

We finally verify by direct computation that the EL equations hold when testing with $\mathfrak{v}$.

Lemma 10.1 *The jet $\mathfrak{v}$ defined by (10.16) and (10.17) has the property*

$$0 = \nabla_{\mathfrak{v}}\ell(\boldsymbol{x}) = \int_N \nabla_{1,\mathfrak{v}}\mathcal{L}(\boldsymbol{x},\boldsymbol{y})\,d\mu(\boldsymbol{y}) - \left(\nabla_{\mathfrak{v}}\mathfrak{s}\right)(\boldsymbol{x}).$$

Proof Since $\mathfrak{s}$ is chosen such that ℓ vanishes, we know that

$$\nabla_{\mathfrak{v}}\ell(\mathbf{x}) = \int_N D_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x},\mathbf{y})\,d\mu(\mathbf{y})$$

$$= \frac{1}{4}\int_N d\mu(\mathbf{y})\int_{-\infty}^{\infty}\epsilon(\tau)\,d\tau\,h_{\alpha\beta}|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,\boldsymbol{\xi}_{\alpha}\frac{\partial}{\partial\boldsymbol{\xi}_{\beta}}\mathcal{L}[\boldsymbol{\xi}].$$

Applying the transformations

$$\mathbf{x} \to \mathbf{x}, \quad \mathbf{y} \to 2\mathbf{x} - \mathbf{y} \quad \text{and} \quad \tau \to -\tau, \tag{10.20}$$

the arguments in the integrand transform according to

$$\boldsymbol{\xi} \to -\boldsymbol{\xi}, \quad \tau\mathbf{y} + (1-\tau)\mathbf{x} = \mathbf{x} + \tau\boldsymbol{\xi} \;\to\; \mathbf{x} + \tau\boldsymbol{\xi} = \tau\mathbf{y} + (1-\tau)\mathbf{x}. \tag{10.21}$$

Therefore, the integrand flips sign. We conclude that the integral vanishes by symmetry. $\square$

10.4 Satisfying the alignment condition

After these preparations, we can now analyze the local alignment condition (5.18). For linear gravity, it simplifies to

$$0 = A_{\alpha}^{(0)}(\boldsymbol{\zeta}) = \int_N \boldsymbol{\xi}_{\alpha}\,\nabla_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x},\mathbf{y})\,d^3\boldsymbol{\xi} = 0. \tag{10.22}$$

Using (10.16), we obtain

$$A_{\alpha}^{(0)}(\boldsymbol{\zeta}) = \frac{1}{2}\int_N \boldsymbol{\xi}_{\alpha}\,h(\mathbf{x})\,\mathcal{L}(\mathbf{x},\mathbf{y})\,d^3\boldsymbol{\xi} \tag{10.23}$$

$$+ \frac{1}{4}\int_N d^3\boldsymbol{\xi}\,\boldsymbol{\xi}_{\alpha}\int_{-\infty}^{\infty}\epsilon(\tau)\,d\tau\,h_{\beta\gamma}|_{\boldsymbol{\zeta}+(\tau-\frac{1}{2})\boldsymbol{\xi}}\,\boldsymbol{\xi}_{\beta}\frac{\partial}{\partial\boldsymbol{\xi}_{\gamma}}\mathcal{L}(\mathbf{x},\mathbf{y}). \tag{10.24}$$

In the special case of an infinitesimal coordinate transformation, the jet $\mathfrak{v}$ reduces to the inner solution $\mathfrak{u}$ given in (10.19). In this case, the vector field $A^{(0)}$ was already computed in Lemma 5.5. As explained after this lemma, we can arrange by a suitable choice of the inner solution that the vector field $A^{(0)}$ vanishes. With this in mind, from now on we may assume that the alignment condition (10.22) is satisfied.

We point out that, due to the unbounded line integral in (10.24), the above formula for $A^{(0)}$ is nonlocal. For this reason, it does not seem possible to satisfy the alignment condition (10.22) explicitly or in closed form. Instead, we proceed by working out the consequences of the local alignment condition. To this end, we use that, if the vector field $A^{(0)}$ vanishes identically, then also its divergence is zero. This divergence is indeed a local quantity:

Lemma 10.2 *For the jet $\mathfrak{v}$ given by (10.16),*

$$\frac{\partial}{\partial \zeta_\alpha} A_\alpha^{(0)} = \frac{\mathfrak{s}}{2}\, h(\zeta) + \mathcal{O}\Big(\frac{\delta^2}{\ell_{\mathrm{macro}}^2}\Big).$$

Proof A direct computation using (10.16) yields

$$
\begin{aligned}
\frac{\partial}{\partial \zeta_\alpha} A_\alpha^{(0)} &= \int_N \Big(\boldsymbol{\xi}\,\frac{\partial}{\partial \zeta}\Big) \nabla_{1,\mathfrak{v}} \mathcal{L}(\mathbf{x}, \mathbf{y})\, d^3\boldsymbol{\xi} \\
&= \frac{1}{2} \int_N \Big(\boldsymbol{\xi}\,\frac{\partial}{\partial \zeta}\Big)\, h(\mathbf{x})\, \mathcal{L}[\boldsymbol{\xi}]\, d^3\boldsymbol{\xi} \\
&\quad + \frac{1}{4} \int_N d^3\boldsymbol{\xi} \Big(\boldsymbol{\xi}\,\frac{\partial}{\partial \zeta}\Big) \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, h_{\beta\gamma}|_{\zeta + (\tau - \frac{1}{2})\boldsymbol{\xi}}\, \boldsymbol{\xi}_\beta\, \frac{\partial}{\partial \boldsymbol{\xi}_\gamma} \mathcal{L}[\boldsymbol{\xi}] \\
&= \frac{1}{4} \int_N d^3\boldsymbol{\xi} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, \frac{d}{d\tau} h_{\beta\gamma}|_{\zeta + (\tau - \frac{1}{2})\boldsymbol{\xi}}\, \boldsymbol{\xi}_\beta\, \frac{\partial}{\partial \boldsymbol{\xi}_\gamma} \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\Big(\frac{\delta^2}{\ell_{\mathrm{macro}}^2}\Big) \\
&= -\frac{1}{2} \int_N d^3\boldsymbol{\xi}\, h_{\beta\gamma}|_{\zeta - \frac{1}{2}\boldsymbol{\xi}}\, \boldsymbol{\xi}_\beta\, \frac{\partial}{\partial \boldsymbol{\xi}_\gamma} \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\Big(\frac{\delta^2}{\ell_{\mathrm{macro}}^2}\Big) \\
&= \frac{1}{2} \int_N d^3\boldsymbol{\xi}\, h|_{\zeta - \frac{1}{2}\boldsymbol{\xi}}\, \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\Big(\frac{\delta^2}{\ell_{\mathrm{macro}}^2}\Big) = \frac{\mathfrak{s}}{2}\, h(\zeta) + \mathcal{O}\Big(\frac{\delta^2}{\ell_{\mathrm{macro}}^2}\Big),
\end{aligned}
$$

giving the result. $\qquad\square$

This lemma shows that, up to the specified error term, the alignment condition implies that h vanishes. In other words, alignment gives rise to volume conservation.

10.5 Similarity between the quasilocal mass and the Brown–York mass

We assume that the system is weakly gravitating near the boundary $\partial\tilde{\Omega}$, so that the quasilocal mass can be computed linearly via (2.17) with the jet $\mathfrak{v}$ as given by (10.16) and (10.17). Moreover, we assume that the local alignment condition 5.18 is satisfied, i.e.

$$A^{(0)}(\zeta) = 0 \qquad \text{for all } \zeta \in \partial\Omega.$$

Then, according to Corollary 5.2, the quasilocal mass is given by

$$\mathfrak{M}(\tilde{\Omega}) = \int_\Omega A_\alpha^{(1)}(\mathbf{x})\, v_\alpha(\mathbf{x})\, d\mu_{\partial\Omega}(\mathbf{x}) \Big(1 + \mathcal{O}\Big(\frac{\delta}{\ell_{\mathrm{macro}}}\Big)\Big).$$

Therefore, our task is to get a connection between the vector field $A^{(1)}(\mathbf{x})$ for a boundary point $\mathbf{x} \in \partial\Omega$ and the mean curvatures of $\tilde{\Omega}$ and Ω. We first compute $A^{(1)}(\mathbf{x})$ in an expansion in powers of the range δ.

Lemma 10.3 *For the jet $\mathfrak{v}$ given by (10.16) and (10.17),*

$$A_\alpha^{(1)}(\zeta) = \frac{1}{144}\, \delta^2\, \mathfrak{s}_2\Big(2\partial_\beta h_{\alpha\beta}(\zeta) + \partial_\alpha h(\zeta)\Big) + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\mathrm{macro}}^3}\Big).$$

Proof We make the straightforward computation using (5.15)

$$\int_N \boldsymbol{\xi}_\alpha \Big(\boldsymbol{\xi}\,\frac{\partial}{\partial \zeta}\Big)^2 \nabla_{1,\mathfrak{v}} \mathcal{L}(\mathbf{x}, \mathbf{y})\, d^3\boldsymbol{\xi}$$

$$= \frac{1}{2} \int_N \boldsymbol{\xi}_\alpha \left(\boldsymbol{\xi} \, \frac{\partial}{\partial \boldsymbol{\zeta}} \right)^2 h(\mathbf{x}) \, \mathcal{L}[\boldsymbol{\xi}] \, d^3 \boldsymbol{\xi}$$

$$+ \frac{1}{4} \int_N d^3 \boldsymbol{\xi} \, \boldsymbol{\xi}_\alpha \left(\boldsymbol{\xi} \, \frac{\partial}{\partial \boldsymbol{\zeta}} \right)^2 \int_{-\infty}^\infty \epsilon(\tau) \, d\tau \, h_{\beta\gamma}|_{\boldsymbol{\zeta} + (\tau - \frac{1}{2})\boldsymbol{\xi}} \, \boldsymbol{\xi}_\beta \, \frac{\partial}{\partial \boldsymbol{\xi}_\gamma} \mathcal{L}[\boldsymbol{\xi}]$$

$$= \frac{1}{4} \int_N d^3 \boldsymbol{\xi} \, \boldsymbol{\xi}_\alpha \left(\boldsymbol{\xi} \, \frac{\partial}{\partial \boldsymbol{\zeta}} \right)^2 \int_{-\infty}^\infty \epsilon(\tau) \, d\tau \, h_{\beta\gamma}|_{\boldsymbol{\zeta} + (\tau - \frac{1}{2})\boldsymbol{\xi}} \, \boldsymbol{\xi}_\beta \, \frac{\partial}{\partial \boldsymbol{\xi}_\gamma} \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3} \right)$$

$$= \frac{1}{4} \int_N d^3 \boldsymbol{\xi} \, \boldsymbol{\xi}_\alpha \left(\boldsymbol{\xi} \, \frac{\partial}{\partial \boldsymbol{\zeta}} \right) \int_{-\infty}^\infty \epsilon(\tau) \, d\tau \, \frac{d}{d\tau} h_{\beta\gamma}|_{\boldsymbol{\zeta} + (\tau - \frac{1}{2})\boldsymbol{\xi}} \, \boldsymbol{\xi}_\beta \, \frac{\partial}{\partial \boldsymbol{\xi}_\gamma} \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3} \right)$$

$$= -\frac{1}{2} \int_N d^3 \boldsymbol{\xi} \, \boldsymbol{\xi}_\alpha \left(\boldsymbol{\xi} \, \frac{\partial}{\partial \boldsymbol{\zeta}} \right) h_{\beta\gamma}|_{\boldsymbol{\zeta} - \frac{1}{2}\boldsymbol{\xi}} \, \boldsymbol{\xi}_\beta \, \frac{\partial}{\partial \boldsymbol{\xi}_\gamma} \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3} \right)$$

$$= -\frac{1}{2} \int_N d^3 \boldsymbol{\xi} \, \boldsymbol{\xi}_\alpha \, \boldsymbol{\xi}_\delta \, \partial_\delta h_{\beta\gamma}|_{\boldsymbol{\zeta} - \frac{1}{2}\boldsymbol{\xi}} \, \boldsymbol{\xi}_\beta \, \frac{\partial}{\partial \boldsymbol{\xi}_\gamma} \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3} \right)$$

$$= \frac{1}{2} \int_N d^3 \boldsymbol{\xi} \, \boldsymbol{\xi}_\delta \, \partial_\delta h_{\beta\alpha}|_{\boldsymbol{\zeta} - \frac{1}{2}\boldsymbol{\xi}} \, \boldsymbol{\xi}_\beta \, \mathcal{L}[\boldsymbol{\xi}]$$

$$+ \frac{1}{2} \int_N d^3 \boldsymbol{\xi} \, \boldsymbol{\xi}_\alpha \, \partial_\gamma h_{\beta\gamma}|_{\boldsymbol{\zeta} - \frac{1}{2}\boldsymbol{\xi}} \, \boldsymbol{\xi}_\beta \, \mathcal{L}[\boldsymbol{\xi}]$$

$$+ \frac{1}{2} \int_N d^3 \boldsymbol{\xi} \, \boldsymbol{\xi}_\alpha \, \boldsymbol{\xi}_\delta \, \partial_\delta h|_{\boldsymbol{\zeta} - \frac{1}{2}\boldsymbol{\xi}} \, \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3} \right)$$

$$= \frac{1}{6} \, \delta^2 \, \mathfrak{s}_2 \Big(\partial_\beta h_{\beta\alpha}(\boldsymbol{\zeta}) + \partial_\gamma h_{\alpha\gamma}(\boldsymbol{\zeta}) + \partial_\alpha h(\boldsymbol{\zeta}) \Big) + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3} \right)$$

$$= \frac{1}{6} \, \delta^2 \, \mathfrak{s}_2 \Big(2\partial_\beta h_{\alpha\beta}(\boldsymbol{\zeta}) + \partial_\alpha h(\boldsymbol{\zeta}) \Big) + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3} \right) .$$

Applying (5.6) gives the result.   □

Let (N, g) be the Riemannian manifold with metric $g_{\alpha\beta}$ given by (10.5). Given $\Omega \subset N$, we denote the first fundamental form (i.e. the induced Riemannian metric) on $\partial\Omega$ by $g|_{\partial\Omega}$. Moreover, we denote the outer normal on $\partial\Omega$ by ν, and the second fundamental form by $k|_{\partial\Omega}$. For the definition of the Brown–York mass (see [4, 37, 39]), one considers an isometric embedding of $\partial\tilde{\Omega}$ into N and sets

$$\mathfrak{M}_{\text{BY}} := \frac{1}{8\pi\kappa} \int_{\partial\Omega} \left(\text{tr} \, \tilde{k}|_{\partial\Omega} - \text{tr} \, k|_{\partial\Omega} \right) d\mu_{\partial\Omega}$$

(where κ is Newton's constant). In our linearized description near the boundary, the isometry of the embedding means that $\delta g|_{\partial\Omega} = 0$. On the other hand, Lemma 10.2 shows that local alignment yields volume preservation up to an error term. Putting all these properties together, our quasilocal mass indeed coincides with the Brown–York mass, as is made precise in the next theorem.

Theorem 10.4 *Assume that the induced metric on $\partial\Omega$ is preserved by the variation, i.e.*

$$\delta g|_{\partial\Omega} = 0 . \tag{10.25}$$

Moreover, assume that the volume form is preserved approximately in the sense that

$$h_{\beta\beta} = \mathcal{O}\left(\frac{\delta^2}{\ell_{\text{macro}}^2} \right) . \tag{10.26}$$

Then the flux of $A^{(1)}$ through $\partial\Omega$ is related to the first variation of the integral of mean curvature by

$$\int_{\partial\Omega} A^{(1)}_\alpha \, \nu_\alpha \, d\mu_{\partial\Omega} = -\frac{1}{36} \, \delta^2 \, \mathfrak{s}2 \int_{\partial\Omega} \operatorname{tr} \delta k|_{\partial\Omega} \, d\mu_{\partial\Omega} + \mathcal{O}\!\left(\frac{\delta^3}{\ell^3_{\text{macro}}}\right). \qquad (10.27)$$

Proof Given a point $\mathbf{x}_0 \in \partial\Omega$, we choose specific local coordinates near $\mathbf{x}_0$ corresponding to the unperturbed metric. To this end, we first choose a Gaussian coordinate system $(\mathbf{x}_1, \mathbf{x}_2)$ on $\partial\Omega$. Thus, denoting tensor indices on $\partial\Omega$ by $i, j \in \{1, 2\}$, the first fundamental form can be written as

$$g^{(0)}_{ij}(\mathbf{x}_0) = \delta_{ij} \qquad \text{and} \qquad \partial_k g^{(0)}_{ij}(\mathbf{x}_0) = 0$$

(here the superscript (0) clarifies that we consider the unperturbed metric). Next, we extend these coordinates to coordinates on N by choosing $\mathbf{x}_3$ as the arc-length parametrization of geodesics which leave $\partial\Omega$ orthogonally (in the literature, these coordinates are sometimes referred to as Fermi coordinates or a tubular neighborhood). In these coordinates, the unperturbed metric takes the form

$$g^{(0)}_{\alpha\beta} = \begin{pmatrix} g^{(0)}_{11} & g^{(0)}_{12} & 0 \\ g^{(0)}_{21} & g^{(0)}_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In these coordinates, the variation of the metric δg looks as follows. The fact that the first fundamental form is preserved means that

$$\delta g_{ij}|_{(\partial\Omega)\cap U} = 0.$$

Moreover, since the volume form is approximately preserved, we know that $\delta g_{\alpha\alpha} = \mathcal{O}(\delta^2/\ell^2_{\text{macro}})$, so that

$$\delta g_{\alpha\beta}|_{(\partial\Omega)\cap U} = \begin{pmatrix} 0 & 0 & \delta g_{13} \\ 0 & 0 & \delta g_{23} \\ \delta g_{31} & \delta g_{32} & \mathcal{O}(\delta^2/\ell^2_{\text{macro}}) \end{pmatrix}.$$

We next compute the normal vector field ν and the second fundamental form. Clearly, the unperturbed normal is given by

$$\nu^{(0)}_\alpha = (0, 0, 1).$$

Moreover, since δg_{33} is approximately zero, we know that

$$\delta\nu^{(0)}_3 = \mathcal{O}\!\left(\frac{\delta^2}{\ell^2_{\text{macro}}}\right).$$

Let $e_\alpha = \partial_{\mathbf{x}_\alpha}$ be the basis of $T_{\mathbf{x}_0}\partial\Omega$. Then the second fundamental form and its first variation take the form

$$k_{ij}(\mathbf{x}_0) = -g(\nu, \nabla_i e_j) = -\Gamma^k_{ij} \, \nu_k$$

$$\delta k_{ij}(\mathbf{x}_0) = -\left(\delta\Gamma^k_{ij}\right) \nu_k + \Gamma^k_{ij} \, \delta\nu_k = -\delta\Gamma^3_{ij} + \Gamma^\gamma_{ij} \, \delta\nu_\gamma = -\delta\Gamma^3_{ij} + \mathcal{O}\!\left(\frac{\delta^2}{\ell^2_{\text{macro}}}\right)$$

$$= -\frac{1}{2} \left(\partial_i \delta g_{3j} + \partial_j \delta g_{3i} - \partial_3 \delta g_{ij}\right) + \mathcal{O}\!\left(\frac{\delta^2}{\ell^2_{\text{macro}}}\right)$$

$$\operatorname{tr} \delta k(\mathbf{x}_0) = \delta k_{ii}(\mathbf{x}_0) = -\partial_i \delta g_{3i} + \frac{1}{2}\, \partial_3 \delta g_{ii} + \mathcal{O}\Big(\frac{\delta^2}{\ell_{\text{macro}}^2}\Big)\,.$$

In general coordinates, the last identity can be written as

$$\operatorname{tr} \delta k = -\big(\nabla_i \delta g_\alpha^i\big)\, \nu^\alpha + \frac{1}{2}\, D_\nu \delta g_i^i + \mathcal{O}\Big(\frac{\delta^2}{\ell_{\text{macro}}^2}\Big)\,. \tag{10.28}$$

On the other hand, applying Lemma 10.3, in the chosen coordinates around $\mathbf{x}_0$ we have

$$\begin{aligned}
A_\alpha^{(1)}(\mathbf{x}_0)\, \nu^\alpha &= \frac{1}{72}\, \delta^2\, \mathfrak{s}_2\, \partial_\beta \delta g_{3\beta} + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big) \\
&= \frac{1}{72}\, \delta^2\, \mathfrak{s}_2\, \big(\partial_i \delta g_{3i} + \partial_3 \delta g_{33}\big) + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big) \\
&= \frac{1}{72}\, \delta^2\, \mathfrak{s}_2\, \big(\partial_i \delta g_{3i} - \partial_3 \delta g_{ii}\big) + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big)\,.
\end{aligned}$$

Thus in general coordinates,

$$A_\alpha^{(1)}\, \nu^\alpha = \frac{1}{72}\, \delta^2\, \mathfrak{s}_2\, \Big(\big(\nabla_i \delta g_\alpha^i\big)\nu^\alpha - D_\nu \delta g_i^i\Big) + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big)\,. \tag{10.29}$$

The identities (10.28) and (10.29) are not multiples of each other because of the different relative prefactors of the two summands. Therefore, we must make use of the fact that, integrating over $\partial\Omega$, gives an additional identity. Indeed,

$$\begin{aligned}
&\int_{\partial\Omega} \big(\nabla_i \delta g_\alpha^i\big)\, \nu^\alpha\, d\mu_{\partial\Omega} \\
&= \int_{\partial\Omega} \nabla_i\big(\delta g_\alpha^i\, \nu^\alpha\big)\, d\mu_{\partial\Omega} - \int_{\partial\Omega} \delta g_\alpha^i\, \nabla_i \nu^\alpha\, d\mu_{\partial\Omega}\,.
\end{aligned}$$

The first term vanishes by the Gauss divergence theorem. The last summand, on the other hand, also vanishes because $\nabla_i \nu^\alpha$ is non-zero only if $\alpha \in \{1, 2\}$, in which case δg_α^i vanishes. We conclude that

$$\int_{\partial\Omega} \big(\nabla_i \delta g_\alpha^i\big)\, \nu^\alpha\, d\mu_{\partial\Omega} = 0\,.$$

Using this identity in (10.28) and (10.29) gives the result.   □

We finally explain this result. The formula (10.27) shows that, under the assumptions of the theorem, the quasilocal mass coincides with the Brown–York mass (up to an irrelevant prefactor and the error term). This result gives a connection between the quasilocal mass for causal variational principles and corresponding notions of differential geometry. This connection is surprising, because the positivity proof used here is completely different from the proof for the Brown–York mass in [39]. However, we point out that, despite these similarities, there are also major differences between our quasilocal mass and the Brown–York mass. In order to explain this point, we need to discuss the assumptions of Theorem 10.4. The preservation of the first fundamental form (10.25) is the linearized statement of the isometric embedding used in the Brown–York mass. Likewise, the approximate preservation of the volume form (10.26) is a consequence of the alignment condition (see Lemma 10.2). However, it is not clear how the isometry property (10.25) comes about in the setting of causal variational principles. In order to clarify this point, it seems necessary to take into account

that the quasilocal mass involves a minimization over the symmetry transformations $\Phi \in \mathcal{G}$. It might be that, as a consequence of this minimization process, the identification of $\tilde{\Omega}$ with Ω is an isometry of the boundary. However, verifying this conjecture goes beyond the scope of the present work.

10.6 Computing the synthetic scalar curvature

We now compute synthetic scalar curvature. We again consider the linearized description. Following (9.1), the expression in (9.3) can be written as

$$\left(\nabla_\mathfrak{v}\mathfrak{s} - \Delta\mathfrak{v}\right)(\mathbf{x}) = \int_N \left(\nabla_{1,\mathfrak{v}} - \nabla_{2,\mathfrak{v}}\right)\mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) \,.$$

Applying Theorem 5.1, we can rewrite this quantity as the divergence of the vector field A,

$$\left(\nabla_\mathfrak{v}\mathfrak{s} - \Delta\mathfrak{v}\right)(\mathbf{x}) = \operatorname{div} A(\mathbf{x}) \,.$$

As a consequence of the local alignment condition 9.2, the vector field $A^{(0)}$ vanishes. Therefore, using the scaling (5.7), we obtain

$$\left(\nabla_\mathfrak{v}\mathfrak{s} - \Delta\mathfrak{v}\right)(\mathbf{x}) = \operatorname{div} A^{(1)}(\mathbf{x}) + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3}\right) \,.$$

This divergence is computed in the following lemma.

Lemma 10.5

$$\left(\nabla_\mathfrak{v}\mathfrak{s} - \Delta\mathfrak{v}\right)(x) = \frac{1}{3} \, \delta^2 \, \mathfrak{s}_2 \, \partial_{\alpha\beta}h_{\alpha\beta}(x) + \frac{1}{6} \, \delta^2 \, \mathfrak{s}_2 \, \Delta_{\mathbb{R}^3}h(x) + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3}\right) \,. \quad (10.30)$$

Proof We compute the divergence of $A^{(1)}$ according to

$$\frac{\partial}{\partial\zeta_\alpha} \int_N \boldsymbol{\xi}_\alpha \left(\boldsymbol{\xi}\,\frac{\partial}{\partial\boldsymbol{\zeta}}\right)^2 \nabla_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x}, \mathbf{y}) \, d^3\boldsymbol{\xi}$$

$$= \int_N \left(\boldsymbol{\xi}\,\frac{\partial}{\partial\boldsymbol{\zeta}}\right)^3 \nabla_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x}, \mathbf{y}) \, d^3\boldsymbol{\xi}$$

$$= \frac{1}{2} \int_N \left(\boldsymbol{\xi}\,\frac{\partial}{\partial\boldsymbol{\zeta}}\right)^3 h(\mathbf{x}) \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\boldsymbol{\xi}$$

$$+ \frac{1}{4} \int_N d^3\boldsymbol{\xi} \left(\boldsymbol{\xi}\,\frac{\partial}{\partial\boldsymbol{\zeta}}\right)^3 \int_{-\infty}^\infty \epsilon(\tau) \, d\tau \, h_{\alpha\beta}|_{\boldsymbol{\zeta}+(\tau-\frac{1}{2})\boldsymbol{\xi}} \, \boldsymbol{\xi}_\alpha \, \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}]$$

$$= \frac{1}{4} \int_N d^3\boldsymbol{\xi} \left(\boldsymbol{\xi}\,\frac{\partial}{\partial\boldsymbol{\zeta}}\right)^3 \int_{-\infty}^\infty \epsilon(\tau) \, d\tau \, h_{\alpha\beta}|_{\boldsymbol{\zeta}+(\tau-\frac{1}{2})\boldsymbol{\xi}} \, \boldsymbol{\xi}_\alpha \, \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3}\right)$$

$$= \frac{1}{4} \int_N d^3\boldsymbol{\xi} \left(\boldsymbol{\xi}\,\frac{\partial}{\partial\boldsymbol{\zeta}}\right)^2 \int_{-\infty}^\infty \epsilon(\tau) \, d\tau \, \frac{d}{d\tau}h_{\alpha\beta}|_{\boldsymbol{\zeta}+(\tau-\frac{1}{2})\boldsymbol{\xi}} \, \boldsymbol{\xi}_\alpha \, \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3}\right)$$

$$= -\frac{1}{2} \int_N d^3\boldsymbol{\xi} \left(\boldsymbol{\xi}\,\frac{\partial}{\partial\boldsymbol{\zeta}}\right)^2 h_{\alpha\beta}|_{\boldsymbol{\zeta}-\frac{1}{2}\boldsymbol{\xi}} \, \boldsymbol{\xi}_\alpha \, \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3}\right)$$

$$= -\frac{1}{2} \int_N d^3\boldsymbol{\xi} \, \boldsymbol{\xi}_\gamma\boldsymbol{\xi}_\delta \, \partial_{\gamma\delta}h_{\alpha\beta}|_{\boldsymbol{\zeta}-\frac{1}{2}\boldsymbol{\xi}} \, \boldsymbol{\xi}_\alpha \, \frac{\partial}{\partial\boldsymbol{\xi}_\beta}\mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3}\right)$$

$$= \int_N d^3\xi \, \xi_\gamma \, \partial_{\gamma\beta} h_{\alpha\beta}|_{\zeta - \frac{1}{2}\xi} \, \xi_\alpha \, \mathcal{L}[\xi]$$

$$+ \frac{1}{2} \int_N d^3\xi \, \xi_\gamma \xi_\delta \, \partial_{\gamma\delta} h|_{\zeta - \frac{1}{2}\xi} \, \mathcal{L}[\xi] + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big)$$

$$= \frac{1}{3} \, \delta^2 \, \mathfrak{s}_2 \, \partial_{\alpha\beta} h_{\alpha\beta}(\zeta) + \frac{1}{6} \, \delta^2 \, \mathfrak{s}_2 \, \Delta_{\mathbb{R}^3} h(\zeta) + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big) .$$

This gives the result.   □

In order to relate the obtained expression to the scalar curvature of the spatial metric, we first note that, according to Lemma 10.2, linear alignment gives rise to volume preservation. Therefore, taking into account the error term, we can rewrite (10.30) as

$$\big(\nabla_\mathfrak{v}\mathfrak{s} - \Delta\mathfrak{v}\big)(\mathbf{x}) = \frac{1}{3} \, \delta^2 \, \mathfrak{s}_2 \, \Big(\partial_{\alpha\beta} h_{\alpha\beta}(\mathbf{x}) - \Delta_{\mathbb{R}^3} h(\mathbf{x})\Big) + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big)$$

$$= \frac{1}{3} \, \delta^2 \, \mathfrak{s}_2 \, \text{scal}_g(\mathbf{x}) + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big),$$

where scal_g is the linearized scalar curvature of the Riemannian metric g (10.5) (see for example [1, Theorem 1.174 (e)]). Since this relation does not depend on the choice of Φ and Ω, in (9.3) we can take the infimum to obtain the following result.

Theorem 10.6 *For an ultrastatic spacetime with linearized gravity, the synthetic scalar curvature (see Definition 9.1) agrees, up to a constant and an error term, with the scalar curvature of the Riemannian metric,*

$$\text{scal}(\mathbf{x}) = \frac{1}{3} \, \delta^2 \, \mathfrak{s}_2 \, \text{scal}_g(\mathbf{x}) + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big) .$$

Appendix A: Example: Schwarzschild spacetime

This appendix is devoted to a detailed computation of the total mass and the asymptotic alignment in the prime example of the Schwarzschild geometry.

A.1 The static Lagrangian in the Schwarzschild geometry

We begin with the Schwarzschild metric in Schwarzschild coordinates,

$$ds^2 = \Big(1 - \frac{2M}{r}\Big) dt^2 - \Big(1 - \frac{2M}{r}\Big)^{-1} dr^2 - r^2 \, d\vartheta^2 - r^2 \sin^2 \vartheta \, d\varphi^2 .$$

The volume form is given by

$$d\rho = \sqrt{|\det g|} \, d^4x = dt \, r^2 \, dr \, \sin\vartheta \, d\vartheta \, d\varphi .$$

We write it as

$$d\rho = dt \, d\mu \quad \text{with} \quad d\mu := r^2 \, dr \, \sin\vartheta \, d\vartheta \, d\varphi .$$

We also write $x = (t, \mathbf{x})$, where $\mathbf{x} = (r, \vartheta, \varphi)$ are the spatial coordinates.

The static causal fermion system in the Schwarzschild geometry is obtained following the general construction in [12, 15]; for details see [30, Section 6.1]. The *static Lagrangian* $\mathcal{L}(\mathbf{x}, \mathbf{y})$ is obtained again by integrating over one time variable (10.3). We point out that here we do *not* show that $\tilde{\mu}$ and μ are asymptotically close and that $\tilde{\mu}$ is asymptotically flat (see Definitions 2.3 and 2.5). These technical questions will be analyzed elsewhere. Instead, we here restrict attention to the linearized description near infinity. Our goal is to compute jets, alignments and the total mass explicitly and in detail.

A.2 The Euler–Lagrange equations in linearized gravity

Denoting the radial coordinates by

$$R := |\mathbf{x}| \quad \text{and} \quad r := |\mathbf{y}| \,,$$

the function ℓ defined in (2.5) takes the form

$$\ell(\mathbf{x}) + \mathfrak{s} = \int_{2M}^{\infty} r^2 \, dr \int_0^{2\pi} d\varphi \int_0^{\pi} \sin \vartheta \, d\vartheta \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \,.$$

The EL equations (2.6) state that, for a suitable choice of the parameter $\mathfrak{s} > 0$, this function must be positive and must vanish for all $x \in N$.

Since we are only interested in the asymptotic behavior near infinity, it suffices to consider *linearized gravity*. Following the presentation in [35, §105], we write the metric as

$$g_{jk}(\mathbf{x}) = \eta_{jk} + h_{jk}(\mathbf{x}) \,, \tag{A.1}$$

where $\eta_{jk} = \text{diag}(1, -1, -1, -1)$ is the Minkowski metric. We only consider the static setting; this is why h_{jk} depends only on the spatial coordinates $\mathbf{x}$. We always raise and lower the metric with respect to the Minkowski metric. In order to describe the Schwarzschild metric in this formalism, we work in space with Cartesian coordinates $\mathbf{x} \in \mathbb{R}^3$. Then, introducing the abbreviation

$$V(\mathbf{x}) := -\frac{2M}{|\mathbf{x}|} \,, \tag{A.2}$$

the linearized metric is given by

$$h_{00}(x) = V(\mathbf{x}) \,, \quad h_{\alpha\beta}(x) = V(\mathbf{x}) \, \hat{\mathbf{x}}_\alpha \, \hat{\mathbf{x}}_\beta \,, \tag{A.3}$$

where $\hat{\mathbf{x}}$ denotes the unit vector in the direction of $\mathbf{x}$, i.e.

$$\hat{\mathbf{x}} := \frac{\mathbf{x}}{|\mathbf{x}|} \,.$$

The tensor h_{ij} is trace-free but not divergence-free, as one sees directly from the computation (see also [35, §105])

$$h = h_{00} - h_{\alpha\alpha} = 0 \tag{A.4}$$

$$\partial_\alpha h_{\alpha\beta} = -\partial_\alpha \left(\frac{2M}{R^3} \mathbf{x}_\alpha \mathbf{x}_\beta \right) = -\frac{2M}{R^3} \mathbf{x}_\beta \left(-3 + 3 + 1 \right)$$

$$= -\frac{2M}{R^2} \hat{\mathbf{x}}_\beta = \partial_\beta \left(\frac{2M}{R} \right) = -\partial_\beta V(\mathbf{x}) \tag{A.5}$$

$$\partial^j h_{jk} = -\sum_{\alpha=1}^{3} \partial_\alpha h_{\alpha k} = \partial_k V(\mathbf{x}) \,. \tag{A.6}$$

Our next step is to compute the first variation of the Lagrangian and to work out the EL equations. Here we make use of the light-cone expansion of the kernel of the fermionic projector as developed in [7, 8]; see also [11, Section 2.3 and §4.5.2]. More specifically, the Lagrangian was computed in the linearized Schwarzschild geometry in [30, Appendix B]. It was shown that any homogeneous, spherically symmetric and static kernel $\mathcal{L}(x, y)$ is perturbed linearly according to (see [30, eq. (B.6)])

$$\delta\mathcal{L}(x, y) = \frac{1}{2} \int_0^1 d\tau \, h^i{}_k\big|_{\tau y + (1-\tau) x} \, \xi^k \, \frac{\partial}{\partial y^i} \mathcal{L}(x, y) \,. \tag{A.7}$$

Exactly as explained for ultrastatic spacetimes in Sect. 10.3, this formula holds up to corrections involving the curvature tensor. All these correction terms will be negligible in our computations, because, again following the considerations after Proposition 5.7, they can be absorbed into the error term (5.23).

The variation of the static Lagrangian (10.3) can be computed integrating by parts,

$$
\begin{aligned}
\delta\mathcal{L}(\mathbf{x}, \mathbf{y}) &= \int_{-\infty}^{\infty} \delta\mathcal{L}\big((0, \mathbf{x}), (t, \mathbf{y})\big) \, dt \\
&= \frac{1}{2} \int_{-\infty}^{\infty} dt \int_0^1 d\tau \, h^i{}_k\big|_{\tau y + (1-\tau) x} \, \xi^k \, \frac{\partial}{\partial y^i} \mathcal{L}\big((0, \mathbf{x}), (t, \mathbf{y})\big) \\
&= -\frac{1}{2} \int_0^1 d\tau \, h_{00}\big|_{\tau y + (1-\tau) x} \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \tag{A.8} \\
&\quad - \frac{1}{2} \int_0^1 d\tau \, h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \xi_\alpha \, \frac{\partial}{\partial \mathbf{y}_\beta} \mathcal{L}[\xi] \,. \tag{A.9}
\end{aligned}
$$

Using that h_{ij} is trace-free (A.4), this formula can be written more compactly as[1]

$$\delta\mathcal{L}(\mathbf{x}, \mathbf{y}) = -\frac{1}{2} \int_0^1 d\tau \, h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \frac{\partial}{\partial \mathbf{y}_\beta} \big(\xi_\alpha \, \mathcal{L}[\xi]\big) \,. \tag{A.10}$$

After these preparation, we can now compute the first variation of the function ℓ introduced in (2.5).

Lemma A.1

$$\delta\ell(\mathbf{x}) = \frac{1}{2} \int_N d^3\mathbf{y} \left(\int_0^1 V\big|_{\tau \mathbf{y} + (1-\tau) \mathbf{x}} \, d\tau - V(\mathbf{y}) \right) \mathcal{L}[\xi] \,.$$

Proof Starting from the variation of the static Lagrangian (A.10), we obtain

$$
\begin{aligned}
\delta\ell(\mathbf{x}) &= \int_N \delta\mathcal{L}(\mathbf{x}, \mathbf{y}) \, d^3 y = -\frac{1}{2} \int_N d^3 y \int_0^1 d\tau \, h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \frac{\partial}{\partial \mathbf{y}_\beta} \big(\xi_\alpha \, \mathcal{L}[\xi]\big) \\
&= \frac{1}{2} \int_N d^3 y \int_0^1 d\tau \, \frac{\partial}{\partial \mathbf{y}_\beta} h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \xi_\alpha \, \mathcal{L}[\xi] \\
&= \frac{1}{2} \int_0^1 d\tau \int_N d^3\mathbf{y} \, \tau \, \big(\partial_\beta h_{\alpha\beta}\big)\big|_{\tau y + (1-\tau) x} \, \xi_\alpha \, \mathcal{L}[\xi]
\end{aligned}
$$

---

[1] To avoid confusion, we note that the sign difference compared to the integral in (10.15) comes about because in the ultrastatic setting, the metric $h_{\alpha\beta}$ is Riemannian, whereas here it is the spatial part of the Lorentzian metric (cf. (10.1), (10.5) and (A.1)).

$$\overset{(A.5)}{=} -\frac{1}{2} \int_0^1 d\tau \int_N d^3\mathbf{y}\, \tau\, \boldsymbol{\xi}_\alpha\, \partial_\alpha V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\, \mathcal{L}[\boldsymbol{\xi}]$$

$$= -\frac{1}{2} \int_N d^3\mathbf{y} \int_0^1 d\tau\, \tau\, \frac{d}{d\tau} V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\, \mathcal{L}[\boldsymbol{\xi}]\,. \tag{A.11}$$

Integrating by parts in τ gives the result. $\qquad\square$

We next expand the obtained formula for $\delta\ell$ asymptotically near infinity and explain our findings.

$$\delta\ell(\mathbf{x})$$

$$= -\frac{1}{2} \int_N d^3\mathbf{y} \left(\int_0^1 \left((\tau-1)\, \boldsymbol{\xi}_\alpha \partial_\alpha V|_{\mathbf{x}} + \frac{1}{2}\, (\tau^2-1)\, \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\beta \partial_{\alpha\beta} V|_{\mathbf{x}} \right) \right) \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\!\left(\frac{M\delta^3}{|\mathbf{x}|^4} \right)$$

$$= \frac{1}{2} \int_N d^3\mathbf{y} \left(\frac{1}{2}\, \boldsymbol{\xi}_\alpha \partial_\alpha V|_{\mathbf{x}} + \frac{1}{3}\, \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\beta \partial_{\alpha\beta} V|_{\mathbf{x}} \right) \mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\!\left(\frac{M\delta^3}{|\mathbf{x}|^4} \right) \tag{A.12}$$

(where δ is again the range of the Lagrangian as introduced after (2.18)). Carrying out the $\mathbf{y}$-integration with the help of (5.15), we obtain

$$\delta\ell(\mathbf{x}) = \frac{1}{18}\, \delta^2\, \mathfrak{s}_2\, \Delta_{\mathbb{R}^3} V|_{\mathbf{x}} + \mathcal{O}\!\left(\frac{M\delta^3}{|\mathbf{x}|^4} \right)$$

Now we can make use of the fact that the potential V in (A.2) is harmonic to conclude that

$$\delta\ell(\mathbf{x}) = \mathcal{O}\!\left(\frac{M\delta^3}{|\mathbf{x}|^4} \right). \tag{A.13}$$

This formula shows that the EL equations are indeed satisfied asymptotically at infinity. This is all we need in order to ensure that the total mass is well-defined. The critical reader might object that the EL equations should hold exactly, not only asymptotically at infinity. In order to understand how the error terms comes about, we first note that the error term in (A.13) includes the leading contributions of curvature, which scale like the Riemann tensor,

$$\mathrm{Riem}(\mathbf{x}) \sim \partial^2 g(\mathbf{x}) \sim \frac{M}{|\mathbf{x}|^3} + \mathcal{O}\!\left(\frac{1}{|\mathbf{x}|^4} \right).$$

In order to describe also the contributions of order $\mathcal{O}(|\mathbf{x}|^{-4})$ in the EL equations, one would have to take into account how the curvature tensor enters the Lagrangian. This goes beyond the scope of the present analysis, where we took the simplified ansatz for the linearized Lagrangian (A.7) as our starting point.

A.3 Description with jets

We next reformulate the previous findings in the jet formalism. To this end, we introduce the jet derivatives

$$D_{1,\mathbf{v}}\mathcal{L}(\mathbf{x},\mathbf{y}) := \frac{1}{4} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, h_{\alpha\beta}\big|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \frac{\partial}{\partial \mathbf{x}_\beta}\big(\boldsymbol{\xi}_\alpha\, \mathcal{L}(\mathbf{x},\mathbf{y}) \big) \tag{A.14}$$

$$D_{2,\mathbf{v}}\mathcal{L}(\mathbf{x},\mathbf{y}) := -\frac{1}{4} \int_{-\infty}^{\infty} \epsilon(1-\tau)\, d\tau\, h_{\alpha\beta}\big|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \frac{\partial}{\partial \mathbf{y}_\beta}\big(\boldsymbol{\xi}_\alpha\, \mathcal{L}(\mathbf{x},\mathbf{y}) \big). \tag{A.15}$$

This jet is constructed such that

$$\left(D_{1,\mathbf{v}} + D_{2,\mathbf{v}}\right)\mathcal{L}(\mathbf{x}, \mathbf{y}) = \delta\mathcal{L}(\mathbf{x}, \mathbf{y})$$

with the variation of the static Lagrangian as computed in (A.10). Similar as explained after (10.18) for ultrastatic spacetimes, we refer to $\mathfrak{v}$ as the *diffeomorphism jet*. Before going on, we point out that this diffeomorphism jet is quite different from the diffeomorphism jet in an ultrastatic spacetime if for the metric $h_{\alpha\beta}$ in (10.16) and (10.17) is chosen as the spatial part of the Schwarzschild metric. In other words, by comparing the diffeomorphism term in the Schwarzschild spacetime with that in ultrastatic spacetimes, one sees that the metric component g_{00} of the Schwarzschild metric has an important effect on the static Lagrangian. This is quite different from mathematical relativity, where the ADM mass of the Schwarzschild spacetime is described purely in terms of the induced Riemannian metric.

Let us verify that the EL equations are satisfied in the vacuum if we test with the jet $\mathfrak{v} = (0, \mathbf{v})$.

Lemma A.2 *The diffeomorphism jet (A.14) satisfies for all $\mathbf{x} \in N$ the equation*

$$\int_N D_{1,\mathbf{v}}\mathcal{L}(\mathbf{x}, \mathbf{y})\, d^3\mathbf{y} = 0\,.$$

Proof By direct computation, we obtain

$$\int_N D_{1,\mathbf{v}}\mathcal{L}(\mathbf{x}, \mathbf{y})\, d^3\mathbf{y} = -\frac{1}{4}\int_N d^3\mathbf{y} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, h_{\alpha\beta}\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}} \frac{\partial}{\partial\mathbf{y}_\beta}\left(\boldsymbol{\xi}_\alpha\, \mathcal{L}[\boldsymbol{\xi}]\right)$$

$$= \frac{1}{4}\int_N d^3\mathbf{y} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, \tau\, \partial_\beta h_{\alpha\beta}\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\, \boldsymbol{\xi}_\alpha\, \mathcal{L}[\boldsymbol{\xi}]$$

$$\overset{(A.5)}{=} -\frac{1}{4}\int_N d^3\mathbf{y} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, \tau\, \frac{d}{d\tau} V\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\, \mathcal{L}[\boldsymbol{\xi}]$$

$$= \frac{1}{4}\int_N d^3\mathbf{y} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, V\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\, \mathcal{L}[\boldsymbol{\xi}]\,.$$

Applying again the transformations (10.20) and (10.21), the integrand flips sign. Therefore, the integral vanishes by symmetry. □

Expressing the result of Lemma A.1 with jets gives a connection to the linearized field operator.

Lemma A.3 *The diffeomorphism jet $\mathbf{v}$ defined by (A.14) and (A.15) satisfies for all $\mathbf{x} \in N$ the equation*

$$\int_N \left(D_{1,\mathbf{v}} + D_{2,\mathbf{v}}\right)\mathcal{L}(\mathbf{x}, \mathbf{y})\, d^3\mathbf{y} = \frac{1}{2}\int_N d^3\mathbf{y}\left(\int_0^1 V\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\, d\tau - V(\mathbf{y})\right)\mathcal{L}[\boldsymbol{\xi}]\,. \quad \text{(A.16)}$$

Proof Noting that

$$\int_N \left(D_{1,\mathbf{v}} + D_{2,\mathbf{v}}\right)\mathcal{L}(\mathbf{x}, \mathbf{y})\, d^3\mathbf{y}$$

$$= -\frac{1}{2}\int_0^1 d\tau \int_N d^3\mathbf{y}\, h_{\alpha\beta}\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}} \frac{\partial}{\partial\mathbf{y}_\beta}\left(\boldsymbol{\xi}_\alpha\, \mathcal{L}[\boldsymbol{\xi}]\right)$$

and comparing with (A.11), we can proceed as in the proof of Lemma A.1. □

In order to better understand this result, it is useful to expand the potential on the right side of (A.16) in power of $\boldsymbol{\xi}$. Proceeding as in (A.12), we obtain

$$\int_N \left(D_{1,\mathbf{v}} + D_{2,\mathbf{v}}\right)\mathcal{L}(\mathbf{x}, \mathbf{y})\, d^3\mathbf{y} = \frac{1}{18}\, \delta^2\, \mathfrak{s}_2\, \Delta_{\mathbb{R}^3} V|_{\mathbf{x}} + \mathcal{O}\left(\frac{M\delta^3}{|\mathbf{x}|^4}\right).$$

The right side vanishes because V is harmonic. In this way, we get a direct connection between the linearized field equations and Newton's law of gravitation.

A.4 Computation of the total mass

We now compute the surface layer integral

$$\mathfrak{M}(R_0) = \int_{R<R_0} d^3\mathbf{x} \int_{r>R_0} d^3\mathbf{y}\, \left(D_{1,\mathbf{v}} - D_{2,\mathbf{v}}\right)\mathcal{L}(\mathbf{x}, \mathbf{y}).$$

Proposition A.4 *For any radius $R_0 > 0$,*

$$\mathfrak{M}(R_0) = \frac{4\pi}{9}\, \delta^2\, \mathfrak{s}_2\, M + \mathcal{O}\left(\frac{M\delta^3}{R_0}\right).$$

Proof Using the anti-symmetry of the integrand together with Lemma A.2, we obtain

$$\mathfrak{M}(R_0) = -\int_{R<R_0} d^3\mathbf{x} \int_N d^3\mathbf{y}\, \left(D_{1,\mathbf{v}} + D_{2,\mathbf{v}}\right)\mathcal{L}(\mathbf{x}, \mathbf{y}).$$

Applying Lemma A.3 gives

$$\mathfrak{M}(R_0) = -\frac{1}{2}\int_{R<R_0} d^3\mathbf{x} \int_N d^3\mathbf{y}\, \left(\int_0^1 V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\, d\tau - V(\mathbf{y})\right)\mathcal{L}[\boldsymbol{\xi}].$$

We now expand the integrand in powers of $\boldsymbol{\xi}$. Proceeding as in (A.12), we obtain

$$\mathfrak{M}(R_0) = \frac{1}{2}\int_{R<R_0} d^3\mathbf{x} \int_N d^3\mathbf{y}\, \left(\left(\frac{1}{2}\, \xi_\alpha \partial_\alpha V|_{\mathbf{x}} + \frac{1}{3}\, \xi_\alpha \xi_\beta \partial_{\alpha\beta} V|_{\mathbf{x}}\right)\mathcal{L}[\boldsymbol{\xi}] + \mathcal{O}\left(\frac{M\delta^3}{|\mathbf{x}|^3}\right)\right).$$

Carrying out the $\mathbf{y}$-integration using the formulas in (5.15), we obtain

$$\mathfrak{M}(R_0) = \frac{1}{18}\, \delta^2\, \mathfrak{s}_2 \int_{R<R_0} \Delta_{\mathbb{R}^3} V|_{\mathbf{x}}\, d^3\mathbf{x} + \mathcal{O}\left(\frac{M\delta^3}{R_0}\right). \tag{A.17}$$

Using (A.2), we get

$$\Delta_{\mathbb{R}^3} V|_{\mathbf{x}} = -2M\, \Delta_{\mathbb{R}^3}\left(\frac{1}{|\mathbf{x}|}\right) = 8\pi M\, \delta^3(\mathbf{x}).$$

Employing this formula in (A.17), we can carry out the $\mathbf{x}$-integral to obtain the result.

We remark that, as an alternative which avoids working with distributions, one can mollify V near the origin. Then the integral in (A.17) can be evaluated with the help of the Gauss divergence theorem to obtain

$$\mathfrak{M}(R_0) = \frac{1}{18}\, \delta^2\, \mathfrak{s}_2 \int_{R=R_0} \hat{\mathbf{x}}_\alpha \partial_\alpha V|_{\mathbf{x}}\, d\mu_R(\mathbf{x}) + \mathcal{O}\left(\frac{M\delta^3}{R_0}\right)$$

$$= \frac{1}{18}\, \delta^2\, \mathfrak{s}_2\, 4\pi R^2\, \frac{2M}{R^2} + \mathcal{O}\left(\frac{M\delta^3}{R_0}\right),$$

which again gives the result. We finally point out that this mollification method is also helpful for clarifying the error term in (A.17). Namely, mollifying the potential such that it vanishes inside a ball of radius $R_0/2$, it becomes clear that integrating the error term $\mathcal{O}(1/|\mathbf{x}|^3)$ over the annulus $R_0/2 < |\mathbf{x}| < R_0$ indeed gives the error term $\mathcal{O}(1/R_0)$.   □

A.5 Analysis of alignments

In this section we shall compute the first two summands in (5.6) in the Schwarzschild geometry and complete the proof of Theorem 5.9.

Lemma A.5 *In Schwarzschild coordinates, the alignment vector field $A^{(0)}$ defined in (5.6) has the form*

$$A_\alpha^{(0)}(x) = \frac{1}{24}\,\delta^2\,\mathfrak{s}2\,\partial_\alpha V(x) \; + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big).$$

Proof Using (A.14) we obtain

$$B_\alpha(\mathbf{x}, \boldsymbol{\xi}) := \xi_\alpha\, D_{1,\mathbf{v}}\mathcal{L}\Big(\zeta - \frac{\xi}{2}, \zeta + \frac{\xi}{2}\Big)$$

$$= -\frac{1}{4}\,\xi_\alpha \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, h_{\beta\gamma}\big|_{\zeta+(\tau-\frac{1}{2})\xi}\, \frac{\partial}{\partial\xi_\beta}\big(\xi_\gamma\, \mathcal{L}[\boldsymbol{\xi}]\big)\,.$$

We take the divergence,

$$\frac{\partial}{\partial\zeta_\alpha} B_\alpha(\mathbf{x}, \boldsymbol{\xi}) = -\frac{1}{4}\,\xi_\alpha \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, \partial_\alpha h_{\beta\gamma}\big|_{\zeta+(\tau-\frac{1}{2})\xi}\, \frac{\partial}{\partial\xi_\beta}\big(\xi_\gamma\, \mathcal{L}[\boldsymbol{\xi}]\big)$$

$$= -\frac{1}{4}\int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, \frac{d}{d\tau} h_{\beta\gamma}\big|_{\zeta+(\tau-\frac{1}{2})\xi}\, \frac{\partial}{\partial\xi_\beta}\big(\xi_\gamma\, \mathcal{L}[\boldsymbol{\xi}]\big) = \frac{1}{2} h_{\beta\gamma}\big|_{\zeta-\frac{1}{2}\xi}\, \frac{\partial}{\partial\xi_\beta}\big(\xi_\gamma\, \mathcal{L}[\boldsymbol{\xi}]\big)\,.$$

Integrating over $\boldsymbol{\xi}$ and using (5.6) gives

$$\frac{\partial}{\partial\zeta_\alpha} A_\alpha^{(0)}(\mathbf{x}) = \frac{1}{2}\int_N h_{\beta\gamma}\big|_{\zeta-\frac{1}{2}\xi}\, \frac{\partial}{\partial\xi_\beta}\big(\xi_\gamma\, \mathcal{L}[\boldsymbol{\xi}]\big)\, d^3\xi$$

$$= \frac{1}{4}\int_N \partial_\beta h_{\beta\gamma}\big|_{\zeta-\frac{1}{2}\xi}\, \xi_\gamma\, \mathcal{L}[\boldsymbol{\xi}]\, d^3\xi \overset{(A.5)}{=} -\frac{1}{4}\int_N \partial_\alpha V\big|_{\zeta-\frac{1}{2}\xi}\, \xi_\alpha\, \mathcal{L}[\boldsymbol{\xi}]\, d^3\xi\,.$$

We now expand the potential in powers of $\boldsymbol{\xi}$,

$$\frac{\partial}{\partial\zeta_\alpha} A_\alpha^{(0)}(\mathbf{x}) = -\frac{1}{4}\int_N \partial_\alpha V(\zeta)\,\xi_\alpha\, \mathcal{L}[\boldsymbol{\xi}]\, d^3\xi$$

$$+ \frac{1}{8}\int_N \partial_{\alpha\beta} V(\zeta)\,\xi_\alpha\,\xi_\beta\, \mathcal{L}[\boldsymbol{\xi}]\, d^3\xi + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big) \qquad (A.18)$$

$$\overset{(5.15)}{=} \frac{1}{24}\,\delta^2\,\mathfrak{s}2\,\Delta_{\mathbb{R}^3} V(\zeta)\, d^3\xi + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)\,. \qquad (A.19)$$

Applying the Gauss divergence theorem and using spherical symmetry, we obtain the result.

We finally remark that, in order to integrate the error term over the ball to obtain (A.18) in a clean way, exactly as explained at the end of the proof of Lemma A.4, one can again smoothen V such that it vanishes inside the ball of radius $|x|/2$.   □

Lemma A.6 *In Schwarzschild coordinates, the vector field $A^{(1)}$ defined in (5.6) has the form*

$$A_\alpha^{(1)}(x) = \frac{1}{72}\, \delta^2\, \mathfrak{s}_2\, \partial_\alpha V(x) + \mathcal{O}\left(\frac{M\delta^3}{|x|^3}\right).$$

Proof Evaluating (5.6) for $k = 1$ gives

$$
\begin{aligned}
A_\alpha^{(1)}(\zeta) &= \frac{1}{4\,3!} \int_N \xi_\alpha \left(\xi\frac{\partial}{\partial\zeta}\right)^2 D_{1,\mathsf{v}}\mathcal{L}\left(\zeta - \frac{\xi}{2}, \zeta + \frac{\xi}{2}\right) d^3\xi \\
&= -\frac{1}{4}\frac{1}{4\,3!} \int_N d^3\xi\, \xi_\alpha \left(\xi\frac{\partial}{\partial\zeta}\right)^2 \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, h_{\beta\gamma}\big|_{\zeta+(\tau-\frac{1}{2})\xi}\, \frac{\partial}{\partial\xi_\beta}\big(\xi_\gamma\,\mathcal{L}[\xi]\big) \\
&= -\frac{1}{96} \int_N d^3\xi\, \xi_\alpha \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau\, \frac{d^2}{d\tau^2}h_{\beta\gamma}\big|_{\zeta+(\tau-\frac{1}{2})\xi}\, \frac{\partial}{\partial\xi_\beta}\big(\xi_\gamma\,\mathcal{L}[\xi]\big) \\
&= \frac{1}{48} \int_N \xi_\alpha\, \xi_\delta\, \partial_\delta h_{\beta\gamma}\big|_{\zeta-\frac{1}{2}\xi}\, \frac{\partial}{\partial\xi_\beta}\big(\xi_\gamma\,\mathcal{L}[\xi]\big)\, d^3\xi \\
&= -\frac{1}{48} \int_N \xi_\delta\, \partial_\delta h_{\alpha\gamma}\big|_{\zeta-\frac{1}{2}\xi}\, \xi_\gamma\, \mathcal{L}[\xi]\, d^3\xi \\
&\quad -\frac{1}{48} \int_N \xi_\alpha\, \partial_\beta h_{\beta\gamma}\big|_{\zeta-\frac{1}{2}\xi}\, \xi_\gamma\, \mathcal{L}[\xi]\, d^3\xi \\
&\quad +\frac{1}{96} \int_N \xi_\alpha\, \xi_\delta\, \partial_{\delta\beta} h_{\beta\gamma}\big|_{\zeta-\frac{1}{2}\xi}\, \xi_\gamma\, \mathcal{L}[\xi]\, d^3\xi \\
&= -\frac{1}{24} \int_N \xi_\alpha\, \partial_\beta h_{\beta\gamma}\big|_{\zeta-\frac{1}{2}\xi}\, \xi_\gamma\, \mathcal{L}[\xi]\, d^3\xi + \mathcal{O}\left(\frac{M\delta^3}{|x|^3}\right) \\
&\overset{(5.15)}{=} -\frac{1}{72} \int_N \partial_\gamma h_{\alpha\gamma}\big|_{\zeta}\, |\xi|^2\, \mathcal{L}[\xi]\, d^3\xi + \mathcal{O}\left(\frac{M\delta^3}{|x|^3}\right) \\
&\overset{(A.5)}{=} \frac{1}{72}\, \delta^2\, \mathfrak{s}_2\, \partial_\alpha V(\zeta) + \mathcal{O}\left(\frac{M\delta^3}{|x|^3}\right).
\end{aligned}
$$

This concludes the proof. $\qquad\square$

Lemma A.7 *For an inner solution $\mathsf{u} = (\operatorname{div}\boldsymbol{u}, \boldsymbol{u})$,*

$$A_\alpha^{(1)}(\zeta) = \frac{1}{72}\, \delta^2\, \mathfrak{s}_2\, \Delta_{\mathbb{R}^3} u_\alpha(\zeta) + \frac{1}{36}\, \delta^2\, \mathfrak{s}_2\, \partial_\alpha \operatorname{div}\boldsymbol{u}(\zeta) + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3}\right).$$

Proof We again evaluate (5.6) for $k = 1$,

$$
\begin{aligned}
A_\alpha^{(1)}(\zeta) &= \frac{1}{24} \int_N \xi_\alpha \left(\xi\frac{\partial}{\partial\zeta}\right)^2 \nabla_{1,\mathsf{u}}\mathcal{L}\left(\zeta - \frac{\xi}{2}, \zeta + \frac{\xi}{2}\right) d^3\xi \\
&= \frac{1}{24} \int_N \xi_\alpha \left(\xi\frac{\partial}{\partial\zeta}\right)^2 \left(\operatorname{div}\boldsymbol{u}(x) + u_\beta(x)\frac{\partial}{\partial x_\beta}\right)\mathcal{L}[\xi]\, d^3\xi \\
&= \frac{1}{24} \int_N \xi_\alpha\, \xi_\gamma \xi_\delta\, \partial_{\gamma\delta}\operatorname{div}\boldsymbol{u}(x)\, \mathcal{L}[\xi]\, d^3\xi \\
&\quad -\frac{1}{24} \int_N \xi_\alpha\, \xi_\gamma \xi_\delta\, \partial_{\gamma\delta}u_\beta(x)\, \frac{\partial}{\partial\xi_\beta}\mathcal{L}[\xi]\, d^3\xi \\
&= -\frac{1}{24} \int_N \xi_\alpha\, \xi_\gamma \xi_\delta\, \partial_{\gamma\delta}u_\beta(\zeta)\, \frac{\partial}{\partial\xi_\beta}\mathcal{L}[\xi]\, d^3\xi + \mathcal{O}\left(\frac{\delta^3}{\ell_{\text{macro}}^3}\right)
\end{aligned}
$$

$$= \frac{1}{24} \int_N \boldsymbol{\xi}_\gamma \boldsymbol{\xi}_\delta \, \partial_{\gamma\delta} \mathbf{u}_\alpha(\boldsymbol{\zeta}) \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\boldsymbol{\xi}$$

$$+ \frac{1}{12} \int_N \boldsymbol{\xi}_\alpha \, \boldsymbol{\xi}_\delta \, \partial_{\gamma\delta} \mathbf{u}_\gamma(\boldsymbol{\zeta}) \, \mathcal{L}[\boldsymbol{\xi}] \, d^3\boldsymbol{\xi} + \mathcal{O}\Big(\frac{\delta^3}{\ell_{\text{macro}}^3}\Big).$$

Expanding in powers of ξ gives the result.   □

Proof of Theorem 5.9 We first compute different contributions to the total mass using the results of Lemmas A.5 and A.6 in the formula of Corollary 5.2.

$$\mathfrak{M}^{(0)}(R_0) := \int_{r=R_0} A_\alpha^{(0)}(\mathbf{x}) \, \hat{\mathbf{x}}_\alpha \, d\mu_{R_0}(\mathbf{x}) = \frac{1}{24} \, \delta^2 \, \mathfrak{s}_2 \int_N \Delta_{\mathbb{R}^3} V(\mathbf{x}) \, d^3\mathbf{x} + \mathcal{O}\Big(\frac{M\delta^3}{R_0}\Big)$$

$$= -\frac{M}{12} \, \delta^2 \, \mathfrak{s}_2 \int_N \Delta_{\mathbb{R}^3}\Big(\frac{1}{|\mathbf{x}|}\Big) \, d^3\mathbf{x} + \mathcal{O}\Big(\frac{M\delta^3}{R_0}\Big)$$

$$= -\frac{M}{12} \, \delta^2 \, \mathfrak{s}_2 \int_N (-4\pi) \, \delta^3(\mathbf{x}) \, d^3\mathbf{x} + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big) = \frac{\pi}{3} \, \delta^2 \, \mathfrak{s}_2 \, M + \mathcal{O}\Big(\frac{M\delta^3}{R_0}\Big)$$

$$\mathfrak{M}^{(1)}(R_0) := \int_{r=R_0} A_\alpha^{(1)}(\mathbf{x}) \, \hat{\mathbf{x}}_\alpha \, d\mu_{R_0}(\mathbf{x}) = \frac{1}{72} \, \delta^2 \, \mathfrak{s}_2 \int_N \Delta_{\mathbb{R}^3} V(\mathbf{x}) \, d^3\mathbf{x} + \mathcal{O}\Big(\frac{M\delta^3}{R_0}\Big)$$

$$= -\frac{M}{36} \, \delta^2 \, \mathfrak{s}_2 \int_N \Delta_{\mathbb{R}^3}\Big(\frac{1}{|\mathbf{x}|}\Big) \, d^3\mathbf{x} + \mathcal{O}\Big(\frac{M\delta^3}{R_0}\Big)$$

$$= -\frac{M}{36} \, \delta^2 \, \mathfrak{s}_2 \int_N (-4\pi) \, \delta^3(\mathbf{x}) \, d^3\mathbf{x} + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big) = \frac{\pi}{9} \, \delta^2 \, \mathfrak{s}_2 \, M + \mathcal{O}\Big(\frac{M\delta^3}{R_0}\Big).$$

In order to choose linear alignment, we must add an inner solution $\mathfrak{u}$. More precisely, comparing Lemma A.5 with Lemma 5.5, one sees that we must choose

$$\mathfrak{u}_\alpha(\mathbf{x}) = -\frac{1}{24 \, \mathfrak{s}} \, \delta^2 \, \mathfrak{s}_2 \, \partial_\alpha V(\mathbf{x}) + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big).$$

Then $A^{(0)}$ and therefore also $\mathfrak{M}^{(0)}$ vanish. According to Lemma A.7, the inner solution changes $A^{(1)}$ only to the order $\mathcal{O}(\delta^3/|\mathbf{x}|^3)$. Therefore, the above formula for $\mathfrak{M}^{(1)}(R_0)$ still holds with alignment, giving (5.24).

It remains to show that the higher order contributions can be absorbed into the error term. More precisely, let us show that

$$\mathfrak{M}^{(k)}(R_0) = \mathcal{O}\Big(\frac{M\delta^4}{R_0^2}\Big) \qquad \text{for } k = 2, 3, \dots . \tag{A.20}$$

To this end, we note that, according to (5.6), these contributions involve at least five factors $\boldsymbol{\xi}$ and at least four $\boldsymbol{\zeta}$-derivatives. Proceeding as in the proofs of Lemmas A.5 and A.6, using the specific form of the jets, one can integrate by parts once in $\boldsymbol{\xi}$ and once in τ. After doing so, at least four factors $\boldsymbol{\xi}$ and at least three β-derivatives remain. Since the $\boldsymbol{\zeta}$-derivatives act on the potential V, we obtain a scaling factor R_0^{-4}. Taking into account that the surface integral gives a scaling factor R_0^2, we obtain (A.20).   □

A.6 Spatially isotropic coordinates

The form of the linearized Schwarzschild metric (A.3) becomes somewhat simpler by transforming to spatially isotropic coordinates (see also [35, Exercise 4 in §100] or [41, Problem 1

in Section 6]). On the other hand, the volume form is no longer preserved, giving rise to jets with scalar components. This is why we decided to carry out all computations in Schwarzschild coordinates. Nevertheless, we now explain how the linearized Schwarzschild metric in spatially isotropic coordinates can be described with jets. This formulation might be of advantage for future computations in asymptotically flat spacetimes.

In order to transform to spatially isotropic coordinates, one introduces the new radial variable $\tilde{r}$ by

$$r = \left(1 + \frac{M}{2\tilde{r}}\right)^2 \tilde{r}\,, \qquad dr = \left(1 + \frac{M}{2\tilde{r}}\right)\left(1 - \frac{M}{2\tilde{r}}\right) d\tilde{r}\,. \tag{A.21}$$

Then the line element becomes

$$ds^2 = \left(\frac{1 - \dfrac{M}{2\tilde{r}}}{1 + \dfrac{M}{2\tilde{r}}}\right)^2 dt^2 - \left(1 + \frac{M}{2\tilde{r}}\right)^4 \left(d\tilde{r}^2 + \tilde{r}^2 d\vartheta^2 + \tilde{r}^2 \sin^2\vartheta\, d\varphi^2\right).$$

Again using the abbreviation (A.2), the linearized metric is given by

$$h'_{00}(x) = V(\mathbf{x})\,, \quad h'_{\alpha\beta} = V(\mathbf{x})\,\delta_{\alpha\beta} \tag{A.22}$$

As already explained for ultrastatic spacetimes in Sect. 10.2 (see (10.6) and (10.7)), in the formalism of linearized gravity, one has the freedom to perform infinitesimal coordinate transformations

$$x'^i = x^i + \zeta^i(\mathbf{x}) \tag{A.23}$$

This transforms the linearized metric according to

$$h'_{jk} = h_{jk} - \partial_j \zeta_k - \partial_k \zeta_j\,. \tag{A.24}$$

This freedom can be understood as the gauge freedom of the gravitational field. It is most convenient to use this gauge freedom to arrange the gauge condition

$$\partial^j h_{jk} = \frac{1}{2}\,\partial_k h \qquad \text{with} \qquad h := h^i_i\,.$$

This gauge condition means that we are working in *harmonic coordinates* (for details see [35, §107]). The spatially isotropic coordinates are indeed harmonic, as one sees directly from the computation (see also [35, §105])

$$h'(x) = h'_{00}(x) - \sum_{\alpha=1}^{3} h'_{\alpha\alpha}(x) = -2V(\mathbf{x}) = 2\,\frac{2M}{R}$$

$$\partial^j h'_{jk}(x) = -\sum_{\alpha=1}^{3} \partial_\alpha h'_{\alpha k}(x) = \delta_{\alpha k}\,\partial_\alpha\!\left(\frac{2M}{R}\right) = \partial_k\!\left(\frac{2M}{R}\right) = \frac{1}{2}\,\partial_k h'(x)\,.$$

The transition from Schwarzschild coordinates to spatially isotropic coordinates is described by the infinitesimal coordinate transformation (A.23) with

$$\zeta_0(x) = 0 \qquad \text{and} \qquad \zeta_\alpha(x) = M\,\frac{x_\alpha}{R}\,, \tag{A.25}$$

as one sees by linearizing (A.21) or, alternatively, from (A.24) using that

$$\partial_\alpha \zeta_\beta(x) + \partial_\beta \zeta_\alpha(x) = \frac{2M}{R}\left(\delta_{\alpha\beta} - \frac{x_\alpha x_\beta}{R^2}\right) = -V(\mathbf{x})\left(\delta_{\alpha\beta} - \hat{\mathbf{x}}_\alpha \hat{\mathbf{x}}_\beta\right). \tag{A.26}$$

The first variation of the Lagrangian can again be described with the help of (A.7) and (A.8), (A.9). Using (A.22), we obtain

$$\delta\mathcal{L}(\mathbf{x}, \mathbf{y}) = -\frac{1}{2}\int_0^1 d\tau\, h'_{00}\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\, \mathcal{L}[\xi] - \frac{1}{2}\int_0^1 d\tau\, h'_{\alpha\beta}\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\, \xi_\alpha\, \frac{\partial}{\partial \mathbf{y}_\beta}\mathcal{L}[\xi]$$

$$= -\frac{1}{2}\int_0^1 d\tau\, V\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\, \mathcal{L}[\xi] - \frac{1}{2}\int_0^1 d\tau\, V\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\, \xi_\alpha\, \frac{\partial}{\partial \mathbf{y}_\alpha}\mathcal{L}[\xi]$$

$$= -\frac{1}{2}\int_0^1 d\tau\, V\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\left(1 + \xi_\alpha\, \frac{\partial}{\partial \mathbf{y}_\alpha}\right)\mathcal{L}[\xi]\,.$$

Consequently, we introduce the jet derivatives

$$D_{1,\mathbf{v}'}\mathcal{L}(\mathbf{x}, \mathbf{y}) := -\frac{1}{4}\int_{-\infty}^\infty \epsilon(\tau)\, d\tau\, V\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\left(1 - \xi_\alpha\, \frac{\partial}{\partial \mathbf{x}_\alpha}\right)\mathcal{L}[\xi] \tag{A.27}$$

$$D_{2,\mathbf{v}'}\mathcal{L}(\mathbf{x}, \mathbf{y}) := -\frac{1}{4}\int_{-\infty}^\infty \epsilon(1-\tau)\, d\tau\, V\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\left(1 + \xi_\alpha\, \frac{\partial}{\partial \mathbf{y}_\alpha}\right)\mathcal{L}[\xi]\,. \tag{A.28}$$

Moreover, we must take into account that the volume form changes. This gives rise to a scalar component of the jet given by

$$b'(\mathbf{x}) = \delta\sqrt{|\det g'|} = \frac{1}{2}\, h'(\mathbf{x}) = -V(\mathbf{x})\,. \tag{A.29}$$

In the next lemma we verify in detail that the jets $\mathfrak{v}$ and $\mathfrak{v}'$ differ by an inner solution describing the infinitesimal coordinate transformation (A.25).

Lemma A.8 *The diffeomorphism jet $\mathfrak{v} = (0, \mathbf{v})$ (with $\mathbf{v}$ according to (A.14) and (A.15)) and the jet $\mathfrak{v}' = (b, \mathbf{v}')$ (with b according to (A.29) and $\mathbf{v}'$ as in (A.27) and (A.28)) satisfy the relation*

$$\mathfrak{v}' = \mathfrak{v} + \mathfrak{u}\,,$$

where $\mathfrak{u}$ is the inner solution corresponding to the vector field $\boldsymbol{\zeta}$ in (A.25), i.e.

$$\mathfrak{u} = (\mathrm{div}\,\boldsymbol{\zeta}, \boldsymbol{\zeta}) = \big(-V(\mathbf{x}), M\hat{\mathbf{x}}\big)\,. \tag{A.30}$$

Proof We first compute the divergence of $\boldsymbol{\zeta}$,

$$\mathrm{div}\,\boldsymbol{\zeta}(x) = \partial_\alpha\left(M\,\frac{x_\alpha}{R}\right) = \frac{M}{R}\,(-1+3) = \frac{2M}{R} = -V(\mathbf{x})\,,$$

proving the last equality in (A.30).

It suffices to prove that

$$\big(\nabla_{1,\mathfrak{v}'} - D_{1,\mathbf{v}}\big)\mathcal{L}(\mathbf{x}, \mathbf{y}) = \nabla_{1,\mathfrak{u}}\mathcal{L}(\mathbf{x}, \mathbf{y})\,, \tag{A.31}$$

because the corresponding relation for the $\mathbf{y}$-derivatives acting is obtained by the replacements $\mathbf{x} \leftrightarrow \mathbf{y}$, $\xi \to -\xi$ and $\tau \to 1 - \tau$. We first consider those terms on the left which involve derivatives of the Lagrangian. According to (A.27) and (A.14), they can be written as

$$A := \frac{1}{4}\int_{-\infty}^\infty \epsilon(\tau)\, d\tau\,\left(V\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\,\xi_\beta - h_{\alpha\beta}\big|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\,\xi_\alpha\right)\frac{\partial}{\partial \mathbf{x}_\alpha}\big(\mathcal{L}[\xi]\big)$$

$$\stackrel{(A.4)}{=} \frac{1}{4}\int_{-\infty}^\infty \epsilon(\tau)\, d\tau\,\left(V(\mathbf{z})\big(\delta_{\alpha\beta} - \hat{\mathbf{z}}_\alpha\,\hat{\mathbf{z}}_\beta\big)\right)\Big|_{\mathbf{z}=\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\,\xi_\alpha\, \frac{\partial}{\partial \mathbf{x}_\beta}\big(\mathcal{L}[\xi]\big)$$

$$\stackrel{(A.26)}{=} -\frac{1}{4} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau \, \left(\partial_\alpha \zeta_\beta + \partial_\beta \zeta_\alpha\right)\big|_{\tau \mathbf{y} + (1-\tau)\mathbf{x}} \, \boldsymbol{\xi}_\alpha \, \frac{\partial}{\partial \mathbf{x}_\beta}\left(\mathcal{L}[\boldsymbol{\xi}]\right).$$

Since the Lagrangian is spherically symmetric, its derivative $\partial_{\mathbf{x}_\beta}\mathcal{L}(x,y)$ is proportional to $\boldsymbol{\xi}_\beta$. Therefore,

$$\left(\partial_\alpha \zeta_\beta + \partial_\beta \zeta_\alpha\right)\big|_{\tau \mathbf{y} + (1-\tau)\mathbf{x}} \, \boldsymbol{\xi}_\alpha \, \frac{\partial}{\partial \mathbf{x}_\beta}\mathcal{L}[\boldsymbol{\xi}] = 2\,\boldsymbol{\xi}_\alpha (\partial_\alpha \zeta_\beta) \, \frac{\partial}{\partial \mathbf{x}_\beta}\mathcal{L}[\boldsymbol{\xi}].$$

We thus obtain

$$A = -\frac{1}{2} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau \, \frac{d}{d\tau} \zeta_\beta \big|_{\tau \mathbf{y} + (1-\tau)\mathbf{x}} \, \frac{\partial}{\partial \mathbf{x}_\beta}\mathcal{L}[\boldsymbol{\xi}] = \zeta_\beta(\mathbf{x}) \, \frac{\partial}{\partial \mathbf{x}_\beta}\mathcal{L}[\boldsymbol{\xi}] = D_{1,\mathbf{u}}\mathcal{L}(\mathbf{x},\mathbf{y}).$$

Taking into account the contributions to the left of (A.31) where the Lagrangian is not differentiated, we obtain from (A.29), (A.27) and (A.14),

$$\left(\nabla_{1,\mathfrak{v}'} - D_{1,\mathbf{v}}\right)\mathcal{L}(\mathbf{x},\mathbf{y}) = A - V(\mathbf{x})\,\mathcal{L}(\mathbf{x},\mathbf{y})$$
$$+ \frac{1}{4} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau \, V\big|_{\tau \mathbf{y} + (1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x},\mathbf{y}) - \frac{1}{4} \int_{-\infty}^{\infty} \epsilon(\tau)\, d\tau \, h_{\alpha\beta}\big|_{\tau \mathbf{y} + (1-\tau)\mathbf{x}} \, \delta_{\alpha\beta} \, \mathcal{L}(\mathbf{x},\mathbf{y}))$$
$$\stackrel{(A.4)}{=} A - V(\mathbf{x})\,\mathcal{L}(\mathbf{x},\mathbf{y}) = \left(-V(\mathbf{x}) + D_{1,\zeta}\right)\mathcal{L}(\mathbf{x},\mathbf{y}) \stackrel{(A.30)}{=} \nabla_{1,\mathbf{u}}\mathcal{L}(\mathbf{x},\mathbf{y}).$$

This concludes the proof. $\qquad \square$

A.7 A spatial integral related to the total mass

In [30, Section 6.5] the total mass was expressed in terms of a spatial integral,

$$\mathfrak{M} = 2\pi \lim_{R \to \infty} R^2\, \mathfrak{P}(R) \tag{A.32}$$

with

$$\mathfrak{P}(R) := \int_N (r - R)\left(D_{1,\mathbf{v}} - D_{2,\mathbf{v}}\right)\mathcal{L}(\mathbf{x},\mathbf{y})\, d^3\mathbf{y}. \tag{A.33}$$

In order to derive this representation, we employed a method used previously in [24, Lemma 5.5] and [13, Lemma 5.2] which consists in integrating the conserved surface layer integral over the radius and taking its mean. More precisely, using the abbreviation

$$A(r, r') = r^2\, r'^2 \int_{S^2} d\omega \int_{S^2} d\omega' \left(D_{1,\mathbf{v}} - D_{2,\mathbf{v}}\right)\mathcal{L}(\mathbf{x},\mathbf{y}),$$

we obtain the identity

$$\mathfrak{M}(R) = \frac{1}{L} \int_{R_{\min}}^{R} dr \int_{R}^{R+L} dr'\, (r' - R)\, A(r, r') \tag{A.34}$$

$$+ \frac{1}{L} \int_{R_{\min}}^{R} dr \int_{R+L}^{\infty} dr'\, L\, A(r, r') \tag{A.35}$$

$$+ \frac{1}{L} \int_{R}^{R+L} dr \int_{r}^{R+L} dr'\, (r' - r)\, A(r, r') \tag{A.36}$$

$$+ \frac{1}{L} \int_{R}^{R+L} dr \int_{R+L}^{\infty} dr'\, (L - r + R)\, A(r, r'). \tag{A.37}$$

In the limit $L \to \infty$, the summands (A.34) and (A.35) obviously vanish. Moreover, the inner integral (A.36) can be shown to converge to $2\pi r^2 \, \mathfrak{P}(r)$. Also taking the limit $R \to \infty$, we obtain the identity

$$\mathfrak{M} = 2\pi \lim_{r \to \infty} r^2 \, \mathfrak{P}(r) + \lim_{R \to \infty} \lim_{L \to \infty} \frac{1}{L} \int_R^{R+L} dr \int_{R+L}^\infty dr' \, (L - r + R) \, A(r, r') \quad (A.38)$$

If the remaining surface layer integral is bounded uniformly in L, the last summand vanishes, giving (A.32). This is indeed the case in the previous applications in [13, 24] as well as for the jet $\mathfrak{v}$ considered in [30]. However, for the jet $\mathfrak{v}$ considered here, the surface layer integral in (A.38) turns out to be linearly divergent in L, so that the last summand is non-zero. In simple terms, this can be understood from the fact that the jet $\mathfrak{v}$ described by (A.14) and (A.15) does not have the necessary decay conditions at infinity (in [30] this issue is avoided with a scaling argument, which however does not capture the higher orders in δ correctly; see [30, Proposition 6.3] and the corresponding footnote). In order to explain in detail why the last summand in (A.38) is indeed non-zero, we now compute $\mathfrak{P}(r)$.

Proposition A.9 *Evaluating the spatial integral (A.32) for $\mathbf{x}$ on a sphere of radius $R > 0$, the spatial integral (A.33) can be computed to obtain*

$$\mathfrak{P}(R) = \frac{M}{R^2} \, \delta^2 \, \mathfrak{s}_2 + \mathcal{O}\!\left(\frac{M\delta^3}{R^3}\right).$$

Comparing with the result of Proposition A.4, one sees explicitly that (A.38) is violated, because the prefactors are different.

The remainder of this appendix is devoted to the proof of Proposition A.9. First, from (A.14) and (A.15), we find

$$\left(D_{1,\mathbf{v}} - D_{2,\mathbf{v}}\right) \mathcal{L}(\mathbf{x}, \mathbf{y}) = -\frac{1}{4} \int_{-\infty}^\infty \epsilon(\tau) \, d\tau \, h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \frac{\partial}{\partial \mathbf{y}_\beta} \left(\boldsymbol{\xi}_\alpha \, \mathcal{L}[\boldsymbol{\xi}]\right)$$

$$+ \frac{1}{4} \int_{-\infty}^\infty \epsilon(1 - \tau) \, d\tau \, h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \frac{\partial}{\partial \mathbf{y}_\beta} \left(\boldsymbol{\xi}_\alpha \, \mathcal{L}[\boldsymbol{\xi}]\right)$$

$$= -\frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0\right) d\tau \, h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \frac{\partial}{\partial \mathbf{y}_\beta} \left(\boldsymbol{\xi}_\alpha \, \mathcal{L}[\boldsymbol{\xi}]\right).$$

As a consequence,

$$\mathfrak{P}(R) = -\frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0\right) d\tau \int_N d^3\mathbf{y} \, (r - R) \, h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \frac{\partial}{\partial \mathbf{y}_\beta} \left(\boldsymbol{\xi}_\alpha \, \mathcal{L}[\boldsymbol{\xi}]\right)$$

$$= \frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0\right) d\tau \int_N d^3\mathbf{y} \, \hat{\mathbf{y}}_\beta \, h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \boldsymbol{\xi}_\alpha \, \mathcal{L}[\boldsymbol{\xi}] \quad (A.39)$$

$$+ \frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0\right) d\tau \int_N d^3\mathbf{y} \, (r - R) \, \tau \, \partial_\beta h_{\alpha\beta}\big|_{\tau y + (1-\tau) x} \, \boldsymbol{\xi}_\alpha \, \mathcal{L}[\boldsymbol{\xi}]. \quad (A.40)$$

In (A.40) we can apply (A.5) and integrate by parts,

$$(A.40) = -\frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0\right) d\tau \int_N d^3\mathbf{y} \, (r - R) \, \tau \, \frac{d}{d\tau} V\big|_{\tau y + (1-\tau) x} \, \mathcal{L}[\boldsymbol{\xi}]$$

$$= \frac{1}{2} \int_N d^3\mathbf{y} \, (r - R) \, V(\mathbf{y}) \, \mathcal{L}[\boldsymbol{\xi}]$$

$$+ \frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, (r - R) \left. V \right|_{\tau \mathbf{y} + (1-\tau)\,\mathbf{x}} \mathcal{L}[\boldsymbol{\xi}] \,.$$

In order to simplify (A.39), we use that, according to (A.3) and (A.2),

$$\partial_\alpha \left(\frac{\mathbf{z}_\beta}{|\mathbf{z}|} \right) = \frac{\delta_{\alpha\beta}}{|\mathbf{z}|} - \frac{\mathbf{z}_\alpha \mathbf{z}_\beta}{|\mathbf{z}|^3}$$

$$h_{\alpha\beta}(\mathbf{z}) = -\frac{2M}{|\mathbf{z}|} \, \hat{\mathbf{z}}_\alpha \, \hat{\mathbf{z}}_\beta = -2M \frac{\mathbf{z}_\alpha \, \mathbf{z}_\beta}{|\mathbf{z}|^3}$$

$$= 2M \, \partial_\alpha \left(\frac{\mathbf{z}_\beta}{|\mathbf{z}|} \right) - 2M \frac{\delta_{\alpha\beta}}{|\mathbf{z}|} = -\partial_\alpha \big(\mathbf{z}_\beta \, V(\mathbf{z}) \big) + \delta_{\alpha\beta} \, V(\mathbf{z}) \,.$$

It follows that

$$(A.39) = -\frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \hat{\mathbf{y}}_\beta \, \frac{d}{d\tau} \big(\mathbf{z}_\beta \, V(\mathbf{z}) \big) \Big|_{\tau \mathbf{y} + (1-\tau)\,x} \mathcal{L}[\boldsymbol{\xi}]$$

$$+ \frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \hat{\mathbf{y}}_\alpha \, \boldsymbol{\xi}_\alpha \left. V \right|_{\tau \mathbf{y} + (1-\tau)\,x} \mathcal{L}[\boldsymbol{\xi}]$$

$$= \frac{1}{2} \int_N d^3\mathbf{y} \, \hat{\mathbf{y}}_\alpha \left(\mathbf{y}_\alpha \, V(\mathbf{y}) + \mathbf{x}_\alpha \, V(\mathbf{x}) \right) \mathcal{L}[\boldsymbol{\xi}]$$

$$+ \frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \hat{\mathbf{y}}_\alpha \, \boldsymbol{\xi}_\alpha \left. V \right|_{\tau \mathbf{y} + (1-\tau)\,x} \mathcal{L}[\boldsymbol{\xi}] \,.$$

Adding all the contributions, we obtain

$$\mathfrak{P}(R) = \frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, (r - R) \left. V \right|_{\tau \mathbf{y} + (1-\tau)\,x} \mathcal{L}[\boldsymbol{\xi}] \tag{A.41}$$

$$+ \frac{1}{2} \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \hat{\mathbf{y}}_\alpha \, \boldsymbol{\xi}_\alpha \left. V \right|_{\tau \mathbf{y} + (1-\tau)\,x} \mathcal{L}[\boldsymbol{\xi}] \tag{A.42}$$

$$+ \frac{1}{2} \int_N d^3\mathbf{y} \, (r - R) \, V(\mathbf{y}) \, \mathcal{L}[\boldsymbol{\xi}] \tag{A.43}$$

$$+ \frac{1}{2} \int_N d^3\mathbf{y} \, \hat{\mathbf{y}}_\alpha \left(\mathbf{y}_\alpha \, V(\mathbf{y}) + \mathbf{x}_\alpha \, V(\mathbf{x}) \right) \mathcal{L}[\boldsymbol{\xi}] \,. \tag{A.44}$$

In the next step we Taylor expand in powers of $\boldsymbol{\xi}$.

Lemma A.10

$$\mathfrak{P}(R) = \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} \left. V \right|_{\tau \mathbf{y} + (1-\tau)\,x} \mathcal{L}[\boldsymbol{\xi}] \tag{A.45}$$

$$+ \frac{3}{4} \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \left(\frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3} \right) \left. V \right|_{\tau \mathbf{y} + (1-\tau)\,x} \mathcal{L}[\boldsymbol{\xi}] \tag{A.46}$$

$$+ \mathfrak{s} \, |\mathbf{x}| \, V(\mathbf{x}) - \frac{1}{6\,|\mathbf{x}|} \, \delta^2 \, \mathfrak{s}_2 \, V(\mathbf{x}) + \mathcal{O} \left(\frac{M\delta^3}{|\mathbf{x}|^3} \right) . \tag{A.47}$$

Proof A direct computation yields

$$r - R = |\mathbf{y}| - |\mathbf{x}| = |\mathbf{x} + \boldsymbol{\xi}| - |\mathbf{x}| = \sqrt{\mathbf{x}^2 + 2\mathbf{x}\boldsymbol{\xi} + \boldsymbol{\xi}^2} - |\mathbf{x}|$$

$$= \frac{1}{2\,|\mathbf{x}|} \left(2\mathbf{x}\boldsymbol{\xi} + \boldsymbol{\xi}^2 \right) - \frac{1}{8\,|\mathbf{x}|^3} \left(2\mathbf{x}\boldsymbol{\xi} \right)^2 + \mathcal{O} \big(\boldsymbol{\xi}^3 \big)$$

$$= \frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} + \frac{\boldsymbol{\xi}^2}{2\,|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{2\,|\mathbf{x}|^3} + \mathcal{O}(\boldsymbol{\xi}^3)$$

$$\hat{\mathbf{y}}_\alpha\,\boldsymbol{\xi}_\alpha = \frac{\mathbf{y}\boldsymbol{\xi}}{|\mathbf{y}|} = \frac{(\mathbf{x}+\boldsymbol{\xi})\boldsymbol{\xi}}{|\mathbf{x}+\boldsymbol{\xi}|} = \frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} + \frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3} + \mathcal{O}(\boldsymbol{\xi}^3)$$

$$\hat{\mathbf{y}}_\alpha\,\mathbf{y}_\alpha = r = |\mathbf{x}| + (r - R)$$

$$= |\mathbf{x}| + \frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} + \frac{\boldsymbol{\xi}^2}{2\,|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{2\,|\mathbf{x}|^3} + \mathcal{O}(\boldsymbol{\xi}^3)$$

$$\hat{\mathbf{y}}_\alpha\,\mathbf{x}_\alpha = \hat{\mathbf{y}}_\alpha\,(\mathbf{y}_\alpha - \boldsymbol{\xi}_\alpha)$$

$$= |\mathbf{x}| + \frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} + \frac{\boldsymbol{\xi}^2}{2\,|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{2\,|\mathbf{x}|^3} - \Big(\frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} + \frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3}\Big) + \mathcal{O}(\boldsymbol{\xi}^3)$$

$$= |\mathbf{x}| - \frac{\boldsymbol{\xi}^2}{2\,|\mathbf{x}|} + \frac{(\mathbf{x}\boldsymbol{\xi})^2}{2\,|\mathbf{x}|^3} + \mathcal{O}(\boldsymbol{\xi}^3)\,.$$

Hence

$(A.41) + (A.42)$

$$= \Big(\int_1^\infty - \int_{-\infty}^0\Big)\,d\tau \int_N d^3\mathbf{y}\,\Big(\frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} + \frac{3}{4}\frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{3}{4}\frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3} + \mathcal{O}(\boldsymbol{\xi}^3)\Big)\,V|_{\tau\mathbf{y}+(1-\tau)\,\mathbf{x}}\,\mathcal{L}[\boldsymbol{\xi}]$$

$(A.43) + (A.44)$

$$= \frac{1}{2}\int_N d^3\mathbf{y}\,(r - R + \hat{\mathbf{y}}_\alpha\,\mathbf{y}_\alpha)\,V(\mathbf{y})\,\mathcal{L}[\boldsymbol{\xi}] + \frac{1}{2}\int_N d^3\mathbf{y}\,\hat{\mathbf{y}}_\alpha\,\mathbf{x}_\alpha\,V(\mathbf{x})\,\mathcal{L}[\boldsymbol{\xi}]$$

$$= \frac{1}{2}\int_N d^3\mathbf{y}\,\Big(|\mathbf{x}| + 2\frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} + \frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3}\Big)\,V(\mathbf{y})\,\mathcal{L}[\boldsymbol{\xi}]$$

$$+ \frac{1}{2}\int_N d^3\mathbf{y}\,\Big(|\mathbf{x}| - \frac{\boldsymbol{\xi}^2}{2\,|\mathbf{x}|} + \frac{(\mathbf{x}\boldsymbol{\xi})^2}{2\,|\mathbf{x}|^3}\Big)\,V(\mathbf{x})\,\mathcal{L}[\boldsymbol{\xi}] + + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)$$

$$= \mathfrak{s}\,|\mathbf{x}|\,V(\mathbf{x}) + \frac{1}{2}\,|\mathbf{x}|\int_N d^3\mathbf{y}\,\frac{1}{2}\,\boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\beta\,\partial_{\alpha\beta}V(\mathbf{x})\,\mathcal{L}[\boldsymbol{\xi}] + \int_N d^3\mathbf{y}\,\frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|}\,\boldsymbol{\xi}_\alpha\,\partial_\alpha V(\mathbf{x})\,\mathcal{L}[\boldsymbol{\xi}]$$

$$+ \frac{1}{2}\int_N d^3\mathbf{y}\,\Big(\frac{\boldsymbol{\xi}^2}{2\,|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{2\,|\mathbf{x}|^3}\Big)\,V(\mathbf{x})\,\mathcal{L}(\mathbf{x},\mathbf{y}) + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)$$

$$= \mathfrak{s}\,|\mathbf{x}|\,V(\mathbf{x}) + \frac{1}{12}\,\delta^2\,\mathfrak{s}_2\,|\mathbf{x}|\,\Delta_{\mathbb{R}^3}V(\mathbf{x}) + \frac{1}{3\,|\mathbf{x}|}\,\delta^2\,\mathfrak{s}_2\,\mathbf{x}_\alpha\partial_\alpha V(\mathbf{x})$$

$$+ \frac{1}{6\,|\mathbf{x}|}\,\delta^2\,\mathfrak{s}_2\,V(\mathbf{x}) + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)\,.$$

We now use that, away from the origin, the potential is harmonic and homogeneous of degree minus one, i.e.

$$\Delta_{\mathbb{R}^3}V(\mathbf{x}) = 0 \quad \text{and} \quad \mathbf{x}_\alpha\,\partial_\alpha V(\mathbf{x}) = -V(\mathbf{x})\,.$$

We thus obtain

$$(A.43) + (A.44) = \mathfrak{s}\,|\mathbf{x}|\,V(\mathbf{x}) - \frac{1}{6\,|\mathbf{x}|}\,\delta^2\,\mathfrak{s}_2\,V(\mathbf{x}) + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)\,.$$

Collecting all the terms gives the result.   □

Our next task is to compute the τ-integrals in $(A.45)$ and $(A.46)$. This will be accomplished in the following two lemmas.

Lemma A.11 *For any* $\mathbf{x}$ *on a sphere of radius* $R > 0$,

$$\left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} \, V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= -\mathfrak{s} \, |\mathbf{x}| \, V(\mathbf{x}) + \frac{1}{6} \, \delta^2 \, \mathfrak{s}_2 \, \frac{1}{|\mathbf{x}|} \, V(\mathbf{x}) + \mathcal{O}\left(\frac{M\delta^3}{|\mathbf{x}|^3} \right).$$

Proof For any $\mathbf{x} \in B_R(0)$, we compute the following divergence using integration by parts,

$$\frac{\partial}{\partial \mathbf{x}_\alpha} \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \boldsymbol{\xi}_\alpha \, V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= -3 \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$+ \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, (1 - \tau) \, \frac{d}{d\tau} V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$- \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \boldsymbol{\xi}_\alpha \, V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \frac{\partial}{\partial \mathbf{y}_\alpha} \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, (1 - \tau) \, \frac{d}{d\tau} V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$+ \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \tau \, \frac{d}{d\tau} V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= \left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \frac{d}{d\tau} V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= - \int_N d^3\mathbf{y} \, \big(V(\mathbf{y}) + V(\mathbf{x}) \big) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= -2\mathfrak{s} \, V(\mathbf{x}) - \int_N d^3\mathbf{y} \, \frac{1}{2} \, \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\beta \partial_{\alpha\beta} V(\mathbf{x}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= -2\mathfrak{s} \, V(\mathbf{x}) - \frac{1}{6} \, \delta^2 \, \mathfrak{s}_2 \, \Delta_{\mathbb{R}^3} V(\mathbf{x}) + \mathcal{O}\left(\frac{M\delta^3}{|\mathbf{x}|^3} \right),$$

where in the last steps we again performed an in powers of ξ and used (5.15).

Applying the Gauss divergence theorem and spherical symmetry, we conclude that

$$\left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \boldsymbol{\xi}_\alpha \, V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= -\mathfrak{s} \, \mathbf{x}_\alpha \, V(\mathbf{x}) - \frac{1}{6} \, \delta^2 \, \mathfrak{s}_2 \, \partial_\alpha V(\mathbf{x}) + \mathcal{O}\left(\frac{M\delta^3}{|\mathbf{x}|^3} \right). \tag{A.48}$$

Note that we used that

$$\partial_\alpha \big(\mathbf{x}_\alpha \, V(\mathbf{x}) \big) = 3 \, V(\mathbf{x}) + \mathbf{x}^\alpha \partial_\alpha V(\mathbf{x}) = 2 \, V(\mathbf{x})$$

(where in the last step we used again that the potential is homogeneous of degree minus one). Multiplying by $\hat{\mathbf{x}}_\alpha$ gives

$$\left(\int_1^\infty - \int_{-\infty}^0 \right) d\tau \int_N d^3\mathbf{y} \, \frac{\mathbf{x}\boldsymbol{\xi}}{|\mathbf{x}|} \, V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}} \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= -\mathfrak{s}\,|\mathbf{x}|\,V(\mathbf{x}) - \frac{1}{6}\,\delta^2\,\mathfrak{s}2\,\frac{\mathbf{x}_\alpha}{|\mathbf{x}|}\,\partial_\alpha V(\mathbf{x}) + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)\,.$$

We finally use again that V is homogeneous of degree minus one.

We remark that the reader who feels uneasy with the singularity of V at the origin can again smoothen V near the origin. For a spherically symmetric smooth function $W(|\mathbf{x}|)$, the divergence equation

$$\partial_\alpha u_\alpha(\mathbf{x}) = W(|\mathbf{x}|)$$

has the explicit solution

$$u_\alpha(\mathbf{x}) = \frac{\mathbf{x}_\alpha}{2\,|\mathbf{x}|^3} \int_0^r r^2\,W(r)\,dr\,.$$

Evaluating the integral for the mollified Newtonian potential, one sees that the mollification can be removed, giving $u_\alpha(\mathbf{x}) = -\mathbf{x}_\alpha V(\mathbf{x})$ for $|\mathbf{x}| = R$. This gives a clean justification of the first summand in (A.48). □

Lemma A.12

$$\frac{3}{4}\left(\int_1^\infty - \int_{-\infty}^0\right) d\tau \int_N d^3\mathbf{y}\,\Big(\frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3}\Big)\,V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,\mathcal{L}(\mathbf{x},\mathbf{y})$$

$$= -\frac{1}{2}\,\delta^2\,\mathfrak{s}2\,\frac{1}{|\mathbf{x}|}\,V(\mathbf{x}) + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)\,.$$

Proof Applying the transformations (10.20) and (10.21), one sees that

$$\int_{-\infty}^\infty \epsilon(\tau)\,d\tau \int_N d^3\mathbf{y}\,\Big(\frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3}\Big)\,V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,\mathcal{L}(\mathbf{x},\mathbf{y}) = 0\,.$$

Hence

$$\left(\int_1^\infty - \int_{-\infty}^0\right) d\tau \int_N d^3\mathbf{y}\,\Big(\frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3}\Big)\,V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,\mathcal{L}(\mathbf{x},\mathbf{y})$$

$$= -\int_0^1 d\tau \int_N d^3\mathbf{y}\,\Big(\frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3}\Big)\,V|_{\tau\mathbf{y}+(1-\tau)\mathbf{x}}\,\mathcal{L}(\mathbf{x},\mathbf{y})$$

$$= -\int_N d^3\mathbf{y}\,\Big(\frac{\boldsymbol{\xi}^2}{|\mathbf{x}|} - \frac{(\mathbf{x}\boldsymbol{\xi})^2}{|\mathbf{x}|^3}\Big)\,V(\mathbf{x})\,\mathcal{L}(\mathbf{x},\mathbf{y}) + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)\,.$$

Using again (5.15) gives the result. □

Collecting all the terms, we obtain

$$\mathfrak{P}(R) = -\mathfrak{s}\,|\mathbf{x}|\,V(\mathbf{x}) + \frac{1}{6}\,\delta^2\,\mathfrak{s}2\,\frac{1}{|\mathbf{x}|}\,V(\mathbf{x}) - \frac{1}{2}\,\delta^2\,\mathfrak{s}2\,\frac{1}{|\mathbf{x}|}\,V(\mathbf{x})$$

$$+ \mathfrak{s}\,|\mathbf{x}|\,V(\mathbf{x}) - \frac{1}{6\,|\mathbf{x}|}\,\delta^2\,\mathfrak{s}2\,V(\mathbf{x}) + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)$$

$$= -\frac{1}{2}\,\delta^2\,\mathfrak{s}2\,\frac{1}{|\mathbf{x}|}\,V(\mathbf{x}) + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big) = \frac{M}{|\mathbf{x}|^2}\,\delta^2\,\mathfrak{s}2 + \mathcal{O}\Big(\frac{M\delta^3}{|\mathbf{x}|^3}\Big)\,.$$

Using this result in (A.32) concludes the proof of Proposition A.9.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study. The authors have read and accept the Publisher's Data Use Privacy Policy and the Aries Privacy Policy.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no conflict of interest or conflict of interest to declare that are relevant to the content of this article. The authors comply to the highest scientific and ethical standards.

References

1. Besse, A.L.: Einstein Manifolds, Classics in Mathematics. Springer, Berlin (2008). (**Reprint of the (1987) edition**)
2. Bogachev, V.I.: Measure Theory, vol. I. Springer, Berlin (2007)
3. Bray, H.L.: Proof of the Riemannian Penrose inequality using the positive mass theorem. J. Differ. Geom. **59**(2), 177–267 (2001)
4. Brown, J.D., York, J.W., Jr.: Quasilocal energy and conserved charges derived from the gravitational action. Phys. Rev. D (3) **47**(4), 1407–1419 (1993). arXiv:gr-qc/9209012
5. Curiel, E., Finster, F., Isidro, J.M.: Two-dimensional area and matter flux in the theory of causal fermion systems. Int. J. Mod. Phys. D **29**, 2050098 (2020). arXiv:1910.06161 [math-ph]
6. Finster, F.: Gaussian-type coordinates in the setting of causal variational principles (**in preparation**)
7. Finster, F.: Light-cone expansion of the Dirac sea to first order in the external potential. Mich. Math. J. **46**(2), 377–408 (1999). arXiv:hep-th/9707128
8. Finster, F.: Light-cone expansion of the Dirac sea in the presence of chiral and scalar potentials. J. Math. Phys. **41**(10), 6689–6746 (2000). arXiv:hep-th/9809019
9. Finster, F.: The Principle of the Fermionic Projector, hep-th/0001048, hep-th/0202059, hep-th/0210121, AMS/IP Studies in Advanced Mathematics, vol. 35. American Mathematical Society, Providence, RI (2006)
10. Finster, F.: Causal variational principles on measure spaces. J. Reine Angew. Math. **646**, 141–194 (2010). arXiv:0811.2666 [math-ph]
11. Finster, F.: The Continuum Limit of Causal Fermion Systems, Fundamental Theories of Physics, vol. 186. Springer, Cham (2016). arXiv:1605.04742 [math-ph]
12. Finster, F.: Causal fermion systems: a primer for Lorentzian geometers. J. Phys. Conf. Ser. **968**, 012004 (2018). arXiv:1709.04781 [math-ph]
13. Finster, F.: The causal action in Minkowski space and surface layer integrals. SIGMA Symmetry Integr. Geom. Methods Appl. **16**(091) (2020). arXiv:1711.07058 [math-ph]
14. Finster, F.: Perturbation theory for critical points of causal variational principles. Adv. Theor. Math. Phys. **24**(3), 563–619 (2020). arXiv:1703.05059 [math-ph]
15. Finster, F.: Causal fermion systems: classical gravity and beyond. In: Ruffini, R., Vereshchagin, G. (eds.) The Sixteenth Marcel Grossmann Meeting, pp. 661–678. World Scientific, Singapore (2023) . arXiv:2109.05906 [gr-qc]

16. Finster, F.: Solving the linearized field equations of the causal action principle in Minkowski space. Adv. Theor. Math. Phys. **27**(7) 2087–2217 (2023). arXiv:2304.00965
17. Finster, F., Grotz, A.: A Lorentzian quantum geometry. Adv. Theor. Math. Phys. **16**(4), 1197–1290 (2012). arXiv:1107.2026 [math-ph]
18. Finster, F., Kamran, N.: Spinors on singular spaces and the topology of causal fermion systems. Mem. Am. Math. Soc. **259**(1251), v+83 (2019)
19. Finster, F., Kamran, N.: Complex structures on jet spaces and bosonic Fock space dynamics for causal variational principles. Pure Appl. Math. Q. **17**(1), 55–140 (2021). arXiv:1808.03177 [math-ph]
20. Finster, F., Kamran, N.: Fermionic Fock spaces and quantum states for causal fermion systems. Ann. Henri Poincaré **23**(4), 1359–1398 (2022). arXiv:2101.10793 [math-ph]
21. Finster, F., Kindermann, S.: A gauge fixing procedure for causal fermion systems. J. Math. Phys. **61**(8), 082301 (2020). arXiv:1908.08445 [math-ph]
22. Finster, F., Kindermann, S., Treude, J.-H.: An introductory course on causal fermion systems :An Introduction to Fundamental Structures, Methods and Applications (2023) https://arxiv.org/abs/2411.06450
23. Finster, F., Kleiner, J.: Causal fermion systems as a candidate for a unified physical theory. J. Phys. Conf. Ser. **626**, 012020 (2015). arXiv:1502.03587 [math-ph]
24. Finster, F., Kleiner, J.: Noether-like theorems for causal variational principles. Calc. Var. Partial Differ. Equ. **55:35**(2), 41 (2016). arXiv:1506.09076 [math-ph]
25. Finster, F., Kleiner, J.: A Hamiltonian formulation of causal variational principles. Calc. Var. Partial Differ. Equ. **56:73**(3), 33 (2017). arXiv:1612.07192 [math-ph]
26. Finster, F., Kraus, M.: The regularized Hadamard expansion. J. Math. Anal. Appl. **491**(2), 124340 (2020). arXiv:1708.04447 [math-ph]
27. Finster, F., Langer, C.: Causal variational principles in the σ-locally compact setting: existence of minimizers. Adv. Calc. Var. **15**(3), 551–575 (2022). arXiv:2002.04412 [math-ph]
28. Finster, F., Lottner, M.: Banach manifold structure and infinite-dimensional analysis for causal fermion systems. Ann. Glob. Anal. Geom. **60**(2), 313–354 (2021). arXiv:2101.11908 [math-ph]
29. Finster, F., Lottner, M.: Elliptic methods for solving the linearized field equations of causal variational principles. Calc. Var. Partial Differ. Equ. **61**(4), 133 (2022). arXiv:2111.08261 [math-ph]
30. Finster, F., Platzer, A.: A positive mass theorem for static causal fermion systems. Adv. Theor. Math. Phys. **25**(7), 1735–1818 (2021). arXiv:1912.12995 [math-ph]
31. Finster, F., Röken, C.: Self-adjointness of the Dirac Hamiltonian for a class of non-uniformly elliptic boundary value problems. Ann. Math. Sci. Appl. **1**(2), 301–320 (2016). arXiv:1512.00761 [math-ph]
32. Finster, F., Schiefeneder, D.: On the support of minimizers of causal variational principles. Arch. Ration. Mech. Anal. **210**(2), 321–364 (2013). arXiv:1012.1589 [math-ph]
33. Friedlander, F.G.: The Wave Equation on a Curved Space-Time. Cambridge University Press, Cambridge (1975)
34. Huisken, G., Ilmanen, T.: The inverse mean curvature flow and the Riemannian Penrose inequality. J. Differ. Geom. **59**(3), 353–437 (2001)
35. Landau, L.D., Lifshitz, E.M.: The Classical Theory of Fields, Revised Second Edition. Course of Theoretical Physics, Vol. 2. Translated from the Russian by M. Hamermesh. Pergamon Press, Oxford (1962)
36. Link to web platform on causal fermion systems. www.causal-fermion-system.com
37. Liu, C.-C.M., Yau, S.-T.: Positivity of quasilocal mass. Phys. Rev. Lett. **90**(23), 4 (2003). arXiv:gr-qc/0303019
38. Parker, T., Taubes, C.H.: On Witten's proof of the positive energy theorem. Commun. Math. Phys. **84**(2), 223–238 (1982)
39. Shi, Y., Tam, L.-F.: Positive mass theorem and the boundary behaviors of compact manifolds with non-negative scalar curvature. J. Differ. Geom. **62**(1), 79–125 (2002). arXiv:math/0301047 [math.DG]
40. Taylor, M.E.: Partial Differential Equations. II, second ed., Applied Mathematical Sciences, vol. 116. Springer, New York (2011)
41. Wald, R.M.: General Relativity. University of Chicago Press, Chicago, IL (1984)
42. Witten, E.: A new proof of the positive energy theorem. Commun. Math. Phys. **80**(3), 381–402 (1981)