A positive mass theorem for static causal fermion systems

FELIX FINSTER AND ANDREAS PLATZER

Asymptotically flat static causal fermion systems are introduced. Their total mass is defined as a limit of surface layer integrals which compare the measures describing the asymptotically flat spacetime and a vacuum spacetime near spatial infinity. Our definition does not involve any regularity assumptions; it even applies to singular or generalized "quantum" spacetimes. A positive mass theorem is proven. Our methods and results explain why and how the causal action principle incorporates the nonlinear effects of gravity for static systems.

1. Introduction and outline of results

In Newtonian physics, the total energy is obtained simply by adding the energies of all particles and fields of the system, including the energy of the gravitational field. In general relativity, however, the situation is much more difficult due to the nonlinearity of the gravitational interaction. Nevertheless, as shown by Arnowitt, Deser and Misner [1], it is possible to define total energy and momentum of an isolated gravitational system, expressed in terms of the asymptotics of the metric tensor near infinity. Since in the present paper we restrict attention throughout to the *static* situation, we also recall the definition by Arnowitt, Deser and Misner (ADM) only in this setting. Let $\mathcal{M}$ be a Lorentzian manifold which is the topological product $\mathcal{M} = \mathbb{R} \times \mathcal{N}$. We denote the spacetime points by $x = (t, \mathbf{x})$ with $t \in \mathbb{R}$ and $\mathbf{x} \in \mathcal{N}$. Next, we assume that the Lorentzian metric is time independent (in other words, we assume that ∂_t is a Killing field), that the induced metric g on $\mathcal{N}$ is Riemannian, and that the second fundamental form vanishes on $\mathcal{N}$ (this is sometimes referred to as the time-symmetric case). Finally, we assume that $(\mathcal{N}, g)$ is *asymptotically flat.* Stated for simplicity in three spatial dimensions and with one asymptotic end, this means that there is a compact set $K \subset \mathcal{N}$ such that $\mathcal{N} \setminus K$ is diffeomorphic to the region $\mathbb{R}^3 \setminus \overline{B_R(0)}$ outside a closed ball of radius R. In the chart defined by this diffeomorphism, the metric should be of the form

$$(1.1) \qquad g_{\alpha\beta}(\mathbf{x}) = \delta_{\alpha\beta} + a_{\alpha\beta}(\mathbf{x}) , \qquad \mathbf{x} \in \mathbb{R}^3 \setminus \overline{B_R(0)} ,$$

where $a_{\alpha\beta}$ decays at infinity as

$$a_{\alpha\beta} = \mathcal{O}(1/|\mathbf{x}|) , \quad \partial_\gamma a_{\alpha\beta} = \mathcal{O}(1/|\mathbf{x}|^2) \quad \text{and} \quad \partial_{\gamma\delta} a_{\alpha\beta} = \mathcal{O}(1/|\mathbf{x}|^3) .$$

Under these assumptions, the total energy is also referred to as the *total mass* or *ADM mass.* It is defined by

$$(1.2) \qquad \mathfrak{M}_{\text{ADM}} = \frac{1}{16\pi} \lim_{R \to \infty} \sum_{\alpha,\beta=1}^{3} \int_{S_R} (\partial_\beta g_{\alpha\beta} - \partial_\alpha g_{\beta\beta}) \, \nu^\alpha \, d\Omega ,$$

where $d\Omega$ is the area form on the coordinate sphere S_R, and ν is the normal vector to S_R (both defined in the coordinate chart).

In Newtonian physics, the fact that the energies of all particles and all energy densities of fields are positive implies that the total energy is also positive. Again, in general relativity the connection is much more involved. It

is made precise by the *positive mass theorem* proved by Schoen and Yau [44]. It states that if a suitable *local energy condition* is fulfilled, which in the static case reduces to the condition that the scalar curvature of g is non-negative,

$$\text{scal} \geq 0 \, ,$$

then the total mass non-negative. Moreover, if the total mass vanishes, then $(\mathcal{N}, g)$ is flat. The positive mass theorem makes a profound statement on the nature of the gravitational interaction. We remark that it can be generalized to higher dimensions; see for example the recent papers [38, 45] and the references therein.

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 nontrivial, 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*. In [18, Chapter 4] it is shown that in a specific limiting case, the so-called *continuum limit*, the Euler-Lagrange (EL) equations of the causal action principle give rise to the Einstein equations, up to possible higher order corrections in curvature (which scale in powers of (δ^2 Riem), where δ is the Planck length and Riem is the curvature tensor). In this limiting case, spacetime goes over to a Lorentzian manifold, whereas the gravitational coupling constant $G \sim \delta^2$ is determined by the length scale δ of the microscopic spacetime structure.

The derivation of the Einstein equations in [18, Chapter 4] has two disadvantages. First, it is rather technical, because it relies on the detailed form of the regularized light-cone expansion of the kernel of the fermionic projector. Consequently, the derivation does not give a good intuitive understanding of the underlying mechanisms. Second and more importantly, the Einstein equations are recovered only in the continuum limit, but the methods do not give any insight into the geometric meaning of the EL equations for more general "quantum" spacetimes.

In view of theses disadvantages, it is an important task to study the nature of the gravitational interaction as described by the causal action principle without referring to limiting cases, but instead by analyzing directly the corresponding EL equations. One step in this direction is the recent paper [8], where the connection between area change and matter flux is worked out for two-dimensional surfaces propagating in a null Killing direction. In the present paper we go a step in a different, somewhat complementary direction which aims at understanding gravity for static systems directly

from the causal action principle. We succeed in giving a general definition of the total mass for static causal fermion systems. Moreover, a positive mass theorem is proved which states that if a suitable local energy condition is fulfilled, then the total mass is positive. We also explain how the ADM mass is recovered as a limiting case.

More precisely, our results are stated as follows. In the theory of causal fermion systems, a physical system (consisting of spacetime and all structures therein) is described by a Borel measure ρ on a set of linear operator $\mathcal{F} \subset \mathrm{L}(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ (for details see the preliminaries in Section 2.2). Spacetime M is defined as the support of this measure,

$$M := \mathrm{supp}\,\rho \,.$$

In the *static* situation to be considered here, there is a global time coordinate $t \in \mathbb{R}$, i.e.

$$M = \mathbb{R} \times N \ni x = (t, \mathbf{x}) \,.$$

Moreover, the system should be time independent. This is made precise by a one-parameter group $(\mathcal{U}_t)_{t \in \mathbb{R}}$ of unitary transformations which leaves the measure ρ invariant (for details see Definition 3.1). In particular, the measure ρ can be written as

$$(1.3) \qquad\qquad\qquad d\rho = dt\,d\mu$$

where μ is a Borel measure on N. The dynamics of a causal fermion system is described by a variational principle, the *causal action principle* (see Section 2.2). In the *static* case, it reduces to minimizing the action $\mathcal{S}$ given by

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

under variations of the measure μ, leaving the total volume fixed. Here in $\mathcal{G} := \mathcal{F}/\mathbb{R}$ we divided out the action of the group $(\mathcal{U}_t)_{t \in \mathbb{R}}$, and the Lagrangian $\mathcal{L}_\kappa$ is of the form

$$\mathcal{L}_\kappa = \mathcal{L} + \kappa\,\mathcal{T} \qquad \text{with given functions} \qquad \mathcal{L}, \mathcal{T} \,:\, \mathcal{G} \times \mathcal{G} \to \mathbb{R}_0^+ \,,$$

and $\kappa > 0$ is a Lagrange parameter (for details see Sections 2.2 and 3.2). The functions $\mathcal{L}$ and $\mathcal{T}$ are symmetric, i.e.

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathcal{L}(\mathbf{y}, \mathbf{x}) \quad \text{and} \quad \mathcal{T}(\mathbf{x}, \mathbf{y}) = \mathcal{T}(\mathbf{y}, \mathbf{x}) \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{G} \,.$$

The causal action principle in the static case is a specific example of a *causal variational principle* (for the general context see Section 2.1).

A minimizer μ of the above variational principle satisfies the

$$(1.4) \qquad \textit{Euler-Lagrange (EL) equations} \qquad \ell_\kappa|_N \equiv \inf_{\mathcal{G}} \ell_\kappa = 0 \,,$$

where the function ℓ_κ is defined by

$$(1.5) \qquad \ell_\kappa : \mathcal{G} \to \mathbb{R}_0^+ \,, \qquad \ell_\kappa(\mathbf{x}) = \ell(\mathbf{x}) + \kappa\, \mathfrak{t}(\mathbf{x})$$

with

$$(1.6) \qquad \ell(\mathbf{x}) := \int_N \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) - \mathfrak{s}$$

$$(1.7) \qquad \mathfrak{t}(\mathbf{x}) := \int_N \mathcal{T}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) \,,$$

and $\mathfrak{s}$ is a positive parameter which for convenience is chosen such that the infimum in (1.4) is zero.

Using the above notions, the total mass can be introduced as follows. Let μ and $\tilde{\mu}$ be two measures which are jointly static (meaning that they are both static with respect to the same one-parameter group $(\mathcal{U}_t)_{t\in\mathbb{R}}$; for details see Definition 3.1) and are both minimizers or critical points of the static causal action principle for the same values of the parameters κ and $\mathfrak{s}$ (for details on the definition of critical measures see the paragraph after (2.9) in Section 2.1). We define the functions $\tilde{\ell}$ and $\tilde{\mathfrak{t}}$ by adding tildes to μ, ℓ and $\mathfrak{t}$ in (1.6) and (1.7). In order to compare the measures μ and $\tilde{\mu}$, we relate them by the Lagrangian. To this end, we introduce the functions

$$(1.8) \qquad \begin{cases} n : N \to \mathbb{R}_0^+ \cup \{\infty\} \,, & n(\mathbf{x}) = \displaystyle\int_{\tilde{N}} \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) \, d\tilde{\mu}(\mathbf{y}) \\[3mm] \tilde{n} : \tilde{N} \to \mathbb{R}_0^+ \cup \{\infty\} \,, & \tilde{n}(\mathbf{x}) = \displaystyle\int_N \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) \end{cases}$$

and define the *correlation measures* ν and $\tilde{\nu}$ by

$$(1.9) \qquad d\nu(\mathbf{x}) = n(\mathbf{x}) \, d\mu(\mathbf{x}) \qquad \text{and} \qquad d\tilde{\nu}(\mathbf{x}) = \tilde{n}(\mathbf{x}) \, d\tilde{\mu}(\mathbf{x}) \,.$$

Definition 1.1. *The measures $\tilde{\mu}$ and μ are* **asymptotically close** *if they are both σ-finite with infinite total volume,*

$$(1.10) \qquad\qquad \tilde{\mu}(\tilde{N}) = \mu(N) = \infty \,,$$

but

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

We begin with the most general definition of the total mass.

Definition 1.2. *Assume that μ and $\tilde{\mu}$ are asymptotically close. Then the* **total mass** *$\mathfrak{M}$ of $\tilde{\mu}$ relative to μ is defined by*

$$\mathfrak{M}(\tilde{\mu}, \mu) := \lim_{\Omega \nearrow N} \lim_{\tilde{\Omega} \nearrow \tilde{N}} \left(- \mathfrak{s}\Big(\tilde{\mu}(\tilde{\Omega}) - \mu(\Omega)\Big) \right.$$

$$(1.11) \qquad \left. + \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) - \int_{\Omega} d\mu(\mathbf{x}) \int_{\tilde{N} \setminus \tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) \right) ,$$

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

Here the limits exist and are independent of the choice of the exhaustions (see Proposition 4.1).

This definition is extremely general because it does not involve any smoothness or continuity assumptions. It even applies to singular or discrete measures. It is not necessary (and would not even be possible in this generality) to specify the dimension of spacetime. Clearly, in order to compare this notion of total mass with the ADM mass, we need to specialize the setting. We now explain step by step how this can be done. Along the way, we will also explain the structure of the formula (1.11).

The first step is to assume that μ or $\tilde{\mu}$ is a continuous measure in the sense that it is non-atomic (for details see Definition 4.2). In this case, in (1.11) one may restrict attention to sets Ω and $\tilde{\Omega}$ with the same measure, making it possible to write the total mass as

$$(1.12) \quad \mathfrak{M}(\tilde{\mu}, \mu) = \lim_{\Omega_n \nearrow N, \, \tilde{\Omega}_n \nearrow \tilde{N} \text{ with } \mu(\Omega_n) = \tilde{\mu}(\tilde{\Omega}_n) < \infty}$$

$$\times \left(\int_{\tilde{\Omega}_n} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega_n} d\mu(\mathbf{y}) \, \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) - \int_{\Omega_n} d\mu(\mathbf{x}) \int_{\tilde{N} \setminus \tilde{\Omega}_n} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) \right) ,$$
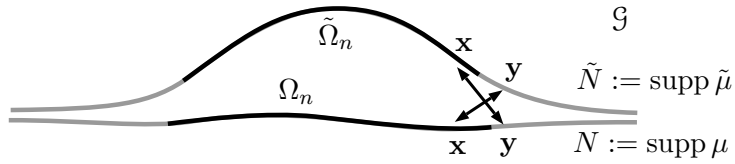
Figure 1: The surface layer integral defining the total mass.

where $(\Omega_n)_{n\in\mathbb{N}}$ and $(\tilde{\Omega}_n)_{n\in\mathbb{N}}$ are exhaustions of N and $\tilde{N}$, respectively (for details see Proposition 4.3).

Let us briefly explain the structure of the above formulas for the total mass. The double integrals in (1.11) and (1.12) are so-called *surface layer integrals*, which generalize surface integrals to the setting of causal variational principles (see [25, Section 2.3], [28, Section 7.1] or Sections 2.1.3 and 2.1.5 in the preliminaries). The main point is that of the two arguments $\mathbf{x}$ and $\mathbf{y}$ of the Lagrangian $\mathcal{L}_\kappa$, one lies in the interior region (Ω_n or $\tilde{\Omega}_n$), whereas the other lies in the exterior region of the other spacetime ($\tilde{N} \setminus \tilde{\Omega}_n$ and $N \setminus \Omega_n$, respectively), as is indicated by the arrows in Figure 1. Since $\mathcal{L}_\kappa$ typically decays if its arguments are far apart, the main contribution to the integrals is obtained when both $\mathbf{x}$ and $\mathbf{y}$ are close to the boundaries $\partial\Omega_n \subset N$ or $\partial\tilde{\Omega}_n \subset \tilde{N}$. Therefore, each of the two double integrals (1.11) and (1.12) can be thought of as an integral over a "thin strip" around the boundaries of $\partial\Omega_n \subset N$ or $\partial\tilde{\Omega}_n \subset \tilde{N}$. Moreover, it is important that in (1.11) and (1.12) we take the difference of these double integrals. This is needed for getting a connection to the general conservation law as first derived in [22, Section 4 and Appendix A].

Using this intuitive picture, it is evident that the expressions (1.11) and (1.12) depend only on the geometry (as encoded in the measures μ and $\tilde{\mu}$) near infinity, but not on the geometry in any compact subset. This statement will be made mathematically precise in Theorem 4.6 in Section 4.2. However, it is important to observe that the *inner volume* does come into play, as one sees from the term $\tilde{\mu}(\tilde{\Omega}) - \mu(\Omega)$ in (1.11) and the constraint $\tilde{\mu}(\tilde{\Omega}_n) = \mu(\Omega_n)$ in (1.12). Therefore, our definition of the total mass can be understood as a limit of surface layer integrals to be evaluated on the boundaries of large subsets $\Omega \subset N$ and $\tilde{\Omega} \subset \tilde{N}$ which have the same volume.

At first sight, the fact that the inner volume comes into play seems to be a major difference to the ADM mass (1.2), which is computable purely from the geometric data near infinity. However, as will be explained later in

this introduction (after the statement of Theorem 1.10), for causal fermion systems constructed in a Lorentzian spacetime (so-called static Dirac system; see Section 2.3), the inner volume drops out of the computation, giving formulas for $\mathfrak{M}$ purely in terms of the geometry near infinity. This can be understood non-technically as follows: The main contribution to the surface layer integral in (1.11) is $\mathfrak{s}$ times the difference of the volumes of $\tilde{\Omega}$ and Ω. If this contribution vanishes, the next-to-leading contribution comes into play. This next-to-leading order contribution gives the total mass. For static Dirac systems, it can be computed purely from the geometric data at infinity, without referring to the volumes of $\tilde{\Omega}$ or Ω. This argument will be made precise in Proposition 6.3 and Appendix A.

In order to avoid confusion, we point out that the "volume" considered here refers to the spatial volume measure μ obtained by rewriting the space-time volume measure ρ in the form (1.3) (and similarly for $\tilde{\mu}$). In examples of spacetimes described by static Lorentzian manifolds (like the Schwarzschild geometry), this measure does *not* agree with the volume element of the induced Riemannian metric on N, but it coincides instead with the volume form of spacetime contracted with the Killing field. This will be explained in more detail in Section 6.

We next outline our positivity results for the mass. In order to prove our results, we need to impose stronger assumptions on the asymptotics at infinity, We first state these assumptions and explain the afterward.

Definition 1.3. *The measure μ has* **one asymptotic end of dimension** *$k \in \mathbb{N}$ if there is a relatively compact open set $I \subset N$ and a diffeomorphism*

$$(1.13) \qquad\qquad \Phi \, : \, N \setminus I \to \mathbb{R}^k \setminus B_R$$

(where B_R is the open ball of radius $R > 0$) with the property that the push-forward of μ is of the form

$$d(\Phi_* \mu)(\mathbf{x}) = h(\mathbf{x}) \, d^k \mathbf{x}$$

with a smooth function $h \in C^\infty(\mathbb{R}^k \setminus B_R, \mathbb{R}^+)$ with $\lim_{\mathbf{x} \to \infty} h(\mathbf{x}) = 1$.
A sequence $(\mathbf{x}_n)_{n \in \mathbb{N}}$ in N **tends to infinity**, *$\mathbf{x}_n \to \infty$, if almost all elements of the sequence are in $N \setminus I$, and if the image sequence $\Phi(\mathbf{x}_n)$ tends to infinity in $\mathbb{R}^k$. A function g on N converges at infinity if the sequence $f(\mathbf{x}_n)$ converges for every $\mathbf{x}_n$ which tends to infinity. The limit is denoted by $\lim_{N \ni \mathbf{x} \to \infty} g(x)$.*

For simplicity, 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.

Definition 1.4. *The measure μ is a **vacuum measure** of dimension $k \in \mathbb{N}$ if it satisfies the following conditions:*

(i) *It has one asymptotic end of dimension k (see Definition 1.3).*

(ii) *The function ℓ is constant on N,*

(1.14) $$\ell(\mathbf{x}) =: \ell_\infty \qquad \text{for all } \mathbf{x} \in N .$$

Definition 1.5. *The measure $\tilde{\mu}$ is **asymptotically flat** of dimension $k \in \mathbb{N}$ if it satisfies the following conditions:*

(i) *It is asymptotically close to a vacuum measure μ of dimension $k \in \mathbb{N}$ (see Definitions 1.1 and 1.4).*

(ii) *It has one asymptotic end of dimension k (see Definition 1.3). The following limit exists,*

$$\lim_{\tilde{N} \ni \mathbf{x} \to \infty} \tilde{\ell}(\mathbf{x}) =: \tilde{\ell}_\infty \in \mathbb{R} .$$

The function $\tilde{\ell} - \tilde{\ell}_\infty$ is integrable, i.e.

$$\int_{\tilde{N}} \left| \tilde{\ell}(\mathbf{x}) - \tilde{\ell}_\infty \right| d\tilde{\mu}(\mathbf{x}) < \infty .$$

(iii) *There is a mapping $F : N \to \tilde{N}$ such that*

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

(where $F_*\mu$ is the push-forward measure defined by $(F_*\mu)(\tilde{\Omega}) = \mu(F^{-1}(\tilde{\Omega}))$). When restricted to $N \setminus I$ (with I as in Definition 1.4), this mapping is a diffeomorphism to its image. Moreover, it tends to the identity at infinity in the sense that the surface layer integral in (1.12) can be linearized to obtain

(1.15) $$\mathfrak{M}(\tilde{\mu}, \mu) = \lim_{\Omega \nearrow N} \int_\Omega d\mu(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \left(\mathcal{L}_\kappa\big(F(\mathbf{x}), \mathbf{y}\big) - \mathcal{L}_\kappa\big(\mathbf{x}, F(\mathbf{y})\big) \right)$$

(1.16) $$= \lim_{\Omega \nearrow N} \int_\Omega d\mu(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \left(D_{1,w} - D_{2,w} \right) \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) \right)$$

for a suitable vector field $w \in \Gamma(N, T\mathcal{G})$ on $\mathcal{G}$ along N.

We remark for clarity that (1.15) follows immediately from (1.12) using the definition of the push-forward measure (together with the fact that the limits in (1.12) exist). The directional derivatives in (1.16) act on the first and second argument of the Lagrangian, respectively. We also note that, setting $\tilde{I} = F(I)$, the mapping $\tilde{\Phi} := \Phi \circ F^{-1}|_{\tilde{N} \setminus \tilde{I}}$ is a diffeomorphism from $\tilde{N} \setminus \tilde{I}$ to $\mathbb{R}^k \setminus B_R$ with the property that it changes the measure $\tilde{\mu}$ according to

$$d(\tilde{\Phi}_* \tilde{\mu})(\mathbf{x}) = \tilde{h}(\mathbf{x})\, d^k \mathbf{x} \qquad \text{with} \qquad \lim_{\mathbf{x} \to \infty} \tilde{h}(\mathbf{x}) = 1 \,.$$

The assumptions so far can be regarded as the analogs of the decay conditions for the metric in (1.1) in the setting of asymptotically flat Riemannian manifolds. We next introduce another important assumption, which means in words that both measures μ and $\tilde{\mu}$ should be extendable to families of critical measures $(\mu_\tau)_\tau$ and $(\tilde{\mu}_\tau)_\tau$ for a variable value of the parameter $\kappa = \kappa(\tau)$.

Definition 1.6. *The measure μ is κ-**extendable** if the following conditions hold:*

(i) *There is a family of measures $(\mu_\tau)_{\tau \in (-1,1)}$ of the form*

$$\mu_\tau = (F_\tau)_* \mu \,,$$

each of which satisfies the EL equations (1.4) with a parameter $\kappa(\tau)$ and

$$F_0 = \mathrm{id}_N \qquad \text{and} \qquad \kappa'(0) \neq 0 \,.$$

(ii) *For every $x \in N$, the curve $F_\tau(x)$ is differentiable at $\tau = 0$, giving rise to a vector field*

$$(1.17) \qquad v := \frac{d}{d\tau} F_\tau \Big|_{\tau=0} \in \Gamma(N, T\mathcal{G}) \,.$$

For a convenient normalization, we always choose the parametrization such that

$$(1.18) \qquad \frac{d}{d\tau} \log \kappa(\tau) \Big|_{\tau=0} = -1 \,.$$

We remark that all our results hold as well for families of critical measures if one merely replaces the EL equations (1.4) in (i) by the weak EL equations (see (2.9) in Section 2.1 with ℓ carrying an additional subscript κ).

Definition 1.7. *Let $\tilde{\mu}$ be asymptotically flat with respect to the vacuum measure μ. Then $\tilde{\mu}$ is said to be κ-**scalable** if both $\tilde{\mu}$ and μ are κ-extendable and if, for suitable choices of the mappings F_τ and $\tilde{F}_\tau$ in Definition 1.6, the vector field w in (1.16) is related to the vector fields v and $\tilde{v}$ in (1.17) by*

$$(1.19) \qquad\qquad w = g\left(\tilde{v} - v\right)$$

*with a constant $g \in \mathbb{R}$, referred to as the **gravitational coupling constant**.*

We can now state our main result.

Definition 1.8. *The asymptotically flat measure $\tilde{\mu}$ satisfies the **local energy condition** if*

$$(1.20) \qquad\qquad \tilde{\ell}(\mathbf{x}) \geq \tilde{\ell}_\infty \qquad \text{for all } \mathbf{x} \in \tilde{N} .$$

Theorem 1.9. (Positive mass theorem) *Assume that $\tilde{\mu}$ is asymptotically flat (see Definition 1.5) and κ-scalable (see Definition 1.7). Then the total mass can be written as*

$$(1.21) \qquad\qquad \mathfrak{M}(\tilde{\mu}, \mu) = g \int_{\tilde{N}} \left(\tilde{\ell} - \tilde{\ell}_\infty\right) d\tilde{\mu} .$$

If $\tilde{\mu}$ satisfies the local energy condition and the gravitational coupling constant g is positive, then the total mass is non-negative,

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

Moreover, if the total mass vanishes, then $\tilde{\mu}$ is a vacuum measure (see Definition 1.4).

For the proof of this theorem, we consider a linear integral equation, referred to as the *equations of linearized gravity* (see (5.3) in Section 5.1). These equations are obtained by linearizing the EL equations of the causal action principle in the parameter κ. Working with linear equations is inspired by the spinor proof of the positive energy theorem (see [42, 46]). The reason why these spinorial methods extend to the setting of causal fermion systems can be understood from the fact that in interesting examples (see the static Dirac systems described below), the causal fermion system is built up of solutions of the Dirac equation. The integral formula (1.21) for the total

mass can be regarded as the analog of the formula

$$\mathfrak{M}_{\mathrm{ADM}} = c(k) \int_{\mathcal{N}} \left(|\nabla \psi|^2 + \frac{\mathrm{scal}}{4} |\psi|^2 \right) d\mu_{\mathcal{N}} \, ,$$

where ψ is the Witten spinor.

Clearly, the above theorem leaves the following questions open:

(a) How is $\mathfrak{M}(\tilde{\mu}, \mu)$ related to the ADM mass in (1.2)?

(b) Why is the measure $\tilde{\mu}$ κ-scalable (see Definition 1.7)?

(c) Why is the gravitational constant as defined in (1.19) positive?

(d) Why is the local energy condition satisfied?

(e) Why do vacuum measures (see Definition 1.4) describe a flat space-time?

In order to address these questions, we need to be more specific and focus on causal fermion systems constructed in a static Lorentzian spacetime, re-ferred to as *static Dirac systems* (for details see [18, Section 1.2], [19] or Section 2.3 in the preliminaries). For such systems, we can answer question (a) as follows:

Theorem 1.10. *For a static Dirac system describing a four-dimensional spacetime which is asymptotically Schwarzschild (for details see the beginning of Section 6), the total mass is proportional to the ADM mass*

$$\mathfrak{M} = c \, \mathfrak{M}_{\mathrm{ADM}} \, ,$$

where the constant c can be computed from the Lagrangian in the regularized, spherically symmetric Minkowski vacuum via the formula

(1.22) $$c = \frac{1}{4\pi} \int_{\mathcal{M}} |\mathbf{y}|^2 \, \mathcal{L}_\kappa \big(0, (t, \mathbf{y}) \big) \, dt \, d^3 \mathbf{y} \; > \; 0 \, .$$

Our method of proof also explains why for static Dirac systems, the total mass is computable purely from the knowledge of the metric near infinity.

In order to address the other questions **(b)**–**(e)**, one must make use of the scaling behavior of static Dirac systems as worked out in Section 7, based on previous results in [18, §4.2.5], [20] and [8, Appendix A]. The relevant

length scales are given by

$$(1.23) \qquad \begin{cases} \text{Compton length} & m^{-1} \\ \text{Planck length} & \delta \\ \text{regularization length} & \varepsilon \,, \end{cases}$$

where m denotes the mass of the Dirac particles. Moreover, the measure μ determines a length scale as the radius of a ball of μ-volume one (here we fix the freedom in rescaling μ by specifying $\mathfrak{s}$ as well as the parameter c of the local trace; for details see Section 3.4). Working on this fixed length scale, we have three dimensionless parameters m, δ and ε.

(b) For static Dirac systems, the relation (1.19) can be derived as follows. In Definition 1.7 we consider a family $(\mu_\tau)_{\tau \in (-1,1)}$ of critical measures for a decreasing value of κ, (1.18). In this family, the parameters in (1.23) will in general change; we denote them by $m(\tau)$, $\delta(\tau)$ and $\varepsilon(\tau)$. Similarly, the family $(\tilde{\mu}_\tau)_{\tau \in (-1,1)}$ of measures which describes curved spacetime, involves the parameters $\tilde{m}(\tau)$, $\tilde{\delta}(\tau)$ and $\tilde{\varepsilon}(\tau)$. Since these parameters are constant in spacetime, their τ-dependence can be determined asymptotically near infinity, where spacetime goes over to Minkowski space. In other words, the τ-dependence of these parameters is the same in curved and in flat spacetime, i.e. $\tilde{m}(\tau) = m(\tau)$, $\tilde{\delta}(\tau) = \delta(\tau)$ and $\tilde{\varepsilon}(\tau) = \varepsilon(\tau)$. Therefore, taking the difference of the jets $\tilde{v}$ and v which describe the infinitesimal variations in curved and flat spacetime, respectively, on the right side of (1.19) the variation of the parameters m, δ and ε drops out. But the fact that $\tilde{v}$ changes these parameters in the spacetime described by $\tilde{\mu}$ has the effect that this spacetime is modified depending on the distribution of matter and gravity. This can be understood most easily in the case of the so-called *natural scaling of δ* where the dimensionless parameter $m\delta$ remains constant. In this case, the system remains unchanged in Planck units where the gravitational constant is fixed. However, in these units the length scale determined by $\tilde{\mu}$ changes. In other words, the gravitational system is fixed, but the volume $\tilde{\mu}$ changes. Since the $\tilde{\mu}$-volume is fixed in the definition of the total mass, this implies that in Planck units, the radius of the set $\tilde{\Omega}$ in (1.16) changes. This in turn gives rise to a change of the total mass, explaining (1.19). This argument is given in more detail and for more general scalings in Section 7.

(c) Working out the above scalings in more detail, in Section 7 we find that the gravitational constant has the same sign as

$$\frac{d}{d\tau}\big(m\delta^2\big)\big|_{\tau=0}$$

(again for variations normalized according to (1.18)). Moreover, we know from the general structure of the causal action principle that the parameter m decreases. Therefore, the gravitational coupling constant is positive for a natural scaling of δ. More generally, the gravitational coupling constant is positive if the scaling of δ does not differ too much from natural scaling. More details are given in Remark 7.1.

(d) One method to deal with the local energy condition in Definition 1.8 is to proceed as in general relativity by taking it as a condition motivated from physical observations which is to be verified case by case for different types of matter. But in the context of causal fermion systems, one can go a step further and try to explain the inequality (1.20) from the minimality of the causal action. In Remark 7.2, this argument is worked out for homogeneous matter distributions. More generally, this argument shows that the energy conditions hold for all matter densities which are nearly constant on the Compton scale. However, at present we cannot rule out the possibility that the energy density might be negative on microscopic scales.

(e) For static Dirac systems, it can be analyzed in detailed what the vacuum condition (1.14) means. By direct computation, one verifies that curvature increases ℓ. Therefore, the condition (1.14) is satisfied only for systems in flat Minkowski space. This argument is given in more detail in Remark 7.3.

The paper is organized as follows. Section 2 provides the necessary preliminaries on causal variational principles and causal fermion systems. In Section 3 we specialize causal fermion systems to the static case and explain how they fit to the general setting of causal variational principles. In Section 4 we prove that the total mass is well-defined and independent of exhaustions, the inner geometry and the identifications of the Hilbert spaces of the two causal fermion systems. In Section 5 we introduce the equations of linearized gravity and use them to prove Theorem 1.9. In Section 6 we compute the total mass in the example of an asymptotically Schwarzschild spacetime and prove Theorem 1.10. In Section 7 we work out the relevant

scalings. Moreover, we discuss the sign of the gravitational coupling constant (Remark 7.1), show that homogeneous perturbations of Minkowski space must satisfy the local energy condition (Remark 7.2) and specify the assumptions under which a vacuum measure describes a flat spacetime (Remark 7.3). In the appendices, we work out the scaling behavior of the total mass for static Dirac systems (Appendix A) and analyze the fermionic projector and the relevant surface layer integrals for linearized gravity (Appendix B).

2. Preliminaries

We now recall the basics on causal variational principles in the setting needed here. More details can be found in [26, 28]. We use a slightly different notation in order to get consistency with the causal variational principle in the static case as will be introduced in Section 3.

2.1. Causal variational principles in the non-compact setting

We consider causal variational principles in the non-compact setting as introduced in [26, 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

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

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

$$\big|\tilde{\mu} - \mu\big|(\mathcal{G}) < \infty \qquad \text{and} \qquad \big(\tilde{\mu} - \mu\big)(\mathcal{G}) = 0$$

(where $|.|$ denotes the total variation of a measure). These variations of the causal action are well-defined. The existence theory for minimizers is developed in [30]. It is shown in [26, 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

$$(2.2) \qquad \qquad \ell(\mathbf{x}) := \int_{\mathcal{G}} \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) - \mathfrak{s}$$

is minimal and vanishes on the support of μ,

$$(2.3) \qquad \qquad \ell|_N \equiv \inf_{\mathcal{G}} \ell = 0 \, .$$

For further details we refer to [26, Section 2].

2.1.1. The weak Euler-Lagrange equations and jet spaces.   We denote the support of μ by N,

$$(2.4) \qquad \qquad N := \operatorname{supp} \mu \, \subset \, \mathcal{G} \, .$$

The EL equations (2.3) 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 *weak EL equations* introduced in [26, 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 weak 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.

$$(2.5) \qquad\qquad \ell|_N \equiv 0 \qquad \text{and} \qquad D\ell|_N \equiv 0$$

(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 $\mathfrak{u} := (a, u)$ consisting of a real-valued function a on N and a vector field u on $T\mathcal{G}$ along N, and to denote the combination of multiplication and directional derivative by

$$(2.6) \qquad\qquad \nabla_\mathfrak{u}\ell(\mathbf{x}) := a(\mathbf{x})\,\ell(\mathbf{x}) + \big(D_u\ell\big)(\mathbf{x}) \,.$$

Then the equations (2.5) imply that $\nabla_\mathfrak{u}\ell(\mathbf{x})$ vanishes for all $\mathbf{x} \in N$. The pair $\mathfrak{u} = (a, u)$ is referred to as a *jet*.

In the general lower-continuous setting, one must be careful because the directional derivative $D_u\ell$ in (2.6) 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

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

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 a smooth extension 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,

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

This gives rise to the jet space

$$(2.7) \qquad\qquad \mathfrak{J}^{\mathrm{diff}} := C^\infty(N, \mathbb{R}) \oplus \Gamma^{\mathrm{diff}} \subset \mathfrak{J} \,.$$

For the jets in $\mathfrak{J}^{\mathrm{diff}}$, the combination of multiplication and directional derivative in (2.6) is well-defined. We choose a linear subspace $\mathfrak{J}^{\mathrm{test}} \subset \mathfrak{J}^{\mathrm{diff}}$ with the

property that its scalar and vector components are both vector spaces,

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

and the scalar component is nowhere trivial in the sense that

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

Then the *weak EL equations* read (for details cf. [26, (eq. (4.10)]])

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

The purpose of introducing $\mathfrak{J}^{\mathrm{test}}$ 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 $\mathfrak{J}^{\mathrm{test}}$ to be composed of jets pointing in directions where the microscopic fluctuations of ℓ are disregarded.

We finally point out that the weak EL equations (2.9) do not hold only for minimizers, but also for critical points of the causal action. With this in mind, all methods and results of this paper do not apply only to minimizers, but more generally to critical points of the causal variational principle. For brevity, we also refer to a measure with satisfies the weak EL equations (2.9) as a *critical measure*.

We conclude this section by introducing a few jet spaces and specifying differentiability conditions which will be needed later on. We begin with the spaces $\mathfrak{J}^{\ell}$, where $\ell \in \mathbb{N}_0 \cup \{\infty\}$ can be thought of as the order of differentiability if the derivatives act simultaneously on both arguments of the Lagrangian:

Definition 2.1. *For any $\ell \in \mathbb{N}_0 \cup \{\infty\}$, the jet space $\mathfrak{J}^{\ell} \subset \mathfrak{J}$ is defined as the vector space of test jets with the following properties:*

(i) *For all $\mathbf{y} \in N$ and all $\mathbf{x}$ in an open neighborhood of N, directional derivatives*

(2.10) $$\left(\nabla_{1,\mathfrak{v}_1} + \nabla_{2,\mathfrak{v}_1}\right) \cdots \left(\nabla_{1,\mathfrak{v}_p} + \nabla_{2,\mathfrak{v}_p}\right) \mathcal{L}(\mathbf{x}, \mathbf{y})$$

(computed componentwise in distinguished charts around $\mathbf{x}$ and $\mathbf{y}$; for details see [31, Section 5.2]) exist for all $p \in \{1, \ldots, \ell\}$ and all $\mathfrak{v}_1, \ldots, \mathfrak{v}_p \in \mathfrak{J}^{\ell}$.

(ii) *The functions in* (2.10) *are μ-integrable in the variable* **y**, *giving rise to locally bounded functions in* **x**. *More precisely, these functions are in the space*

$$L_{\mathrm{loc}}^{\infty}\Big(M, L^1\big(M, d\rho(y)\big); d\rho(x)\Big) .$$

(iii) *Integrating the expression* (2.10) *in* **y** *over N with respect to the measure μ, the resulting function (defined for all* **x** *in an open neighborhood of N) is continuously differentiable in the direction of every jet $\mathfrak{u} \in \mathfrak{J}^{\mathrm{test}}$.*

Here and throughout this paper, we use the following conventions for partial derivatives and jet derivatives:

▸ Partial and jet derivatives with an index $i \in \{1,2\}$, as for example in (2.10), only act on the respective variable of the function $\mathcal{L}$. This implies, for example, that the derivatives commute,

$$\nabla_{1,\mathfrak{v}}\nabla_{1,\mathfrak{u}}\mathcal{L}(\mathbf{x},\mathbf{y}) = \nabla_{1,\mathfrak{u}}\nabla_{1,\mathfrak{v}}\mathcal{L}(\mathbf{x},\mathbf{y}) .$$

▸ The partial or jet derivatives which do not carry an index act as partial derivatives on the corresponding argument of the Lagrangian. This implies, for example, that

$$\nabla_{\mathfrak{u}} \int_{\mathcal{G}} \nabla_{1,\mathfrak{v}}\, \mathcal{L}(\mathbf{x},\mathbf{y})\, d\mu(\mathbf{y}) = \int_{\mathcal{G}} \nabla_{1,\mathfrak{u}}\nabla_{1,\mathfrak{v}}\, \mathcal{L}(\mathbf{x},\mathbf{y})\, d\mu(\mathbf{y}) .$$

We point out that (in contrast to the method and conventions used in [26]) *jets are never differentiated.*

In order for all integral expressions to be well-defined, we impose throughout that the space $\mathfrak{J}^{\mathrm{test}}$ has the following properties (for details see [27, Section 3.5]).

Definition 2.2. *Let $m \in \mathbb{N}$. The jet space $\mathfrak{J}^{\mathrm{test}}$ is* **surface layer regular** *if $\mathfrak{J}^{\mathrm{test}} \subset \mathfrak{J}^2$ (see Definition 2.1) and if for all $\mathfrak{u}, \mathfrak{v} \in \mathfrak{J}^{\mathrm{test}}$ and all $p \in \{1,2\}$ the following conditions hold:*

(i) *The directional derivatives*

(2.11) $$\nabla_{1,\mathfrak{u}} \big(\nabla_{1,\mathfrak{v}} + \nabla_{2,\mathfrak{v}}\big)^{p-1}\mathcal{L}(\mathbf{x},\mathbf{y})$$

exist.

(ii) *The functions in* (2.11) *are μ-integrable in the variable* **y**, *giving rise to locally bounded functions in* **x**. *More precisely, these functions are in*

the space

$$L^\infty_{\text{loc}}\Big(L^1\big(N, d\mu(\mathbf{y})\big), d\mu(\mathbf{x})\Big) .$$

(iii) *The $\mathfrak{u}$-derivative in (2.11) may be interchanged with the $\mathbf{y}$-integration,
i.e.*

$$\int_N \nabla_{1,\mathfrak{u}} \big(\nabla_{1,\mathfrak{v}} + \nabla_{2,\mathfrak{v}}\big)^{p-1} \mathcal{L}(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y})$$

$$= \nabla_\mathfrak{u} \int_N \big(\nabla_{1,\mathfrak{v}} + \nabla_{2,\mathfrak{v}}\big)^{p-1} \mathcal{L}(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y}) .$$

2.1.2. The linearized field equations. Usually, linearized fields are ob-
tained by considering a family of nonlinear solutions and linearizing with
respect to a parameter τ which describes the field strength. The analogous
notion in the setting of causal fermion systems is a linearization of a family
of measures $(\tilde\mu_\tau)_{\tau \in [0,1)]}$ which all satisfy the weak EL equations (2.9) (for
fixed values of the parameters κ and $\mathfrak{s}$). It turns out to be fruitful to con-
struct this family of measures by multiplying a given critical measure μ by
a weight function f_τ and then "transporting" the resulting measure with a
mapping F_τ. More precisely, we consider the ansatz

$$(2.12) \qquad \tilde\mu_\tau = (F_\tau)_* \big(f_\tau\, \mu\big) ,$$

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 [6, Section 3.6]).

The property of the family of measures $(\tilde\mu_\tau)_{\tau \in [0,1)}$ of the form (2.12) to
satisfy the weak EL equation for all τ means infinitesimally in τ that the
jet $\mathfrak{v}$ defined by

$$(2.13) \qquad \mathfrak{v} = (b, v) := \frac{d}{d\tau}(f_\tau, F_\tau)\big|_{\tau=0}$$

satisfies the *linearized field equations*. We now recall the main step of the
construction. Using the definition of the push-forward measure, we can write
the weak EL equations (2.9) for the measure $\tilde\mu_\tau$ as

$$\nabla_\mathfrak{u}\bigg(\int_N \mathcal{L}\big(F_\tau(\mathbf{x}), F_\tau(\mathbf{y})\big)\, f_\tau(\mathbf{y})\, d\mu - \mathfrak{s}\bigg) = 0 .$$

Since the function ℓ vanishes on the support, we may multiply by $f_\tau(\mathbf{x})$ to
obtain

$$(2.14) \qquad \nabla_\mathfrak{u}\bigg(\int_N f_\tau(\mathbf{x})\, \mathcal{L}\big(F_\tau(\mathbf{x}), F_\tau(\mathbf{y})\big)\, f_\tau(\mathbf{y})\, d\mu - f_\tau(\mathbf{x})\, \mathfrak{s}\bigg) = 0 .$$

At this point, the technical complication arise that one must specify the τ-dependence of the jet spaces, and moreover the last transformation makes it necessary to transform the jet spaces. Here we do not enter the details but refer instead to the rigorous derivation in [21, Section 3.3] or to the simplified presentation in the smooth setting in the textbook [28, Chapter 6]. Differentiating (2.14) with respect to τ gives the linearized field equations

$$(2.15) \qquad \langle \mathfrak{u}, \Delta \mathfrak{v} \rangle |_N = 0 \qquad \text{for all } \mathfrak{u} \in \mathfrak{J}^{\text{test}} \,,$$

where

$$(2.16) \qquad \langle \mathfrak{u}, \Delta \mathfrak{v} \rangle (\mathbf{x}) := \nabla_{\mathfrak{u}} \left(\int_N \left(\nabla_{1,\mathfrak{v}} + \nabla_{2,\mathfrak{v}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) - \nabla_{\mathfrak{v}} \, \mathfrak{s} \right) .$$

We denote the vector space of all solutions of the linearized field equations by $\mathfrak{J}^{\text{lin}} \subset \mathfrak{J}^1$.

2.1.3. A conserved surface layer integral for linearized solutions.
In the setting of causal fermion systems, the usual integrals over hypersurfaces in spacetime are undefined. Instead, one considers so-called *surface layer integrals*, being double integrals of the form

$$(2.17) \qquad \int_\Omega d\mu(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, (\cdots) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \,,$$

where Ω is a Borel subset of N, and $(\cdots)$ stands for a differential operator acting on the Lagrangian. The structure of such surface layer integrals can be understood most easily in the special situation that the Lagrangian is of short range in the sense that $\mathcal{L}(\mathbf{x}, \mathbf{y})$ vanishes unless $\mathbf{x}$ and $\mathbf{y}$ are close together. In this situation, we get a contribution to the double integral (2.17) only if both $\mathbf{x}$ and $\mathbf{y}$ are close to the boundary $\partial \Omega$. With this in mind, surface layer integrals can be understood as an adaptation of surface integrals to the setting of causal variational principles (for a more detailed explanation see [25, Section 2.3]).

Surface layer integrals were first introduced in [25] in order to formulate Noether-like theorems for causal variational principles. In particular, it was shown that there is a conserved surface layer integral which generalizes the Dirac current conservation in relativistic quantum mechanics (see [25, Section 5]). More recently, in [26] another conserved surface layer integral was discovered which gives rise to a symplectic form on the solutions of the linearized field equations (see [26, Sections 3.3 and 4.3]). A systematic study of conservation laws for surface layer integrals is given in [27]. The

conservation law which is most relevant for our purposes is summarized in the next lemma. For a compact subset $\Omega \subset N$ and a jet $\mathfrak{u} \in \mathfrak{J}^1$ we introduce the surface layer integral

$$(2.18) \qquad \gamma_\mu^\Omega(\mathfrak{v}) := \int_\Omega d\mu(\mathbf{x}) \int_{N\setminus\Omega} d\mu(\mathbf{y}) \left(\nabla_{1,\mathfrak{u}} - \nabla_{2,\mathfrak{u}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y})$$

Lemma 2.3. *For every compact $\Omega \subset N$ and any linearized solution $\mathfrak{u} \in \mathfrak{J}^{\mathrm{lin}}$,*

$$(2.19) \qquad \gamma_\mu^\Omega(\mathfrak{v}) = \int_\Omega \nabla_{\mathfrak{u}}\, \mathfrak{s}\, d\mu \, .$$

Proof. In view of the anti-symmetry of the integrand,

$$\int_\Omega d\mu(\mathbf{x}) \int_\Omega d\mu(\mathbf{y}) \left(\nabla_{1,\mathfrak{u}} - \nabla_{2,\mathfrak{u}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) = 0 \, .$$

Adding this equation to the left side of (2.19), we obtain

$$\int_\Omega d\mu(\mathbf{x}) \int_{N\setminus\Omega} d\mu(\mathbf{y}) \left(\nabla_{1,\mathfrak{u}} - \nabla_{2,\mathfrak{u}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$= \int_\Omega d\mu(\mathbf{x}) \int_N d\mu(\mathbf{y}) \left(\nabla_{1,\mathfrak{u}} - \nabla_{2,\mathfrak{u}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y})$$
$$= \int_\Omega d\mu(\mathbf{x}) \left(2 \nabla_{\mathfrak{u}} \left(\ell(\mathbf{x}) + \mathfrak{s} \right) - \left(\Delta \mathfrak{u} \right)(\mathbf{x}) - \nabla_{\mathfrak{u}}\, \mathfrak{s} \right) ,$$

where in the last line we used the definitions of ℓ and Δ (see (2.2) and (2.16)). Applying the weak EL equations (2.9) and the linearized field equations (2.15) gives the result. $\qquad \square$

2.1.4. Inner solutions. We again define N as the support of μ, (2.4). Furthermore we make the following simplifying assumption:

Definition 2.4. *Spacetime $N := \mathrm{supp}\,\mu$ has a* **smooth manifold structure** *if the following conditions hold:*

(i) *N is a k-dimensional smooth, oriented and connected submanifold of $\mathcal{G}$.*

(ii) *In a chart $(\mathbf{x}, U)$ of N, the universal measure is absolutely continuous with respect to the Lebesgue measure with a smooth, strictly positive weight function,*

$$(2.20) \qquad d\mu = h(x)\, d^k x \qquad \text{with} \quad h \in C^\infty(N, \mathbb{R}^+) \, .$$

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

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

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

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

(where, using the Einstein summation convention, we sum over $\alpha = 1, \dots, k$). The jets of the form $\mathfrak{v} := (\operatorname{div} v, v)$ are of particular significance. 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 $f \in C_0^1(N, \mathbb{R})$

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

Integrating by parts formally, one finds that these jets satisfy the linearized field equations,

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

In the last step we used that $\nabla_{D_v \mathfrak{u}} \ell(\mathbf{x})$ vanishes by the EL equations. Moreover, the function $\nabla_\mathfrak{u} \ell$ vanishes identically on M in view of the weak EL equations. Therefore, it is differentiable in the direction of every vector field on M, and this directional derivative is zero.

The above formal computation has two shortcoming. First, it is a-priori not clear whether integrating by parts gives boundary terms. Moreover, we need to be careful because the individual derivatives do not need to exist. This is why, in order to give the above computation a mathematical meaning, we need to impose additional technical assumptions. We now specify these assumptions and prove that the resulting jets are indeed solutions of the linearized field equations. For technical simplicity, we restrict attention to the case that μ has a smooth manifold structure (see Definition 2.4) and one asymptotic end (see Definition 1.3); for a more general presentation see [22,

Section 3]. In preparation, we let $\Gamma_{\mathbf{x}}$ be the subspace of the tangent space spanned by the test jets,

$$(2.21) \qquad \Gamma_{\mathbf{x}} := \big\{ u(\mathbf{x}) \,|\, u \in \Gamma^{\text{test}} \big\} \subset T_{\mathbf{x}}\mathcal{F} \,.$$

Similar to our assumption that the scalar components of the test jets is nowhere trivial (2.8), it is sensible and useful to assume that this subspace of the tangent space contains all the tangent vectors to M,

$$(2.22) \qquad T_{\mathbf{x}}M \subset \Gamma_{\mathbf{x}} \qquad \text{for all } \mathbf{x} \in M \,.$$

We introduce a Riemannian metric $g_{\mathbf{x}}$ on $\mathfrak{J}_{\mathbf{x}}$. This Riemannian metric also induces a pointwise scalar product on the jets. Namely, setting

$$(2.23) \qquad \mathfrak{J}_{\mathbf{x}} := \mathbb{R} \oplus \Gamma_{\mathbf{x}} \,,$$

we obtain the scalar product on $\mathfrak{J}_{\mathbf{x}}$

$$(2.24) \quad \langle \mathfrak{v}, \tilde{\mathfrak{v}} \rangle_{\mathbf{x}} : \mathfrak{J}_{\mathbf{x}} \times \mathfrak{J}_{\mathbf{x}} \to \mathbb{R} \,, \qquad \langle \mathfrak{v}, \tilde{\mathfrak{v}} \rangle_{\mathbf{x}} := b(\mathbf{x})\, \tilde{b}(\mathbf{x}) + g_{\mathbf{x}}\big(v(\mathbf{x}), \tilde{v}(\mathbf{x})\big) \,.$$

We denote the corresponding norm by $\|.\|_{\mathbf{x}}$. We assume that this Riemannian metric is *adapted in the asymptotic end* in the sense that its restriction to TN is equivalent to the Euclidean metric in the asymptotic end, i.e. in the chart Φ in Definition 1.3 there is a constant $C > 0$ such that

$$(2.25)$$
$$\frac{1}{C} \sum_{j=1}^{k} \big|u^j\big|^2 \leq g_{\mathbf{x}}(u, u) \leq C \sum_{j=1}^{k} \big|u^j\big|^2 \qquad \text{for all } \mathbf{x} \in N \setminus I \text{ and } u \in T_{\mathbf{x}}N \,.$$

Definition 2.5. *An **inner solution** is a jet $\mathfrak{v}$ of the form*

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

We make the following regularity and decay assumptions:

(i) *The vector field v can be extended to a vector field $\tilde{v} \in \Gamma(U, T\mathcal{F})$ defined in a neighborhood U of N such that the directional derivative $(D_{1,\tilde{v}} + D_{2,\tilde{v}})\mathcal{L}(\mathbf{x}, \mathbf{y})$ exists for all $\mathbf{x} \in U$ and $\mathbf{y} \in N$ and is integrable in $\mathbf{y}$, i.e.*

$$\int_N \big|(D_{1,\tilde{v}} + D_{2,\tilde{v}})\mathcal{L}(\mathbf{x}, \mathbf{y})\big|\, d\mu(\mathbf{y}) < \infty \qquad \text{for all } \mathbf{x} \in U \,.$$

Moreover, the directional derivative $D_{\tilde{v}}\ell(\mathbf{x})$ exists for all $\mathbf{x} \in U$ and is continuous in U.

(ii) *The integral*

$$\int_N \mathcal{L}(\mathbf{x}, \mathbf{y}) \, \|\mathfrak{v}(\mathbf{y})\|_{\mathbf{y}} \, d\mu(\mathbf{y})$$

is finite and bounded locally uniformly in a neighborhood of N (where $\|.\|_{\mathbf{y}}$ is again the norm corresponding to the scalar product (2.24) adapted in the asymptotic end according to (2.25)).

(iii) *For any test jet $\mathfrak{u} \in \mathfrak{J}^{\text{test}}$, the directional derivative $D_v \mathfrak{u}$ (computed in the same charts used for computing the higher derivatives in Definition 2.1) is again in $\mathfrak{J}^{\text{test}}$.*

The vector space of all inner solutions is denoted by $\mathfrak{J}^{\text{in}}$.

Note that (i) implies that every inner solution is in $\mathfrak{J}^1 \cap \mathfrak{J}^{\text{diff}}$ (see (2.7) and Definition 2.1).

The name "inner *solution*" is justified by the following lemma:

Lemma 2.6. *Every inner solution $\mathfrak{v} \in \mathfrak{J}^{\text{in}}$ is a solution of the linearized field equations, i.e.*

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

Proof. We choose a monotone decreasing function $\eta \in C_0^\infty(\mathbb{R})$ with

$$\eta_{[0,1]} \equiv 1 \qquad \text{and} \qquad \text{supp}\, \eta \subset (-2, 2) \, .$$

For $n \in \mathbb{N}$, we introduce the compactly supported cutoff functions

$$\eta_n(\mathbf{x}) := \left\{ \begin{array}{ll} 1 & \text{if } \mathbf{x} \in I \\[2mm] \eta\left(\dfrac{\|\Phi(\mathbf{x})\|_{\mathbb{R}^k} - R}{n} \right) & \text{if } \mathbf{x} \in N \setminus I \end{array} \right.$$

(where R is the radius of the ball in (1.13)). By construction, the functions η_n are monotone increasing and exhaust N in the sense that for any compact set $K \subset N$ there is N with $\eta_n|_K \equiv 1$ for all $n \geq N$. Moreover, using (2.25), the derivatives tend uniformly to zero, i.e.

$$(2.26) \qquad \lim_{n \to \infty} \sup_{\mathbf{x} \in N} \|D\eta_n(\mathbf{x})\|_{\mathbf{x}} = 0 \, .$$

Our first goal is to prove that for any $\mathbf{x} \in U$,

$$(2.27) \qquad \int_N \left(D_{1,\tilde{v}} + \nabla_{2,\mathfrak{v}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\rho(\mathbf{y}) = D_{\tilde{v}}\left(\ell + \mathfrak{s} \right)(\mathbf{x}) \, .$$

Making use of Definition 2.5 (i), we know from Lebesgue's dominated convergence theorem that for any $\mathbf{x} \in U$,

$$A(\mathbf{x}) := \int_N \left(D_{1,\tilde{v}} + \nabla_{2,\tilde{\mathfrak{v}}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y})$$

$$= \lim_{n \to \infty} \int_N \left(D_{1,\tilde{v}} + \nabla_{2,\tilde{\mathfrak{v}}} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \, \eta_n(\mathbf{y}) \, d\mu(\mathbf{y}) \, .$$

Now we can integrate by parts to obtain
(2.28)
$$A(\mathbf{x}) = \lim_{n \to \infty} \left(D_{\tilde{v}} \int_N \mathcal{L}(\mathbf{x}, \mathbf{y}) \, \eta_n(\mathbf{y}) \, d\mu(\mathbf{y}) - \int_N \mathcal{L}(\mathbf{x}, \mathbf{y}) \left(D_v \eta_n(\mathbf{y}) \right) d\mu(\mathbf{y}) \right)$$

(here one needs to pull out the derivative $D_{\tilde{v}}$ before the integral, because the Lagrangian need not be differentiable; the integral, on the other hand, is well-defined because the last integral is). The last integral can be estimated by
(2.29)
$$\left| \int_N \mathcal{L}(\mathbf{x}, \mathbf{y}) \left(D_v \eta_n(\mathbf{y}) \right) d\mu(\mathbf{y}) \right| \leq \sup_N \| D_v \eta_n \| \int_N \| v(\mathbf{y}) \|_{\mathbf{y}} \, \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) \, .$$

According to Definition 2.5 (ii), the obtained integral is bounded locally uniformly in $\mathbf{x}$. Using (2.26), we conclude that the last integral in (2.28) tends to zero as $n \to 0$, locally uniformly in $\mathbf{x}$.

As a consequence, also the first integral in (2.28) converges as $n \to 0$, locally uniformly in $\mathbf{x}$. In order to prove (2.27), it remains to show that this limit is given by

$$(2.30) \qquad \lim_{n \to \infty} \left(D_{\tilde{v}} \int_N \mathcal{L}(\mathbf{x}, \mathbf{y}) \, \eta_n(\mathbf{y}) \, d\mu(\mathbf{y}) \right) = D_{\tilde{v}} \left(\ell + \mathfrak{s} \right)(\mathbf{x}) \, .$$

Assume conversely that this equation does not hold for all $\mathbf{x} \in U$. Then by continuity (note that the left side of (2.30) is continuous as a locally uniform limit of continuous function, as is the right side by Definition 2.5 (i)), the equation (2.30) is violated in an open set. Therefore, we may choose a path $\gamma : [t_0, t_1] \to U$ along the integral curves of $\tilde{v}$ such that

$$\int_{t_0}^{t_1} \lim_{n \to \infty} \left(D_{\tilde{v}} \int_N \mathcal{L}\left(\gamma(t), \mathbf{y} \right) \left(1 - \eta_n(\mathbf{y}) \right) d\mu(\mathbf{y}) \right) dt \neq 0 \, .$$

Due to the locally uniform convergence, we may interchange the integral and the limit to conclude that

$$\lim_{n\to\infty} \int_N \mathcal{L}(\mathbf{x}, \mathbf{y}) \left(1 - \eta_n(\mathbf{y})\right) d\mu(\mathbf{y}) \Big|_{\mathbf{x}=\gamma(t_0)}^{\mathbf{x}=\gamma(t_1)} \neq 0 \,.$$

On the other hand, using assumption (iv) on page 1750, the limits on the left vanish using Lebesgue's dominated convergence theorem. This is a contradiction. Hence (2.30) holds. This concludes the proof of (2.27).

We rewrite (2.27) as

$$\Delta \tilde{\mathfrak{v}}(\mathbf{x}) = \nabla_{\tilde{\mathfrak{v}}} \ell(\mathbf{x}) \qquad \text{for all } \mathbf{x} \in U$$

(where the scalar component of $\mathfrak{v}$ can be extended to U arbitrarily). The next and final step is to show that for any $\mathfrak{u} \in \mathfrak{J}^{\text{test}}$ and $\mathbf{x} \in N$, the jet derivative $\nabla_{\mathfrak{u}}$ of this equation exists and vanishes. To this end, we write the jet derivative of the right side as

$$\nabla^2 \ell|_{\mathbf{x}}(\mathfrak{u}, \mathfrak{v}) = \nabla_{\mathfrak{v}(\mathbf{x})} \left(\nabla_{\mathfrak{u}} \ell(\mathbf{x})\right) - \nabla_{D_v \mathfrak{u}} \ell(\mathbf{x})$$

(where the first summand on the right is an iterated directional derivative). The last summand vanishes because of the weak EL equations, using that $D_v \mathfrak{u} \in \mathfrak{J}^{\text{test}}$ (see Definition 2.5 (iii)). In order to treat the first summand, we note that the function $\nabla_{\mathfrak{u}} \ell$ vanishes identically on N by the weak EL equations. Therefore, this function is differentiable in the direction of every vector field on N, and this directional derivative is zero. This concludes the proof. $\qquad \square$

We next show that for any function a on N one can find an inner solution whose scalar component coincides with a. If N were compact, the analogous statement would be the infinitesimal version of Moser's theorem (see for example [40, Section XVIII, §2]). Here we give a detailed proof if N is non-compact with one asymptotic end, based on [40, Theorem 1.2 in Section XVIII].

Lemma 2.7. *Assume that $N := \operatorname{supp} \nu$ has a smooth manifold structure (see Definition 2.4) and one asymptotic end (see Definition 1.3). Then for any $a \in C^\infty(N)$ there is a vector field $v \in \Gamma(N, TN)$ such that $\operatorname{div} v = a$.*

Proof. We choose a partition of unity $(\phi_n)_{n\in\mathbb{N}}$ of N supported in annuli of our coordinate system Φ. More precisely,

$$\operatorname{supp}\phi_1 \subset I \cup \Phi^{-1}\big(B_{R+2}\big) \qquad \text{and}$$
$$\operatorname{supp}\phi_n \subset \Phi^{-1}\big(B_{R+n+1} \setminus \overline{B_{R+n-1}}\big) \qquad \text{for } n \geq 2 .$$

Due to the smoothness assumption (2.20), the measure μ can be represented by a volume form $\psi \in \Lambda^k(N)$, i.e.

$$\mu(U) = \int_U \psi \qquad \text{for all compact } U \subset N .$$

Likewise, the measure $a\mu$ can be represented by a volume form $\omega \in \Lambda^k(N)$, i.e.

$$\int_U a(\mathbf{x})\, d\mu(\mathbf{x}) = \int_U \omega$$

(again valid for all compact $U \subset N$). We now proceed inductively: We choose a real number c_1 such that

$$\int_N \big(\phi_1\,\omega - c_1\,\phi_1\,\psi\big) = 0 .$$

According to [40, Theorem 1.2 in Section XVIII], there is a compactly supported $(k-1)$-form $\eta_1 \in \Lambda_0^{k-1}(N)$ such that

$$\phi_1\,\omega - c_1\,\phi_1\,\psi = d\eta_1 .$$

In the induction step from n to $n+1$ we choose $c_{n+1} \in \mathbb{R}$ such that

$$\int_N \big(\phi_{n+1}\,\omega + c_n\,\phi_n\,\psi - c_{n+1}\,\phi_{n+1}\,\psi\big) = 0 .$$

Then there exists $\eta_{n+1} \in \Lambda_0^{k-1}(N)$ with

$$\phi_{n+1}\,\omega + c_n\,\phi_n\,\psi - c_{n+1}\,\phi_{n+1}\,\psi = d\eta_{n+1} .$$

By applying [40, Theorem 1.2 in Section XVIII] on the manifold chosen as the above annuli, one sees that the support of the η_n can be arranged to also lie in these annuli, i.e.

$$\operatorname{supp}\eta_1 \subset I \cup \Phi^{-1}\big(B_{R+2}\big) \qquad \text{and}$$
$$\operatorname{supp}\eta_n \subset \Phi^{-1}\big(B_{R+n+1} \setminus \overline{B_{R+n-1}}\big) \qquad \text{for } n \geq 2$$

(here we make use of the fact that all the annuli are connected).

Being locally finite, we can carry out the sum over n. The summands involving $\phi_n \psi$ cancel. We conclude that

$$\eta := \sum_{n=1}^{\infty} \eta_n \qquad \text{satisfies} \qquad d\eta = \sum_{n=1}^{\infty} \phi_n \, \omega = \omega \, .$$

Finally, we need to identify η with a vector field such that $\operatorname{div} v \, d\mu = d\eta$. To this end, we choose a Riemannian metric g on N. By a conformal transformation we can arrange that the corresponding volume form coincides with the measure μ. Now we choose the vector field as

$$v^\alpha = g^{\alpha\beta} \, (*\eta)_\beta \, ,$$

where $* : \Lambda^{k-1}(N) \to \Lambda^1(N)$ is the Hodge star. This concludes the proof. $\quad\square$

We point out that in general, the vector field v constructed in this lemma is not compactly supported, even if the function a is. In particular, if the integral of a is non-zero, then the Gauss divergence theorem implies that the flux of v through large coordinate spheres must be non-zero.

In what follows, we always assume that the vector field constructed in this lemma satisfies all the regularity and decay assumptions in Definition 2.5. We then obtain a corresponding inner solution

$$\mathfrak{v} := (a, v) \in \mathfrak{J}^{\mathrm{in}} \, .$$

In view of Lemma 2.7, given any jet we can arrange by adding a suitable inner solution that the scalar component of the jet vanishes. With this in mind, in what follows we may always restrict attention to jets with vanishing scalar component. However, one must keep in mind that the resulting vector components will not be zero near infinity. Indeed, the asymptotics of these vector fields will encode the total mass.

2.1.5. A nonlinear surface layer integral.   We finally mention another surface layer integral which was first introduced in [22]. Instead of working with linearized solutions, we directly compare the perturbed measure $\tilde{\mu}$, which takes into account the nonlinear interaction, with a vacuum measure μ.

Definition 2.8. *We let μ and $\tilde{\mu}$ be two Borel measures on $\mathcal{G}$ and set $N :=$ $\operatorname{supp}\mu$, $\tilde{N} := \operatorname{supp}\tilde{\mu}$. Given compact subsets $\Omega \subset N$ and $\tilde{\Omega} \subset \tilde{N}$, the* **nonlinear surface layer integral** $\gamma^{\tilde{\Omega},\Omega}(\tilde{\mu},\mu)$ *is defined by*

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

In [22, Section 4 and Appendix A] a conservation law for this nonlinear surface layer integral was derived, and it was used in order to rewrite the dynamics with a norm-preserving linear operator on Fock spaces. This nonlinear surface layer integral was also our starting point when searching for the right definition of the total mass. Indeed, the total mass in (1.11) can be written in the short form

$$(2.31) \qquad \mathfrak{M}(\tilde{\mu},\mu) := \lim_{\Omega\nearrow N}\,\lim_{\tilde{\Omega}\nearrow\tilde{N}} \left(-\mathfrak{s}\Big(\tilde{\mu}(\tilde{\Omega}) - \mu(\Omega)\Big) + \gamma^{\tilde{\Omega},\Omega}(\tilde{\mu},\mu) \right).$$

2.2. Causal fermion systems and the causal action principle

We now recall the basic definitions of a causal fermion system and the causal action principle. The connection to causal variational principles will be made in the static setting in Section 3.3.

Definition 2.9. *(causal fermion system)* Given a separable complex Hilbert space $\mathcal{H}$ with scalar product $\langle.|.\rangle_{\mathcal{H}}$ and a parameter $n \in \mathbb{N}$ (the *"spin dimension"*), we let $\mathcal{F} \subset \mathrm{L}(\mathcal{H})$ be the set of all symmetric[1] operators on $\mathcal{H}$ of finite rank, which (counting multiplicities) have at most n positive and at most n negative eigenvalues. On $\mathcal{F}$ we are given a positive measure ρ (defined on a σ-algebra of subsets of $\mathcal{F}$), the so-called *universal measure*. We refer to $(\mathcal{H},\mathcal{F},\rho)$ as a *causal fermion system*.

A causal fermion system describes a spacetime together with all structures and objects therein. In order to single out the physically admissible causal fermion systems, one must formulate physical equations. To this end, we impose that the universal measure should be a minimizer of the causal action principle, which we now introduce. For any $x,y \in \mathcal{F}$, the product xy is an operator of rank at most $2n$. However, in general it is no longer a

---

[1] Here by a symmetric operator A we mean that $\langle Au|v\rangle_{\mathcal{H}} = \langle u|Av\rangle_{\mathcal{H}}$ for all $u,v \in \mathcal{H}$. Representing the operator in an orthonormal basis, the resulting matrix is Hermitian. For bounded operators as considered here, the notions "symmetric" and "self-adjoint" coincide.

symmetric operator because $(xy)^* = yx$, and this is different from xy unless x and y commute. As a consequence, the eigenvalues of the operator xy are in general complex. We denote these eigenvalues counting algebraic multiplicities by $\lambda_1^{xy}, \ldots, \lambda_{2n}^{xy} \in \mathbb{C}$ (more specifically, denoting the rank of xy by $k \leq 2n$, we choose $\lambda_1^{xy}, \ldots, \lambda_k^{xy}$ as all the non-zero eigenvalues and set $\lambda_{k+1}^{xy}, \ldots, \lambda_{2n}^{xy} = 0$). We introduce the Lagrangian and the causal action by

(2.32)   *Lagrangian:* $$\mathcal{L}(x, y) = \frac{1}{4n} \sum_{i,j=1}^{2n} \left(\left| \lambda_i^{xy} \right| - \left| \lambda_j^{xy} \right| \right)^2$$

(2.33)   *causal action:* $$\mathcal{S}(\rho) = \iint_{\mathcal{F} \times \mathcal{F}} \mathcal{L}(x, y) \, d\rho(x) \, d\rho(y) \,.$$

The *causal action principle* is to minimize $\mathcal{S}$ by varying the measure ρ under the following constraints:

(2.34)   *volume constraint:* $$\rho(\mathcal{F}) = \text{const}$$

(2.35)   *trace constraint:* $$\int_{\mathcal{F}} \text{tr}(x) \, d\rho(x) = \text{const}$$

(2.36)   *boundedness constraint:* $$\iint_{\mathcal{F} \times \mathcal{F}} |xy|^2 \, d\rho(x) \, d\rho(y) \leq C \,,$$

where C is a given parameter, tr denotes the trace of a linear operator on $\mathcal{H}$, and the absolute value of xy is the so-called spectral weight,

$$|xy| := \sum_{j=1}^{2n} \left| \lambda_j^{xy} \right| \,.$$

This variational principle is mathematically well-posed if $\mathcal{H}$ is finite-dimensional. For the existence theory and the analysis of general properties of minimizing measures we refer to [5, 15, 16]. In the existence theory one varies in the class of regular Borel measures (with respect to the topology on $L(\mathcal{H})$ induced by the operator norm), and the minimizing measure is again in this class. With this in mind, here we always assume that

(2.37) ρ is a regular Borel measure .

Let ρ be a *minimizing* measure. *Spacetime* is defined as the support of this measure,

$$M := \text{supp} \, \rho \,.$$

Thus the spacetime points are symmetric linear operators on $\mathcal{H}$. These operators contain a lot of additional information which, if interpreted correctly, gives rise to spacetime structures like causal and metric structures, spinors and interacting fields. We refer the interested reader to [18, Chapter 1].

The only results on the structure of minimizing measures which will be needed in what follows concern the treatment of the trace constraint and the boundedness constraint. As a consequence of the trace constraint, for any minimizing measure ρ the local trace is constant in spacetime, i.e. there is a real constant $c \neq 0$ such that (see [5, Theorem 1.3] or [18, Proposition 1.4.1])

$$(2.38) \qquad\qquad \operatorname{tr} x = c \qquad \text{for all } x \in M .$$

Restricting attention to operators with fixed trace, the trace constraint (2.35) is equivalent to the volume constraint (2.34) and may be disregarded. The boundedness constraint, on the other hand, can be treated with a Lagrange multiplier. More precisely, in [5, Theorem 1.3] it is shown that for every minimizing measure ρ, there is a Lagrange multiplier $\kappa > 0$ such that ρ is a critical point of the causal action with the Lagrangian replaced by

$$\mathcal{L}_\kappa(x, y) := \mathcal{L}(x, y) + \kappa \, |xy|^2 ,$$

leaving out the boundedness constraint.

2.3. Constructing causal fermion systems in static Lorentzian spacetimes

2.3.1. Construction in globally hyperbolic spacetimes. We now recall how, starting from a globally Lorentzian spacetime, one can construct corresponding causal fermion systems. The general method is to choose $\mathcal{H}$ as a subspace of the solution space of the Dirac equation, as we now explain (for more details see [19]). Let $(\mathcal{M}, g)$ be a smooth, globally hyperbolic Lorentzian manifold of dimension $k \geq 2$. For the signature of the metric we use the convention $(+, -, \ldots, -)$. As proven in [4], $\mathcal{M}$ admits a smooth foliation $(\mathcal{N}_t)_{t \in \mathbb{R}}$ by Cauchy hypersurfaces. Thus $\mathcal{M}$ is topologically the product of $\mathbb{R}$ with a $k - 1$-dimensional manifold. In the case $k = 4$ of a four-dimensional spacetime, this implies that $\mathcal{M}$ is spin (for details see [3, 41]). For a general spacetime dimension we need to impose that $\mathcal{M}$ is spin. We let $S\mathcal{M}$ be the spinor bundle on $\mathcal{M}$ and denote the smooth sections of the spinor bundle by $C^\infty(\mathcal{M}, S\mathcal{M})$. Similarly, $C_0^\infty(\mathcal{M}, S\mathcal{M})$ denotes the smooth sections with compact support. The sections of the spinor bundle are also

referred to as wave functions. The fibres $S_x\mathcal{M}$ are endowed with an inner product of signature (n, n) with $n = 2^{[k/2]-1}$ (where $[\cdot]$ is the Gauß bracket; for details see again [3, 41]), which we denote by $\prec.|.\succ_x$. The Lorentzian metric induces a Levi-Civita connection and a spin connection, which we both denote by ∇. Every vector of the tangent space acts on the corresponding spinor space by Clifford multiplication. Clifford multiplication is related to the Lorentzian metric via the anti-commutation relations. Denoting the mapping from the tangent space to the linear operators on the spinor space by γ, we thus have

$$\gamma \,:\, T_x\mathcal{M} \to \mathrm{L}(S_x\mathcal{M}) \qquad \text{with} \qquad \gamma(u)\,\gamma(v) + \gamma(v)\,\gamma(u) = 2\,g(u, v)\,\mathbb{1}_{S_x(\mathcal{M})}\,.$$

We also write Clifford multiplication in components with the Dirac matrices γ^j. The connections, inner products and Clifford multiplication satisfy Leibniz rules and compatibility conditions; we refer to [3, 41] for details.

Combining the spin connection with Clifford multiplication gives the geometric Dirac operator denoted by

$$\mathcal{D} := i\gamma^j\nabla_j \,:\, C^\infty(\mathcal{M}, S\mathcal{M}) \to C^\infty(\mathcal{M}, S\mathcal{M})\,.$$

Given a real parameter m (the "rest mass"), the Dirac equation reads

$$(2.39) \qquad\qquad\qquad (\mathcal{D} - m)\,\psi = 0\,.$$

We mainly consider solutions in the class $C^\infty_{\mathrm{sc}}(\mathcal{M}, S\mathcal{M})$ of smooth sections with spatially compact support. On such solutions, one has the scalar product

$$(2.40) \qquad\qquad (\psi|\phi)_m = 2\pi \int_{\mathcal{N}} \prec\psi \,|\, \gamma(\nu)\,\phi\succ_x d\mu_{\mathcal{N}}(x)\,,$$

where $\mathcal{N}$ denotes any Cauchy surface and ν its future-directed normal (due to current conservation, the scalar product is in fact independent of the choice of $\mathcal{N}$; for details see [32, Section 2]). Forming the completion gives the Hilbert space $(\mathcal{H}_m, (.|.)_m)$.

Next, we choose a closed subspace $\mathcal{H} \subset \mathcal{H}_m$ of the solution space of the Dirac equation. The induced scalar product on $\mathcal{H}$ is denoted by $\langle.|.\rangle_{\mathcal{H}}$. There is the technical difficulty that the wave functions in $\mathcal{H}$ are in general not continuous, making it impossible to evaluate them pointwise. For this reason, we need to introduce an *ultraviolet regularization* on the length scale ε,

described mathematically by a linear

(2.41)   *regularization operator* $\mathfrak{R}_\varepsilon \;:\; \mathcal{H}_m \to C^0(\mathcal{M}, S\mathcal{M})\,.$

In the simplest case, the regularization can be realized by a convolution on a Cauchy surface or in spacetime (for details see [32, Section 4] or [18, Section §1.1.2]). For us, the regularization is not just a technical tool, but it realizes the concept that we want to change the geometric structures on the microscopic scale. With this in mind, we always consider the regularized quantities as those having mathematical and physical significance. Different choices of regularization operators realize different microscopic spacetime structures.

Given $\mathfrak{R}_\varepsilon$, for any spacetime point $x \in \mathcal{M}$ we consider the sesquilinear form

$$b_x \;:\; \mathcal{H} \times \mathcal{H} \to \mathbb{C}\,, \quad b_x(\psi, \phi) = -\!\prec\!(\mathfrak{R}_\varepsilon \psi)(x) | (\mathfrak{R}_\varepsilon \phi)(x)\!\succ_x\,.$$

This sesquilinear form is well-defined and bounded because $\mathfrak{R}_\varepsilon$ maps to the continuous wave functions and because evaluation at x gives a linear operator of finite rank. Thus for any $\phi \in \mathcal{H}$, the anti-linear form $b_x(.,\phi) : \mathcal{H} \to \mathbb{C}$ is continuous. By the Fréchet-Riesz theorem, there is a unique $\chi \in \mathcal{H}$ such that $b_x(\psi, \phi) = \langle \psi | \chi \rangle_\mathcal{H}$ for all $\psi \in \mathcal{H}$. The mapping $\phi \mapsto \chi$ is linear and bounded. We thus obtain a unique bounded linear operator $F^\varepsilon(x)$ on $\mathcal{H}$ which is characterized by the relation

(2.42) $(\psi \,|\, F^\varepsilon(x)\,\phi) = -\!\prec\!(\mathfrak{R}_\varepsilon \psi)(x) | (\mathfrak{R}_\varepsilon \phi)(x)\!\succ_x$ for all $\psi, \phi \in \mathcal{H}\,.$

Taking into account that the inner product on the Dirac spinors at x has signature (n, n), the local correlation operator $F^\varepsilon(x)$ is a symmetric operator on $\mathcal{H}$ of rank at most $2n$, which (counting multiplicities) has at most n positive and at most n negative eigenvalues. Varying the spacetime point, we obtain a mapping

$$F^\varepsilon \;:\; \mathcal{M} \to \mathcal{F} \subset \mathrm{L}(\mathcal{H})\,,$$

where $\mathcal{F}$ again denotes all symmetric operators of rank at most four with at most two positive and at most two negative eigenvalues. Finally, we introduce the

universal measure $d\rho := (F^\varepsilon)_* \, d\mu_\mathcal{M}$

as the push-forward of the volume measure on $\mathcal{M}$ under the mapping F^ε (thus $\rho(\Omega) := \mu_\mathcal{M}((F^\varepsilon)^{-1}(\Omega)))$.

In this way, we obtain a measure ρ on the set $\mathcal{F} \subset \mathrm{L}(\mathcal{H})$ of linear operators on a Hilbert space $\mathcal{H}$. The basic concept is to work exclusively with these objects, but to drop all other structures (like the Lorentzian metric g, the structure of the spinor bundle $S\mathcal{M}$ and the manifold structure of $\mathcal{M}$). This leads us to the structure of a causal fermion system of spin dimension n, as defined abstractly in Definition 2.9 above.

For clarity, we close with a few comments on the underlying physical concepts. The vectors in the subspace $\mathcal{H} \subset \mathcal{H}_m$ have the interpretation as those Dirac wave functions which are realized in the physical system under consideration. Therefore, the vectors in $\mathcal{H}$ are referred to as the *physical wave functions*. If we describe for example a system of one electron, then the wave function of the electron is contained in $\mathcal{H}$. Moreover, $\mathcal{H}$ includes all the wave functions in $\mathcal{H}_-$ which form the so-called Dirac sea (for an explanation of this point see for example [17]). The name causal *fermion* system is motivated by the fact that Dirac particles are fermions. According to (2.42), the local correlation operator $F^\varepsilon(x)$ describes densities and correlations of the physical wave functions at the spacetime point x. Working exclusively with the local correlation operators and the corresponding pushforward measure ρ means in particular that the geometric structures are encoded in and must be retrieved from the physical wave functions. Since the physical wave functions describe the distribution of matter in spacetime, one can summarize this concept by saying that *matter encodes geometry*.

2.3.2. Construction in static spacetimes.   By a *static* spacetime we here mean a globally hyperbolic spacetime where the foliation $(\mathcal{N}_t)_{t\in\mathbb{R}}$ can be chosen such that the timelike vector field ∂_t is a Killing field which is orthogonal to the hypersurfaces $\mathcal{N}_t$. A typical example is the Schwarzschild geometry. Since all Cauchy surfaces $\mathcal{N}_t$ are isometric, we may restrict attention to one of them and omit the subscript t. In a static spacetime, it is most convenient to write the Dirac equation (2.39) in the Hamiltonian form

$$(2.43) \qquad\qquad\qquad i\partial_t \psi = H\psi \,,$$

where H, the *Dirac Hamiltonian*, is an elliptic operator acting on the spatial sections $\Gamma(\mathcal{N}, S\mathcal{M})$. For convenience, we choose its domain as the smooth, compactly supported spinors,

$$\mathcal{D}(H) = C_0^\infty(\mathcal{N}, S\mathcal{M}) \,.$$

Identifying the initial data on $\mathcal{N}$ with the corresponding solution of the Cauchy problem, this domain is a subspace of the Hilbert space $\mathcal{H}_m$ introduced after (2.40). Moreover, the scalar product (2.40) can be expressed in terms of the functions on $\mathcal{N}$ by

$$(2.44) \qquad (\psi|\phi)_{\mathcal{N}} = 2\pi \int_{\mathcal{N}} \prec\!\psi \,|\, \gamma(\nu)\, \phi\!\succ_x d\mu_{\mathcal{N}}(x) \,.$$

As a consequence of current conservation, the Dirac Hamiltonian H with domain $\mathcal{D}(H)$ is a symmetric operator on $\mathcal{H}_m$. Using finite propagation speed, the general method by Chernoff [7] yields that the Dirac Hamiltonian is essentially self-adjoint (for details in the more general case with boundary conditions see [33]). We denote the unique self-adjoint extension again by H. The spectral theorem gives a decomposition

$$(2.45) \qquad H = \int_{\sigma(H)} \omega \, dE_\omega \,,$$

where E is a projection-valued measure. In simple terms, the spectral parameter $\omega \in \mathbb{R}$ is the frequency of the usual separation with a plane wave ansatz,

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

Consequently, the solution space $\mathcal{H}_m$ splits into the direct sum of the solutions of positive and negative frequency,

$$\mathcal{H}_m = \mathcal{H}_+ \oplus \mathcal{H}_- \qquad \text{with} \qquad \mathcal{H}_+ := E_{[0,\infty)}(\mathcal{H}_m), \ \mathcal{H}_- := E_{(-\infty,0)}(\mathcal{H}_m) \,.$$

In this paper we always choose the subspace $\mathcal{H}$ used in the above construction of the causal fermion system as a subspace which differs from $\mathcal{H}_-$ by a finite-dimensional subspace. More precisely, we assume that there are finite-dimensional subspaces $\mathcal{H}_p \subset \mathcal{H}_+$ and $\mathcal{H}_{ap} \subset \mathcal{H}_- \cap \mathcal{H}^\perp$ such that

$$(2.46) \qquad \mathcal{H} \oplus \mathcal{H}_{ap} = \mathcal{H}_- \oplus \mathcal{H}_p \,.$$

We refer to the resulting causal fermion systems as *static Dirac systems*. We remark that the choice (2.46) means that we consider a system involving a finite number of particles and anti-particles (whose states span $\mathcal{H}_p$ and $\mathcal{H}_a$, respectively).

   We finally point out that in general, the static causal fermion systems obtained in this way are *not* minimizers of the causal action. But, as worked out in detail in [18], they are critical points of the causal action in a limiting

case where $\varepsilon \to 0$, referred to as the *continuum limit*, provided that the classical field equations (Maxwell, Yang-Mills and Einstein equations) hold. Therefore, static Dirac systems are suitable examples for understanding the connection between the total mass of static causal fermion systems and the ADM mass.

3. The causal action principle in the static case

We now specialize the setting of causal fermion systems to the static case (Section 3.1). Adapting the causal action principle to static causal fermion systems and imposing a regularity condition, we get into the setting of causal variational principles (Sections 3.2 and 3.3). We finally explain the scaling freedom for static causal fermion systems (Section 3.4).

3.1. Static causal fermion systems

In this paper, we shall restrict attention to causal fermion systems which are time independent in the following sense.

Definition 3.1. *Let $(\mathcal{U}_t)_{t\in\mathbb{R}}$ be a strongly continuous one-parameter group of unitary transformations on the Hilbert space $\mathcal{H}$ (i.e. s-$\lim_{t'\to t}\mathcal{U}_{t'} = \mathcal{U}_t$ and $\mathcal{U}_t\mathcal{U}_{t'} = \mathcal{U}_{t+t'}$). The causal fermion system $(\mathcal{H}, \mathcal{F}, \rho)$ is **static with respect to** $(\mathcal{U}_t)_{t\in\mathbb{R}}$ if it has the following properties:*

(i) *Spacetime $M := \operatorname{supp}\rho \subset \mathcal{F}$ is a topological product,*

$$M = \mathbb{R} \times N \,.$$

We write a spacetime point $x \in M$ as $x = (t, \mathbf{x})$ with $t \in \mathbb{R}$ and $\mathbf{x} \in N$.

(ii) *The one-parameter group $(\mathcal{U}_t)_{t\in\mathbb{R}}$ leaves the universal measure invariant, i.e.*

$$\rho\big(\mathcal{U}_t\,\Omega\,\mathcal{U}_t^{-1}\big) = \rho(\Omega) \qquad \text{for all ρ-measurable } \Omega \subset \mathcal{F} \,.$$

Moreover,

$$\mathcal{U}_{t'}\,(t, \mathbf{x})\,\mathcal{U}_{t'}^{-1} = (t + t', \mathbf{x}) \,.$$

Before going on, we point out that we here restrict attention to spacetimes of *infinite lifetime*. Alternatively, one could also consider static and time-periodic spacetimes, in which case M would be the topological product $S^1 \times$

N. We also remark that our definition of "static" even applies to spacetimes like the Kerr geometry which are not static but *stationary*. In fact, in the above generality without a Lorentzian metric, it is a-priori not clear how to distinguish between static and stationary spacetimes. Since in this paper, we have static spacetimes like the Schwarzschild geometry in mind, it is more appropriate and more modest to refer to our spacetimes as being static. But clearly, the same definition could also be used when extending our work to stationary spacetimes.

Given a static causal fermion system, we also consider the set of operators

$$N := \{(0, \mathbf{x})\} \subset \mathcal{F}.$$

The universal measure induces a measure μ on N defined by

$$\mu(\Omega) := \rho\big([0,1] \times \Omega\big).$$

The fact that the causal fermion system is static implies that $\rho([t_1, t_2] \times \Omega) = (t_2 - t_1)\, \mu(\Omega)$, valid for all $t_1 < t_2$. This can be expressed more conveniently as

(3.1) $$d\rho = dt\, d\mu.$$

3.2. The causal action principle in the static setting

The causal action principle can be formulated in a straightforward manner for static causal fermion systems. The only point to keep in mind is that, when considering families of measures, these measures should all be static with respect to the same group $(\mathcal{U}_t)_{t \in \mathbb{R}}$ of unitary operators (see Definition 3.1). In order to make this point clear, right from the beginning we choose a group of unitary operators $(\mathcal{U}_t)_{t \in \mathbb{R}}$ on $\mathcal{H}$. We denote the equivalence classes of $\mathcal{F}$ under the action of the one-parameter group by

$$\mathcal{F}/\mathbb{R} := \{\mathcal{U}_t\, x\, \mathcal{U}_t^{-1} \mid x \in \mathcal{F}, t \in \mathbb{R}\}.$$

We denote the elements of $\mathcal{F}/\mathbb{R}$ just as the spatial points by $\mathbf{x}$ and $\mathbf{y}$. Next, we define the following functions:

$$(3.2) \qquad \text{static Lagrangian} \quad \mathcal{L}(\mathbf{x},\mathbf{y}) := \int_{-\infty}^{\infty} L\big((0,\mathbf{x}),(t,\mathbf{y})\big)\, dt$$

$$(3.3) \quad \text{static boundedness function} \quad \mathcal{T}(\mathbf{x},\mathbf{y}) := \int_{-\infty}^{\infty} \big|(0,\mathbf{x})\,(t,\mathbf{y})\big|^2\, dt$$

$$(3.4) \qquad \text{static } \kappa\text{-Lagrangian} \quad \mathcal{L}_\kappa(\mathbf{x},\mathbf{y}) := \mathcal{L}(\mathbf{x},\mathbf{y}) + \kappa\, \mathcal{T}(\mathbf{x},\mathbf{y})\,.$$

Due to the unitary invariance of the Lagrangian, this definition does not depend on the choice of representatives. Moreover, the static Lagrangian is again symmetric because

$$\mathcal{L}(\mathbf{x},\mathbf{y}) = \int_{-\infty}^{\infty} L\big((0,\mathbf{x}),\mathcal{U}_t\,(0,\mathbf{y})\,\mathcal{U}_t^{-1}\big)\, dt = \int_{-\infty}^{\infty} L\big(\mathcal{U}_t\,(0,\mathbf{y})\,\mathcal{U}_t^{-1},\,(0,\mathbf{x})\big)\, dt$$

$$= \int_{-\infty}^{\infty} L\big(\mathcal{U}_t^{-1}\,\mathcal{U}_t\,(0,\mathbf{y})\,\mathcal{U}_t^{-1}\,\mathcal{U}_t,\,\mathcal{U}_t^{-1}\,(0,\mathbf{x})\,\mathcal{U}_t\big)\, dt$$

$$= \int_{-\infty}^{\infty} L\big((0,\mathbf{y}),\,\mathcal{U}_t^{-1}\,(0,\mathbf{x})\,\mathcal{U}_t\big)\, dt = \int_{-\infty}^{\infty} L\big((0,\mathbf{y}),\,(-t,\mathbf{x})\big)\, dt$$

$$(3.5) \qquad = \big\{t \to -t\big\} = \int_{-\infty}^{\infty} L\big((0,\mathbf{y}),(t,\mathbf{x})\big)\, dt = \mathcal{L}(\mathbf{y},\mathbf{x})\,,$$

and similarly for $\mathcal{T}(x,y)$. For a measure ρ which is static with respect to $(\mathcal{U}_t)_{t\in\mathbb{R}}$, we introduce the

$$(3.6) \quad \text{static causal action:} \quad \mathcal{S}(\rho) = \int_{\mathcal{F}/\mathbb{R}} d\mu(\mathbf{x}) \int_{\mathcal{F}/\mathbb{R}} d\mu(\mathbf{y})\, \mathcal{L}_\kappa(\mathbf{x},\mathbf{y})\,.$$

The *static causal action principle* is to minimize $\mathcal{S}$ by varying the measure ρ with in the class of measures which are static with respect to $(\mathcal{U}_t)_{t\in\mathbb{R}}$ under the following constraints:

$$(3.7) \qquad \text{volume constraint:} \qquad \mu(\mathcal{F}/\mathbb{R}) = \text{const}$$

$$(3.8) \qquad \text{trace constraint:} \qquad \int_{\mathcal{F}/\mathbb{R}} \text{tr}(\mathbf{x})\, d\mu(\mathbf{x}) = \text{const}\,.$$

Note that the boundedness constraint is taken into account by the Lagrange multiplier term $\kappa\,\mathcal{T}(x,y)$ in (3.4).

3.3. The regular setting as a causal variational principle

We now explain how to get to the setting of causal variational principles introduced in Section 2.1. The main differences between the static causal action principle and the setting of causal variational principles is that the set $\mathcal{F}/\mathbb{R}$ does not need to be a manifold and that there is the additional trace constraint (3.8). As explained after (2.38), the trace constraint can be treated by restricting attention to operators of fixed trace. In order to give the set of operators a manifold structure, we assume that ρ is *regular* in the sense that all operators in its support have exactly n positive and exactly n negative eigenvalues. This leads us to introduce the set $\mathcal{F}^{\mathrm{reg}}$ as the set of all operators F on $\mathcal{H}$ with the following properties:

(i) F is symmetric, has finite rank and (counting multiplicities) has exactly n positive and n negative eigenvalues.

(ii) The trace is constant, i.e. $\mathrm{tr}(F) = c > 0$.

If $\mathcal{H}$ is finite-dimensional, the set $\mathcal{F}^{\mathrm{reg}}$ has a smooth manifold structure (see the concept of a flag manifold in [37] or the detailed construction in [24, Section 3]). Assuming that the action of the group $(\mathcal{U}_t)_{t \in \mathbb{R}}$ on $\mathcal{G}$ is proper and has no fixed points, the quotient is again a manifold. Thus setting

$$\mathcal{G} = \mathcal{F}^{\mathrm{reg}}/\mathbb{R} \, ,$$

we get into the setting of causal variational principles as introduced in Section 2.1. We now verify that also the technical assumptions are satisfied.

Proposition 3.2. *Assume that $\mathcal{H}$ is finite-dimensional and that the action of the group $(\mathcal{U}_t)_{t \in \mathbb{R}}$ on $\mathcal{G}$ is proper and has no fixed points. Then for every minimizing measure, the conditions* (i)–(iv) *in Section 2.1 on page 1749 are satisfied for the static κ-Lagrangian.*

Proof. The symmetry property (i) was already verified for the static Lagrangian (3.4) in (3.5). Since the eigenvalues of a matrix depend continuously on the matrix entries (see for example [39, §II.1]), the Lagrangian $\mathcal{L}_\kappa(x, y)$ is continuous in both arguments. However, this does not mean that the static

Lagrangian (3.4) is also continuous. But Fatou's lemma yields

$$\mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) = \int_{-\infty}^{\infty} \mathcal{L}_\kappa\big((0, \mathbf{x}), (t, \mathbf{y})\big)\, dt = \int_{-\infty}^{\infty} \lim_{\mathbf{x}' \to \mathbf{x}} \mathcal{L}_\kappa\big((0, \mathbf{x}'), (t, \mathbf{y})\big)\, dt$$

$$\leq \liminf_{\mathbf{x}' \to \mathbf{x}} \int_{-\infty}^{\infty} \mathcal{L}_\kappa\big((0, \mathbf{x}'), (t, \mathbf{y})\big)\, dt = \liminf_{\mathbf{x}' \to \mathbf{x}} \mathcal{L}_\kappa(\mathbf{x}', \mathbf{y})\,,$$

showing that the static κ-Lagrangian is lower semi-continuous. This proves (ii).

In order to prove (iii) and (iv), we make use of the EL equations (2.3), which we write as

$$(3.9) \qquad \int_N \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y}) \leq \mathfrak{s} \qquad \text{for all } x \in \mathcal{G}\,.$$

The regularity of μ follows immediately from the regularity of ρ in (2.37). Using that the trace of x is non-zero, the Lagrangian $\mathcal{L}_\kappa(x, x)$ is strictly positive on the diagonal (for details see [15, Proposition 4.3]). By continuity of the Lagrangian, also the static Lagrangian is strictly positive on the diagonal, $\mathcal{L}_\kappa(\mathbf{x}, \mathbf{x}) > 0$. By lower semi-continuity of the static Lagrangian, there is $\varepsilon > 0$ and an open neighborhood U of x where $\mathcal{L}_\kappa(\mathbf{x}, .)$ is larger than ε. Combining this fact with the positivity of the Lagrangian, the inequality (3.9) gives rise to the estimate

$$\mathfrak{s} \geq \int_U \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y}) \geq \varepsilon\, \mu(U)\,,$$

showing that μ is locally finite. This proves (iii). Next, it is obvious from (3.9) that $\mathcal{L}_\kappa(\mathbf{x}, .)$ is μ-integrable and that the resulting function is bounded. It remains to prove that this function is lower semi-continuous. To this end, we can argue similar as in the proof of (ii) above:

$$\int_{\mathcal{G}} \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y}) \leq \int_{\mathcal{G}} \lim_{\mathbf{x}' \to \mathbf{x}} \mathcal{L}_\kappa(\mathbf{x}', \mathbf{y})\, d\mu(\mathbf{y}) \leq \lim_{\mathbf{x}' \to \mathbf{x}} \int_{\mathcal{G}} \mathcal{L}_\kappa(\mathbf{x}', \mathbf{y})\, d\mu(\mathbf{y})$$

where in the first step we used that the static κ-Lagrangian is lower semi-continuous and applied again Fatou's lemma. This concludes the proof of (iv). □

In the infinite-dimensional setting, the set $\mathcal{F}^{\text{reg}}$ is an infinite-dimensional Banach manifold (for details see [31]). For technical simplicity, here we shall not enter the details of the infinite-dimensional analysis. Instead, our method is to restrict attention to a finite-dimensional submanifold of $\mathcal{F}^{\text{reg}}$. Clearly,

this submanifold must contain the supports of both measures ρ and $\tilde{\rho}$, and the unitary group $(\mathcal{U}_t)_{t\in\mathbb{R}}$ must map the submanifold to itself. Moreover, the vector fields of the jets needed for the analysis must all be tangential to this submanifold. Restricting the Lagrangian to this submanifold, all the results of Proposition 3.2 must hold. Then we simply choose $\mathcal{G}$ as the equivalence classes of this submanifold under the action of the group $(\mathcal{U}_t)_{t\in\mathbb{R}}$. This procedure will be illustrated in Section 6 in the example of static Dirac systems in asymptotically Schwarzschild spacetimes.

3.4. Freedom in rescaling solutions of the Euler-Lagrange equations

Let ρ be a critical measure of the static causal action principle for given values of the parameters c and κ. Then for a suitable Lagrange multiplier $\mathfrak{s} > 0$, the equations (2.38) and (2.3) hold. For clarity, we add a subscript κ to ℓ and write (2.2) as

$$\ell_\kappa(\mathbf{x}) := \ell(\mathbf{x}) + \kappa \, \mathfrak{t}(\mathbf{x}) \,,$$

where

$$(3.10) \qquad \ell(\mathbf{x}) := \int_M \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) \, d\rho(\mathbf{y}) - \mathfrak{s}$$

$$(3.11) \qquad \mathfrak{t}(\mathbf{x}) := \int_M \mathcal{T}(\mathbf{x}, \mathbf{y}) \, d\rho(\mathbf{y}) \,.$$

There is a two-parameter family of rescalings which again give critical measures. Indeed, the new measure $\hat{\mu}$ defined by

$$(3.12) \qquad \hat{\mu}(\Omega) = \sigma \, \mu\!\left(\frac{\Omega}{\lambda}\right) \qquad \text{with } \lambda, \sigma > 0$$

again satisfies the EL equation with new Lagrange multipliers

$$\hat{c} = \lambda \, c \quad \text{and} \quad \hat{\mathfrak{s}} = \sigma \, \lambda^4 \, \mathfrak{s} \,.$$

This rescaling freedom could be fixed for example by imposing that

$$c = \mathfrak{s} = 1 \,.$$

Note that the Lagrange multiplier κ remains unchanged; it is a dimensionless parameter which characterizes the solution independent of the values of c and $\mathfrak{s}$.

4. General properties of the total mass

In this section we work out a few general properties of the total mass as introduced in Definition 1.2.

4.1. Independence of the exhaustion

We first verify that the total mass does not depend on the choice of the exhaustions of N and $\tilde{N}$.

Proposition 4.1. *Assume that μ and $\tilde{\mu}$ are asymptotically close (see Definition 1.1). Then the limits $\Omega \nearrow N$ and $\tilde{\Omega} \nearrow \tilde{N}$ in (1.11) exist and are independent of the exhaustions. The total mass is finite. It can also be written as the difference of the spatial integrals*

$$(4.1) \qquad \mathfrak{M}(\tilde{\mu}, \mu) = \int_{\tilde{N}} \big(\tilde{n}(\mathbf{x}) - \mathfrak{s}\big) \, d\tilde{\mu}(\mathbf{x}) - \int_{N} \big(n(\mathbf{x}) - \mathfrak{s}\big) \, d\mu(\mathbf{x}) \, .$$

Proof. Using that the Lagrangian is symmetric, the integral expression in Definition 1.2 can be rewritten with the help of the correlation measures (1.9) as

$$\int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N \setminus \Omega} d\mu(\mathbf{y}) \, \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y}) - \int_{\Omega} d\mu(\mathbf{x}) \int_{\tilde{N} \setminus \tilde{\Omega}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y})$$

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

$$= \tilde{\nu}\big(\tilde{\Omega}\big) - \nu(\Omega) \, ,$$

where in the last step we used the definition of the correlation measures (1.9). We thus obtain the compact formula for the total mass

$$\mathfrak{M}(\tilde{\mu}, \mu) = \lim_{\Omega \nearrow N} \lim_{\tilde{\Omega} \nearrow \tilde{N}} \Big(- \mathfrak{s}\big(\tilde{\mu}(\tilde{\Omega}) - \mu(\Omega)\big) + \tilde{\nu}\big(\tilde{\Omega}\big) - \nu(\Omega) \Big)$$

$$(4.2) \qquad = \lim_{\Omega \nearrow N} \lim_{\tilde{\Omega} \nearrow \tilde{N}} \left(\int_{\tilde{\Omega}} \big(\tilde{n}(\mathbf{x}) - \mathfrak{s}\big) \, d\tilde{\mu}(\mathbf{x}) - \int_{\Omega} \big(n(\mathbf{x}) - \mathfrak{s}\big) \, d\mu(\mathbf{x}) \right)$$

with n and $\tilde{n}$ as defined in (1.8). Since μ and $\tilde{\mu}$ are asymptotically close, the integrands in (4.2) are in L^1. Therefore, the limits $\Omega \nearrow N$ and $\tilde{\Omega} \nearrow \tilde{N}$ exist by Lebesgue's dominated convergence theorem. We thus obtain (4.1). $\square$

We next recall the definition of non-atomic measures (see for example [36, Section 40]).

Definition 4.2. *A Borel set $\Omega \subset \mathcal{M}$ is called an* **atom** *of the Borel measure μ if $\mu(\Omega) > 0$ and if every Borel subset $K \subset \Omega$ with $\mu(K) < \mu(\Omega)$ has measure zero. A Borel measure is said to be* **non-atomic** *if it has no atoms.*

Proposition 4.3. *If μ or $\tilde{\mu}$ is non-atomic, then the total mass (see Definition 1.2) can be written equivalently in the form* (1.12).

Proof. Clearly, for every exhaustions $(\Omega_n)_{n \in \mathbb{N}}$ of N and $(\tilde{\Omega}_n)_{N \in \mathbb{N}}$ of $\tilde{N}$ which satisfy the condition $\mu(\Omega_n) = \tilde{\mu}(\tilde{\Omega}_n) < \infty$ for all $n \in \mathbb{N}$, the formula (1.11) reduces to (1.12). In view of the independence of the choice of exhaustions (Proposition 4.1), it remains to show that there are exhaustions $(\Omega_n)_{n \in \mathbb{N}}$ and $(\tilde{\Omega}_n)_{n \in \mathbb{N}}$ with $\mu(\Omega_n) = \tilde{\mu}(\tilde{\Omega}_n) < \infty$ for all n.

To this end, assume for example that μ is non-atomic. Let $(U_n)_{n \in \mathbb{N}}$ and $(\tilde{U}_n)_{n \in \mathbb{N}}$ be exhaustions of N and $\tilde{N}$, respectively, by sets of finite volume. In view of (1.10), the volumes of these sets tends to infinity. Therefore, we can choose subsequences (for simplicity we again denoted by $(U_n)_{n \in \mathbb{N}}$ and $(\tilde{U}_n)_{n \in \mathbb{N}}$) such that

$$\mu(U_1) \leq \tilde{\mu}(\tilde{U}_1) \leq \mu(U_2) \leq \tilde{\mu}(\tilde{U}_2) \leq \cdots .$$

Since μ is a non-atomic and regular Borel measure, we can find measurable sets V_n with

$$U_n \subset V_n \subset U_{n+1} \qquad \text{and} \qquad \mu(V_n) = \tilde{\mu}(\tilde{U}_n) \,.$$

Clearly, the sequence $(V_n)_{n \in \mathbb{N}}$ is again an exhaustion of N by sets of finite measure. Therefore, the sets $\Omega_n := V_n$ and $\tilde{\Omega}_n := \tilde{U}_n$ have all the required properties. $\qquad \square$

4.2. Independence of the inner geometry

We now want to verify and make precise that the total mass in Definition 1.2 depends only on the geometry near infinity and on the inner volume, but not on the inner geometry. To this end, we shall introduce a more general notion of total mass which only involves the geometry outside an arbitrarily chosen compact set (see Definition 4.4). Then we shall prove that this notion of mass coincides with the total mass of Definition 1.2.

We describe the "inner regions" of our spacetimes by relatively compact open subsets $I \subset M$ and $\tilde{I} \subset \tilde{M}$. In order to disregard the geometry of these

subsets, we modify the surface layer integral of Definition 2.8 to

$$\gamma_{\tilde{I},I}^{\tilde{\Omega},\Omega}(\tilde{\mu},\mu) := \int_{\tilde{\Omega}} d\tilde{\mu}(\mathbf{x}) \int_{N\setminus(I\cup\Omega)} d\mu(\mathbf{y})\, \mathcal{L}(\mathbf{x},\mathbf{y}) - \int_{\Omega} d\mu(\mathbf{x}) \int_{\tilde{N}\setminus(\tilde{I}\cup\tilde{\Omega})} d\tilde{\mu}(\mathbf{y})\, \mathcal{L}(\mathbf{x},\mathbf{y})\,,$$

where Ω and $\tilde{\Omega}$ are now subsets of $N\setminus I$ and $\tilde{N}\setminus\tilde{I}$, respectively.

Definition 4.4. *The **total mass** $\mathfrak{M}$ on $\tilde{N}\setminus\tilde{I}$ relative to $N\setminus I$ is defined in generalization of* (2.31) *by*

$$(4.3)\quad \mathfrak{M}_{\tilde{I},I}(\tilde{\mu},\mu) := \lim_{\Omega\nearrow N\setminus I}\ \lim_{\tilde{\Omega}\nearrow \tilde{N}\setminus\tilde{I}} \left(-\mathfrak{s}\Big(\tilde{\mu}(\tilde{\Omega}) - \mu(\Omega)\Big) + \gamma_{\tilde{I},I}^{\tilde{\Omega},\Omega}(\tilde{\mu},\mu) \right),$$

where the sets Ω and $\tilde{\Omega}$ form exhaustions of $N\setminus I$ and $\tilde{N}\setminus\tilde{I}$ by sets of finite volume, respectively.

This version of total mass can again be rewritten in terms of correlation measures. Indeed, introducing the *correlation measures* $\nu_{\tilde{I}}$ on $N\setminus I$ and $\tilde{\nu}_I$ on $\tilde{N}\setminus\tilde{I}$ by

$$d\nu_{\tilde{I}}(\mathbf{x}) = n_{\tilde{I}}(\mathbf{x})\, d\mu(\mathbf{x}) \qquad \text{and} \qquad d\tilde{\nu}_I(\mathbf{x}) = \tilde{n}_I(\mathbf{x})\, d\tilde{\mu}(\mathbf{x})$$

with functions

$$(4.4)\quad \begin{cases} n_{\tilde{I}}\ :\ N \to \mathbb{R}_0^+ \cup \{\infty\}\,, \qquad n_{\tilde{I}}(\mathbf{x}) = \displaystyle\int_{\tilde{N}\setminus\tilde{I}} \mathcal{L}(\mathbf{x},\mathbf{y})\, d\tilde{\mu}(\mathbf{y}) \\[2ex] \tilde{n}_I\ :\ \tilde{N} \to \mathbb{R}_0^+ \cup \{\infty\}\,, \qquad \tilde{n}_I(\mathbf{x}) = \displaystyle\int_{N\setminus I} \mathcal{L}(\mathbf{x},\mathbf{y})\, d\mu(\mathbf{y})\,, \end{cases}$$

a short computation yields

$$\gamma_{\tilde{I},I}^{\tilde{\Omega},\Omega}(\tilde{\mu},\mu) = \tilde{\nu}_I(\tilde{\Omega}) - \nu_{\tilde{I}}(\Omega)\,.$$

We now generalize the result of Proposition 4.1:

Proposition 4.5. *Assume that μ and $\tilde{\mu}$ are asymptotically close. Then the limits $\Omega\nearrow N\setminus I$ and $\tilde{\Omega}\nearrow \tilde{N}\setminus\tilde{I}$ in* (4.3) *exist and are independent of the exhaustions. The total mass on $\tilde{N}\setminus\tilde{I}$ relative to $N\setminus I$ is finite. It can be written as*

$$(4.5)\quad \mathfrak{M}_{\tilde{I},I}(\tilde{\mu},\mu) = \int_{\tilde{N}\setminus\tilde{I}} \big(\tilde{n}_I(\mathbf{x}) - \mathfrak{s}\big)\, d\tilde{\mu}(\mathbf{x}) - \int_{N\setminus I} \big(n_{\tilde{I}}(\mathbf{x}) - \mathfrak{s}\big)\, d\mu(\mathbf{x})\,.$$

Proof. Following the proof of Proposition 4.1, it suffices to show that the functions $n_{\tilde{I}} - \mathfrak{s}$ and $\tilde{n}_I - \mathfrak{s}$ are integrable. We only consider the latter function, because the proof for the first function is analogous. Thus our task is to show that

$$\int_{N \setminus I} \left| n_{\tilde{I}} - \mathfrak{s} \right| d\mu < \infty .$$

Comparing (4.4) with (1.8), we can rewrite the function in the integrand as

$$(4.6) \qquad n_{\tilde{I}}(\mathbf{x}) - \mathfrak{s} = \left[n(\mathbf{x}) - s \right] - \int_{\tilde{I}} \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\tilde{\mu}(\mathbf{y}) .$$

The square bracket is integrable because μ and $\tilde{\mu}$ are asymptotically close (see Definition 1.1). Integrating the last summand, the fact that the Lagrangian is non-negative makes it possible to apply Tonelli's theorem,

$$\int_N \left| \int_{\tilde{I}} \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\tilde{\mu}(\mathbf{y}) \right| d\mu(\mathbf{x}) = \int_{\tilde{I}} \left(\int_N \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{x}) \right) d\tilde{\mu}(\mathbf{y}) \in \mathbb{R}_0^+ \cup \{\infty\} .$$

According to condition (iv) in Section 2.1 on page 1750, the $\mathbf{x}$-integration gives a lower semi-continuous and bounded function on $\mathcal{G}$. Since $\tilde{I}$ is compact and $\tilde{\mu}$ is locally finite, the $\mathbf{y}$-integral exists and is finite. We conclude that the last summand in (4.6) is integrable over N, concluding the proof. $\qquad \square$

We can now state and prove the main result of this section.

Theorem 4.6. *Assume that μ and $\tilde{\mu}$ are asymptotically close. Then the total mass depends only on the volumes of the relatively compact open subsets $I \subset N$ and $\tilde{I} \subset \tilde{N}$. More precisely,*

$$\mathfrak{M}_{\tilde{I}, I}(\tilde{\mu}, \mu) = \mathfrak{M}(\tilde{\mu}, \mu) + \mathfrak{s} \left(\tilde{\mu}(\tilde{I}) - \mu(I) \right) .$$

Proof. We rewrite (4.5) as follows,

$$\mathfrak{M}_{\tilde{I}, I}(\tilde{\mu}, \mu) = \int_{\tilde{N} \setminus \tilde{I}} \left(\tilde{n}(\mathbf{x}) - \mathfrak{s} \right) d\tilde{\mu}(\mathbf{x}) - \int_{N \setminus I} \left(n(\mathbf{x}) - \mathfrak{s} \right) d\mu(\mathbf{x})$$

$$- \int_{\tilde{N} \setminus \tilde{I}} d\tilde{\mu}(\mathbf{x}) \int_I d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) + \int_{N \setminus I} d\mu(\mathbf{x}) \int_{\tilde{I}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y})$$

$$= \mathfrak{M}(\tilde{\mu}, \mu) - \int_{\tilde{I}} \left(\tilde{n}(\mathbf{x}) - \mathfrak{s} \right) d\tilde{\mu}(\mathbf{x}) + \int_I \left(n(\mathbf{x}) - \mathfrak{s} \right) d\mu(\mathbf{x})$$

$$- \int_{\tilde{N} \setminus \tilde{I}} d\tilde{\mu}(\mathbf{x}) \int_I d\mu(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) + \int_{N \setminus I} d\mu(\mathbf{x}) \int_{\tilde{I}} d\tilde{\mu}(\mathbf{y}) \, \mathcal{L}(\mathbf{x}, \mathbf{y}) .$$

We now transform the last line: Since the Lagrangian is symmetric and non-negative, by Tonelli's theorem we may interchange the orders of integration. Moreover, using again the symmetry of the Lagrangian, we may replace the integration range $\tilde{N} \setminus \tilde{I}$ by $\tilde{N}$ and $N \setminus I$ by N. Then the last line becomes

$$- \int_I n(\mathbf{y}) \, d\mu(\mathbf{y}) + \int_{\tilde{I}} \tilde{n}(\mathbf{y}) \, d\tilde{\mu}(\mathbf{y}) \, .$$

Collecting all the terms gives the result.   □

4.3. Independence of the identification of Hilbert spaces

In Section 1, the total mass was introduced for two measures ρ and $\tilde{\rho}$ defined on a set of linear operators $\mathcal{F}$ on a Hilbert space ρ. In most applications, however, the two spacetimes are described by two causal fermion systems $(\mathcal{H}, \mathcal{F}, \rho)$ and $(\tilde{\mathcal{H}}, \tilde{\mathcal{F}}, \tilde{\rho})$ which are defined on two different Hilbert spaces $\mathcal{H}$ and $\tilde{\mathcal{H}}$. Both spacetimes are static in the sense that ρ is static with respect to a group unitary transformations $(\mathcal{U}_t)_{t \in \mathbb{R}}$ on $\mathcal{H}$ (see Definition 3.1), whereas $\tilde{\rho}$ is static with respect to a unitary group $(\tilde{\mathcal{U}}_t)_{t \in \mathbb{R}}$ on $\tilde{\mathcal{H}}$. In order to get into the setting of the introduction, the two Hilbert spaces must be identified by a unitary transformation $V : \mathcal{H} \to \tilde{\mathcal{H}}$, in such a way that the spacetimes become jointly static, i.e.

$$(4.7) \qquad\qquad \tilde{\mathcal{U}}_t = V \, \mathcal{U}_t \, V^{-1} \qquad \text{for all } t \in \mathbb{R} \, .$$

In Section 6.2, this construction will be explained in more detail in the example of the Schwarzschild geometry. Here our point of interest is that the condition (4.7) does not determine V uniquely. Indeed, V is unique only up to the transformations $V \to VW$, where W is a unitary transformation on $\mathcal{H}$ which is *static* in the sense that it commutes with the time evolution,

$$(4.8) \qquad\qquad \mathcal{U}_t \, W = W \, \mathcal{U}_t \qquad \text{for all } t \in \mathbb{R} \, .$$

We now prove that, within the class of asymptotically flat spacetimes, the total mass does not depend on the choice of W. To this end, we let $(\mathcal{H}, \mathcal{F}, \rho)$ and $(\mathcal{H}, \mathcal{F}, \tilde{\rho})$ be two causal fermion systems which are both static with respect to a group $(\mathcal{U}_t)_{t \in \mathbb{R}}$ of unitary transformations of $\mathcal{H}$ (see Definition 3.1). Given a unitary transformation W on $\mathcal{H}$ which is static (4.8), the

unitarily transformed static measure $W\tilde{\rho}$ defined by

$$(4.9) \qquad (W\tilde{\rho})(\Omega) := \tilde{\rho}(W^{-1}\,\Omega\,W) \quad \text{for } \Omega \subset \mathcal{F}$$

is again static, making it possible to decompose it similar to (3.1) as

$$d(W\tilde{\rho}) = dt\,d(W\tilde{\mu})\,.$$

Theorem 4.7. *If both $\tilde{\mu}$ and $W\tilde{\mu}$ are asymptotically flat with respect to the vacuum measure μ (see Definition 1.4 and 1.5), then the total masses of $\tilde{\mu}$ and $W\tilde{\mu}$ coincide,*

$$\mathfrak{M}(\tilde{\mu}, \mu) = \mathfrak{M}\big(W\tilde{\mu}, \mu\big)\,.$$

The proof of this theorem is based on the unitary invariance of the causal action. In preparation of the proof, we must introduce the necessary concepts. The causal action principle is *unitarily invariant* in the following sense. Let $W \in \mathrm{U}(\mathcal{H})$ be a unitary transformation. Given a measure ρ on $\mathcal{F}$, we can unitarily transform the measure similar to (4.9) by setting $(W\rho)(\Omega) := \rho(W^{-1}\,\Omega\,W)$. Since the eigenvalues of an operator are invariant under unitary transformations, a universal measure ρ is a minimizer or critical point of the causal action principle if and only if $W\rho$ is. Next, we specialize again to the setting that all objects are static with respect to a unitary group $(\mathcal{U}_t)_{t\in\mathbb{R}}$. Then, as explained in Sections 3.1 and 3.2, we can work with the static causal action principle. If also W is static (in the sense (4.8)), the resulting static measure $W\mu$ defined by $d(W\rho) = dt\,d(W\mu)$ is a critical point of the static causal action if and only if μ is. This fact can be used to construct solutions of the linearized field equations, as we now explain. Let μ be a critical static measure. Moreover, let $(W_s)_{s\in[0,\tau_{\max}]}$ be a smooth and strongly continuous family of static unitary transformations with generator

$$\mathcal{A} := -i\frac{d}{ds}\,W_s\big|_{s=0}\,.$$

Lemma 4.8. *Assume that the jet*

$$(4.10) \qquad \mathfrak{u} := (0, u) \qquad \text{with} \qquad u(\mathbf{x}) := i\big[\mathcal{A}, \mathbf{x}\big]$$

has the property

$$(4.11) \qquad D_v u \in \mathfrak{J}^{\text{test}} \qquad \text{for all } \mathfrak{v} \in \mathfrak{J}^{\text{test}}$$

(where the directional derivatives are computed in the distinguished charts mentioned in Definition 2.1). Then the jet $\mathfrak{u}$ is a solution of the linearized

field equations, i.e.

(4.12) $\nabla_{\mathfrak{v}} \int_N \left(D_{1,u} + D_{2,u} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\mu(\mathbf{y}) = 0$ *for all* $\mathfrak{v} \in \mathfrak{J}^{\text{test}}$.

Proof. One method of proof would be to differentiate through the EL equations. However, this would involve a transformation of the space of test jets (similar as explained for example in [27, Section 3.1]). Here we prefer to show that the integrand of (4.12) vanishes identically. Indeed, due to the unitary invariance of the Lagrangian,

$$\mathcal{L}\left(\mathcal{U}_\tau \, \mathbf{x} \, \mathcal{U}_\tau^{-1}, \ \mathcal{U}_\tau \, \mathbf{y} \, \mathcal{U}_\tau^{-1} \right) = \mathcal{L}(\mathbf{x}, \mathbf{y}) \, .$$

Differentiating with respect to s gives

$$\left(D_{1,u} + D_{2,u} \right) \mathcal{L}(\mathbf{x}, \mathbf{y}) \, d\rho(\mathbf{y}) = 0 \, .$$

Hence the integrand in (4.12) vanishes for all $\mathbf{x}, \mathbf{y} \in \mathcal{F}/\mathbb{R}$. As a consequence, the integral in (4.12) vanishes for all $x \in \mathcal{F}/\mathbb{R}$. Consequently, also its derivative in the direction of u vanishes. Using our convention that the jet derivatives act only on the Lagrangian (see the end of Section 2.1), the directional derivative differs from the derivative by the term $D_{D_v u}\ell(\mathbf{x})$. This term vanishes in view of (4.11) and the weak EL equations (2.9).   □

Due to the commutator in (4.10), we refer to jets of this form as **static commutator jets**. For more details on commutator jets we refer to [23].

Proof of Theorem 4.7. Since both $\tilde{\mu}$ and $W\tilde{\mu}$ are asymptotically flat with respect to μ, their total mass can be expressed via (1.16) in terms of a vector field w near infinity. The freedom to perform static unitary transformations means for the jets that w can be changed by infinitesimal unitary transformations, which by (4.10) correspond to static commutator jets. Therefore, the freedom to perform static unitary transformations of $\tilde{\mu}$ means that in (1.16) we have the freedom to transform w according to $w \to w + u$ with u a static commutator jet. Since u is a linearized solution without scalar component (see Lemma 4.8), we can apply the conservation law of Lemma 2.3 to conclude that u does not contribute to the surface layer integral (1.16).   □

5. The positive mass theorem

The goal of this section is to prove Theorem 1.9. In preparation, we need to derive and analyze the equations of linearized gravity which were already mentioned in words in the introduction.

5.1. The equations of linearized gravity

We saw in Section 2.1.2 that, given a family $(\tilde{\mu}_\tau)_{\tau \in [0,1)}$ of solutions of the EL equations of the form (2.12) for fixed values of the parameters κ and $\mathfrak{s}$, the infinitesimal generator $\mathfrak{v}$ of this family is a solution of the linearized field equations (see (2.13) and (2.15)). Keeping the parameter $\mathfrak{s}$ fixed is a matter of convenience, because the rescaling freedom in Section 3.4 makes it possible to give this parameter an arbitrary value. However, the situation is different for the parameter κ, which is dimensionless and scaling invariant. For this reason, it is interesting to also consider families $(\tilde{\mu}_\tau)_{\tau \in [0,1)}$ where $\kappa(\tau)$ depends on τ. It is most convenient to arrange by a reparametrization of τ that

$$(5.1) \qquad \frac{d}{d\tau} \log \kappa(\tau)\big|_{\tau=0} = -1 \,.$$

Moreover, as will become clear below, it is preferable to also choose $\mathfrak{s}$ as a function of τ. Writing the EL equations (2.14) for the κ-Lagrangian in (3.4),

$$(5.2) \qquad \nabla_\mathfrak{u} \left(\int_N f_\tau(\mathbf{x}) \left(\mathcal{L}_\kappa \big(F_\tau(\mathbf{x}), F_\tau(\mathbf{y}) \big) + \kappa(\tau)\, \mathcal{T}(\mathbf{x}, \mathbf{y}) \right) f_\tau(\mathbf{y})\, d\mu(\mathbf{y}) \right.$$
$$\left. - f_\tau(\mathbf{x})\, \mathfrak{s}(\tau) \right) = 0 \,,$$

differentiating with respect to τ and evaluating at $\tau = 0$ gives in generalization of (2.15) the equation

$$\langle \mathfrak{u}, \Delta \mathfrak{v} \rangle|_N = -\kappa'(0)\, \nabla_\mathfrak{u} \mathfrak{t} + \mathfrak{s}'(0)\, \nabla_\mathfrak{u} 1 \qquad \text{for all } \mathfrak{u} \in \mathfrak{J}^{\text{test}} \,,$$

where Δ is defined by (2.16), but for the κ-Lagrangian,

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

and with $\mathfrak{t}$ as in (1.7).

By assumption, the measure μ is asymptotically flat and satisfies the EL equations. Using these properties, we know according to Definition 1.5 (ii) that the function ℓ has a limit at infinity. Moreover, the EL equations (1.4) and (1.5) imply that the same is true for the function $\mathfrak{t}$. We set

$$\mathfrak{t}_\infty := \lim_{\mathbf{x} \to \infty} \mathfrak{t}(\mathbf{x}) > 0 .$$

We apply (5.1) and choose $\mathfrak{s}'(0)$ as

$$\mathfrak{s}'(0) = -\kappa(0)\, \mathfrak{t}_\infty < 0 .$$

We thus obtain the inhomogeneous linearized field equations

$$\langle \mathfrak{u}, \Delta\mathfrak{v} \rangle|_N = \kappa\, \nabla_{\mathfrak{u}}\big(\mathfrak{t} - \mathfrak{t}_\infty\big)\big|_N \qquad \text{for all } \mathfrak{u} \in \mathfrak{J}^{\text{test}} .$$

Using the EL equations, we can rewrite the inhomogeneity in terms of ℓ,

$$\langle \mathfrak{u}, \Delta\mathfrak{v} \rangle|_N = -\nabla_{\mathfrak{u}}\big(\ell - \ell_\infty\big)\big|_N \qquad \text{for all } \mathfrak{u} \in \mathfrak{J}^{\text{test}} .$$

By adding a suitable inner solution (see Section 2.1.4) we can arrange that the jet $\mathfrak{v} = (0, v)$ has no scalar component. We refer to the resulting equation

$$(5.3) \qquad \langle \mathfrak{u}, \Delta v \rangle|_N = -\nabla_{\mathfrak{u}}\big(\ell - \ell_\infty\big)\big|_N \qquad \text{for all } \mathfrak{u} \in \mathfrak{J}^{\text{test}}$$

as the *equations of linearized gravity*. We point out that the solution v of the equations of linearized gravity will in general not be unique. For example, one can add to v any divergence-free vector field on N. But this has no effect on the following conservation law:

Proposition 5.1. *For a jet $v \in \mathfrak{J}^1$ without scalar component which satisfies the equations of linearized gravity* (5.3), *the surface layer integral* (2.18) *(again for the κ-Lagrangian) is computed for an exhaustion of N by*

$$\lim_{\Omega \nearrow N} \gamma_\mu^\Omega(v) = \int_N \big(\ell - \ell_\infty\big)\, d\mu .$$

Proof. A direct computation using (5.3) yields

$$\gamma_\mu^\Omega(v) = \int_\Omega d\mu(\mathbf{x}) \int_{M\setminus\Omega} d\mu(\mathbf{y}) \left(D_{1,v} - D_{2,v}\right)\mathcal{L}_\kappa(\mathbf{x}, \mathbf{y})$$

$$= \int_\Omega d\mu(\mathbf{x}) \int_M d\mu(\mathbf{y}) \left(D_{1,v} - D_{2,v}\right)\mathcal{L}_\kappa(\mathbf{x}, \mathbf{y})$$

$$= \int_\Omega \left(2D_v\ell(\mathbf{x}) - (\Delta\mathfrak{v})(\mathbf{x})\right) d\mu(\mathbf{x}) = \int_\Omega \left(\ell - \ell_\infty\right) d\mu \,.$$

Taking the limit $\Omega \nearrow N$ gives the result. $\qquad\square$

5.2. Proof of the positive mass theorem

We are now in the position to prove our positive mass theorem.

Proof of Theorem 1.9. Using (1.19) in (1.16) gives

$$\mathfrak{M}(\tilde\mu, \mu) = g \lim_{\Omega\nearrow N} \gamma_\mu^\Omega\left(\tilde v - v\right),$$

where we used the notation (2.18). For the computation of the surface layer integral $\gamma_\mu^\Omega(\tilde v)$ at infinity, we can work just as well with the measure $\tilde\mu$, i.e.

$$\mathfrak{M}(\tilde\mu, \mu) = g \lim_{\Omega\nearrow N} \left(\gamma_{\tilde\mu}^{F(\Omega)}\left(\tilde v\right) - \gamma_\mu^\Omega(v)\right).$$

This makes it possible to apply Proposition 5.1 to obtain

$$\mathfrak{M}(\tilde\mu, \mu) = g \int_{\tilde N} \left(\tilde\ell - \tilde\ell_\infty\right) d\tilde\mu - g \int_N \left(\ell - \ell_\infty\right) d\mu \,.$$

The last integral vanishes because μ is a vacuum measure (see Definition 1.4 (ii)). This concludes the proof. $\qquad\square$

6. Example: Asymptotically Schwarzschild spacetimes

In this section we establish the correspondence between the total mass and the ADM mass by proving Theorem 1.10.

6.1. Construction of static Dirac systems

Let $(\mathcal{M}, g)$ be a four-dimensional static globally hyperbolic Lorentzian manifold $\mathcal{M}$ with one asymptotic end which is asymptotically Schwarzschild. Thus we assume that the manifold is the topological product

$$\mathcal{M} = \mathbb{R} \times \mathcal{N} .$$

Moreover, denoting the coordinates by $x = (t, \mathbf{x})$ with $t \in \mathbb{R}$ and $\mathbf{x} \in \mathcal{N}$, we assume that the metric takes the form

$$ds^2 = L(\mathbf{x})^2 \, dt^2 - g_{\mathcal{N}} ,$$

where $L(\mathbf{x})$ is the lapse function and $g_{\mathcal{N}}$ is a complete Riemannian metric on $\mathcal{N}$. Finally, we assume that outside a compact set $K \subset \mathcal{N}$, the metric coincides with the Schwarzschild metric. Thus choosing polar coordinates $\mathbf{x} = (r, \vartheta, \varphi)$ on $\mathcal{N} \setminus K$, the line element becomes

$$(6.1) \quad ds^2 = \left(1 - \frac{2M_{\mathrm{S}}}{r}\right) dt^2 - \left(1 - \frac{2M_{\mathrm{S}}}{r}\right)^{-1} dr^2 - r^2 \left(d\theta^2 + \sin^2 \theta \, d\varphi^2\right) .$$

We now recall a few basics on the Dirac equation in the Schwarzschild geometry (for details and explicit formulas see [34] or [43]). For our purpose, it is most convenient to write the Dirac equation in the Hamiltonian form (2.43) and to take its spectral decomposition (2.45). The essential spectrum is determined from the asymptotic behavior of the metric at infinity. Therefore,

$$\sigma_{\mathrm{ess}}(H) = (-\infty, -m] \cup [m, \infty) .$$

In addition, there could be a point spectrum in $[-m, m]$ describing bound states of the system.

Following the general procedure in Section 2.3, we now construct static Dirac systems. The *Dirac sea* is defined as the negative spectral subspace of the essential spectrum,

$$\mathcal{H}_- := E_{(-\infty, m]}(\mathcal{H}_m) .$$

Following the general procedure in (2.46), we choose $\mathcal{H}$ as a subspace which differs from $\mathcal{H}_-$ by a finite-dimensional subspace. As a consequence, the kernel of the fermionic projector of our system differs from that of the Dirac

sea by smooth contributions,

$$P(x, y) = P^{\mathrm{sea}}(x, y) + (\text{smooth contributions}) \,.$$

Moreover, describing bound states, these smooth contributions decay rapidly near spatial infinity. Due to this rapid decay, the smooth contributions drop out when computing the surface layer integral near infinity. With this in mind, we may disregard the smooth contributions and simply choose

$$(6.2) \qquad\qquad \mathcal{H} = \mathcal{H}_- \,.$$

Moreover, for technical simplicity we choose the regularization operator (2.41) as a spatial convolution operator in the asymptotic end, i.e. for all $\mathbf{x} \in \mathcal{N} \setminus K$,

$$(6.3) \qquad
\begin{aligned}
\mathfrak{R}_\varepsilon \ &: \ \mathcal{H}_m \to C^0(\mathcal{M}, S\mathcal{M}) \cap \mathcal{H}_m \,, \\
(\mathfrak{R}_\varepsilon \psi)(\mathbf{x}) &= \int_{\mathcal{N}} h_\varepsilon\big(|\mathbf{x} - \mathbf{y}|\big)\, \psi(\mathbf{y})\, d\mu(\mathbf{y}) \,,
\end{aligned}$$

where $h_\varepsilon \in C_0^\infty(\mathbb{R}^3, \mathbb{R}_0^+)$ is a mollifier and $|\mathbf{x} - \mathbf{y}|$ denotes the Euclidean norm in our coordinate system near infinity (for the general context of this regularization method see [32, Section 4]). As already mentioned at the end of Section 2.3.2, the resulting static causal fermion systems are critical points of the causal action principle in the continuum limit, provided that the Einstein equations are satisfied. With this in mind, in what follows we assume that the static causal fermion systems satisfy the weak EL equations (2.9) (for the κ-Lagrangian).

6.2. Identifying the Dirac solution spaces

In preparation of the computation of the total mass, we need to identify the Dirac solution spaces in Minkowski space and in our asymptotically flat spacetime. This needs to be done in such a way that both spacetimes are jointly static (meaning that they are both static with respect to the same one-parameter group $(\mathcal{U}_t)_{t \in \mathbb{R}}$ in Definition 3.1). In order to distinguish the spacetimes, we again denote all objects of the asymptotically flat spacetime with a tilde, whereas the objects without a tilde refer to Minkowski space. Clearly, the corresponding one-parameter groups are the time evolution operators, i.e.

$$\mathcal{U}_t = e^{-itH} \ : \ \mathcal{H} \to \mathcal{H} \qquad \text{and} \qquad \tilde{\mathcal{U}}_t = e^{-it\tilde{H}} \ : \ \tilde{\mathcal{H}} \to \tilde{\mathcal{H}} \,,$$

where $\mathcal{H}$ and $\tilde{\mathcal{H}}$ are both chosen as the respective Dirac seas (6.2). Since both Hamiltonians have the same essential spectrum and no point spectrum, they can be mapped to each other by a unitary transformation, i.e. there is a

(6.4)   unitary $V : \mathcal{H} \to \tilde{\mathcal{H}}$ with $\tilde{H} = V H V^{-1}$.

Identifying $\mathcal{H}$ and $\tilde{\mathcal{H}}$ by this unitary transformation, the one-parameter groups $(\mathcal{U}_t)_{t \in \mathbb{R}}$ and $(\tilde{\mathcal{U}}_t)_{t \in \mathbb{R}}$ are also mapped to each other. In this way, the corresponding causal fermion systems become jointly static. Dividing out the group action, we get into the setting of causal variational principles.

    We point out that the above unitary transformation V is not unique, because it involves the freedom in unitarily transforming the Dirac solutions for every fixed energy. However, we proved in Theorem 4.7 that the total mass does not depend on the choice of V. With this in mind, in what follows we may choose the unitary transformation V in a convenient way. A particularly convenient choice of identification is obtained by a perturbative treatment, as we now explain (for other identifications of the Hilbert spaces see [43]).

6.3. Perturbative description near infinity

Since the metric $\tilde{g}$ is asymptotically flat, its effect can be treated asymptotically in first order perturbation theory. To this end, we decompose the Dirac Hamiltonian in the gravitational field as

$$\tilde{H} = H + \Delta H$$

and treat ΔH as a static perturbation in the Dirac equation for fixed energy ω. One must keep in mind that changing the metric also modifies the spatial integration measure in the scalar product (2.44). Compensating for this fact by a a local rescaling of the Dirac wave functions, one can work in a fixed Hilbert space (for details on this procedure see [12, Appendix A]). This has the advantage that we get a natural identification of $\mathcal{H}$ and $\tilde{\mathcal{H}}$. Then the above-mentioned non-uniqueness of the identification operator V reduces to the gauge freedom of the Dirac operator (for details see [11]). For a specific gauge, the Dirac operator is given explicitly in [34]. For the details of the perturbative treatment of the operator ΔH we refer again to [12, Appendix A] or, more generally, to [18, Appendix F]. The perturbation expansion is gauge covariant, meaning that gauge invariant quantities

like the closed chain do not depend on the choice of the gauge. The methods in [12] or, more generally, in [13] (for an introduction see [18, Section 2.2]) also give explicit formulas for the kernel of the fermionic projector. Non-perturbatively, these formulas correspond do the Hadamard expansion of the bi-distribution $P(x, y)$ (see for example the textbook [2]). In the presence of a regularization (6.3), the resulting *regularized Hadamard expansion* is worked out in [29].

In order to illustrate these results, we now state a few formulas which will be needed later on and make the connection to the jet formalism. Working with jets corresponds to linear perturbations by gravity. Thus we consider a metric g of the form

$$g_{ij} = \eta_{ij} + h_{ij} \,,$$

where η is the Minkowski metric and h_{ij} is the linear perturbation. The gravitational field modifies the light cones. The corresponding modification of the singularities of the unregularized kernel $P(x, y)$ on the light cone is described by the formula

$$(6.5) \qquad \Delta P(x, y) = -\frac{1}{2} \int_0^1 d\alpha \, h^i_j|_{\alpha y + (1-\alpha) x} \, \xi^j \, \frac{\partial}{\partial y^i} P(x, y) \,,$$

where we set $\xi = y - x$. This formula is derived in [12, Appendix A]. A more geometric way of understanding this formula is to integrate the geodesic equation; for details see Appendix B. In the static and spherically symmetric situation, the formula (6.5) remains valid for the regularized kernel (for details see again Appendix B),

$$(6.6) \qquad \Delta P^\varepsilon(x, y) = -\frac{1}{2} \int_0^1 d\alpha \, h^i_j|_{\alpha y + (1-\alpha) x} \, \xi^j \, \frac{\partial}{\partial y^i} P^\varepsilon(x, y) \,.$$

In addition to this effect of the "deformation of the light cone," there are effects by curvature. This becomes apparent in the formulas of the light cone expansion by terms which involve the Riemann tensor and its derivatives. Here we do not need the detailed form of these expressions. It suffices to keep in mind that these contributions are less singular on the light cone than (6.6).

In the jet formalism, the linear perturbation by gravity is described by a jet $\mathfrak{v}$. Note that the volume form in the Schwarzschild geometry in Schwarzschild coordinates does not depend on the mass, because

$$(6.7) \qquad d\tilde\rho = \sqrt{|\det \tilde g|} \, d^4x = dt \, d\mu \qquad \text{with} \qquad d\mu := r^2 \, dr \, d\omega \,,$$

where $d\omega := d\varphi\, d\cos\vartheta$ is the volume measure on the unit sphere. Therefore, the jet describing the linear perturbation by gravity has no scalar component,

$$(6.8) \qquad\qquad \mathfrak{v} = (0, v)\,.$$

The perturbation in (6.6) is obtained by perturbing the wave functions at both points x and y in the same way, i.e.

$$\Delta P^\varepsilon(x, y) = \left(D_{1,v} + D_{2,v}\right) P^\varepsilon(x, y)\,.$$

The surface layer integral needed for the computation of the total mass (see (1.16)) has a different form, because it involves the difference $D_1 - D_2$ of jet derivatives (this comes about because the arguments $\mathbf{x}$ and $\mathbf{y}$ in the formula for the total mass (1.11) lie in different spacetimes, one with and one without gravitational field). Therefore, one must extend the perturbative treatment to the case where the wave functions are perturbed only at x, but not at y or vice versa. This case is treated in [14, Appendix F] and [20, Section 5.1]. It amounts to replacing the bounded line integral in (6.5) by suitable unbounded line integrals. In the static and spherically symmetric case, the resulting formulas simplify to (for more details see Appendix B)

$$(6.9) \qquad
\begin{aligned}
D_{1,v} P^\varepsilon(x, y) &= \frac{1}{4} \int_{-\infty}^{\infty} d\alpha\, \epsilon(\alpha)\, h^i_j|_{\alpha y + (1-\alpha) x}\, \xi^j\, \frac{\partial}{\partial x^i} P^\varepsilon(x, y) \\
D_{2,v} P^\varepsilon(x, y) &= \frac{1}{4} \int_{-\infty}^{\infty} d\alpha\, \epsilon(\alpha - 1)\, h^i_j|_{\alpha y + (1-\alpha) x}\, \xi^j\, \frac{\partial}{\partial y^i} P^\varepsilon(x, y)\,,
\end{aligned}$$

where ϵ is the sign function. Clearly, there are again additional contributions involving curvature and its derivatives, but they are all less singular on the light cone.

Let us come back to the freedom in identifying the Hilbert spaces $\mathcal{H}$ and $\tilde{\mathcal{H}}$ as already mentioned after (6.4). The above perturbative procedure gives a canonical way to identify the Hilbert spaces in linearized gravity. However, it does *not* give a canonical identification of the Hilbert space $\mathcal{H}$ and $\tilde{\mathcal{H}}$ for two spacetimes whose gravitational fields coincide only asymptotically. Namely, in this case, the perturbative treatment is admissible only near spatial infinity. Consequently, it only gives an identification of the Hilbert spaces in the asymptotic ends. In general, this identification does not extend canonically to a unitary transformation which also preserves the Hamiltonians (6.4). In view of Theorem 4.7, the resulting non-uniqueness of the identifications has no effect on the total mass.

In more technical terms, the just-mentioned non-uniqueness of the identification of $\mathcal{H}$ and $\tilde{\mathcal{H}}$ becomes manifest in the fact that the unbounded line integrals in (6.9) are well-defined only in the asymptotic region where gravity can be treated linearly. One method of extending (6.9) to the spacetime regions with strong gravity is to replace the line integrals by integrals along null geodesics and h^i_j by a first order variation of the metric. But this procedure is not canonical. Here we do not need to worry about this issue because we already know from Theorem 4.7 that the total mass does not depend on the identification of Hilbert spaces. Correspondingly, we shall see in Proposition 6.3 below that the unbounded line integrals will drop out of our computations.

6.4. Implementing the volume constraint

After the above preparations, we can enter the computation of the total mass. It is most convenient to work with the formula (1.16). Thus we consider exhaustions Ω_n of N and $\tilde{\Omega}_n$ of $\tilde{N}$, subject to the volume constraint in (1.12)

$$(6.10) \qquad \mu(\Omega_n) = \tilde{\mu}(\tilde{\Omega}_n) < \infty \,.$$

Due to the independence of the choice of the exhaustion (Proposition 4.1) we may choose the sets as coordinate balls, i.e.

$$(6.11) \qquad \Omega_n = I \cup \Phi^{-1}\big(B_{R_n}\big) \qquad \text{and} \qquad \tilde{\Omega}_n = \tilde{I} \cup \tilde{\Phi}^{-1}\big(B_{\tilde{R}_n}\big)$$

(where we used the notation in Definition 1.3). For convenience, we choose the coordinates in the asymptotic end as the spatial part of the Schwarzschild coordinates (6.1) with mass $M_S = 0$ (for μ) and $M_S \neq 0$ (for $\tilde{\mu}$).

Since the inner volumes $\mu(I)$ and $\tilde{\mu}(\tilde{I})$ will in general be different, the volume constraint (6.10) forces us to also choose the radii differently, i.e. $R_n \neq \tilde{R}_n$. But, using that the volume form is independent of the mass (see (6.7)), the radii will coincide asymptotically in the sense that their difference decays quadratically,

$$|R_n - \tilde{R}_n| \lesssim \frac{1}{R_n^2} \,.$$

Working again with the identification of the spacetimes used in the perturbative description in the previous section, the mapping from Ω_n to $\tilde{\Omega}_n$ can be described near infinity by an infinitesimal volume-preserving diffeomorphism, i.e. by a divergence-free vector field u. We combine the corresponding

inner solution $\mathfrak{u} = (0, u)$ with the jet v in (6.8),

$$(6.12) \qquad\qquad\qquad w := v + u\,.$$

The resulting jet w describes the change of the metric near infinity, taking into account the volume constraint (6.10). Therefore, it can be used to compute the total mass via (1.16). We note for clarity that, while the jet v is given explicitly by (6.9), the inner solution u is *not* known, because it depends on the difference of the inner volumes of μ and $\tilde{\mu}$. We will come back to this issue in Section 6.6.

6.5. Reduction of the surface layer integral to a spacetime integral

Our task is to compute the surface layer integral in (1.16) for the jet w in (6.12). We write this surface layer integral in the short form

$$(6.13) \qquad \mathfrak{M}(\tilde{\mu}, \mu) = \lim_{R \to \infty} \int_{R_{\min}}^{R} dr \int_{R}^{\infty} dr'\, A(r, r') \qquad \text{with}$$

$$(6.14) \qquad A(r, r') = r^2\, r'^2 \int_{S^2} d\omega \int_{S^2} d\omega'\, \big(D_{1,w} - D_{2,w}\big) \mathcal{L}(\mathbf{x}, \mathbf{y})\,,$$

where $R_{\min}$ is any radius for which the sphere is contained in the domain of the coordinate chart at infinity (for example, one could choose $R_{\min} = R_1$ with R_1 as in (6.11)). Since the jet w is a solution of the linearized field equations without scalar component, the surface layer integral (6.13) does not depend on R. Using this conservation law, we can follow the procedure in [20, Lemma 4.2] (a similar method was first used in [25, proof of Lemma 5.5]) to rewrite the surface layer integral as an average of integrals over all of space:

Lemma 6.1. *The surface layer integral satisfies the relation*

$$\int_{R_{\min}}^{R} dr \int_{R}^{\infty} dr'\, A(r, r') = \lim_{L \to \infty} \frac{1}{2L} \int_{R}^{R+L} dr \int_{R_{\min}}^{\infty} dr'\, (r' - r)\, A(r, r')\,.$$

Proof. We closely follow [20, proof of Lemma 4.2]. Since the surface layer integral does not depend on R, we may take an average over R,

$$
\int_{R_{\min}}^{R} dr \int_{R}^{\infty} dr' \, A(r, r') = \frac{1}{L} \int_{0}^{L} d\ell \int_{R_{\min}}^{R+\ell} dr \int_{R+\ell}^{\infty} dr' \, A(r, r')
$$

$$
= \frac{1}{L} \int_{0}^{L} d\ell \int_{R_{\min}}^{\infty} dr \int_{R_{\min}}^{\infty} dr \, \Theta(R + \ell - r) \, \Theta(r' - R - \ell) \, A(r, r')
$$

$$
= \frac{1}{L} \int_{R_{\min}}^{\infty} dr \int_{R_{\min}}^{\infty} dr \, A(r, r') \left(\int_{0}^{L} \Theta(R + \ell - r) \, \Theta(r' - R - \ell) \, d\ell \right).
$$

Computing the integral over the Heaviside functions involves different cases. A straightforward calculation yields

(6.15)
$$
\int_{R_{\min}}^{R} dr \int_{R}^{\infty} dr' \, A(r, r') = \frac{1}{L} \int_{R_{\min}}^{R} dr \int_{R}^{R+L} dr' \, (r' - R) \, A(r, r')
$$

(6.16)
$$
+ \frac{1}{L} \int_{R_{\min}}^{R} dr \int_{R+L}^{\infty} dr' \, L \, A(r, r')
$$

(6.17)
$$
+ \frac{1}{L} \int_{R}^{R+L} dr \int_{r}^{R+L} dr' \, (r' - r) \, A(r, r')
$$

(6.18)
$$
+ \frac{1}{L} \int_{R}^{R+L} dr \int_{R+L}^{\infty} dr' \, (L - r + R) \, A(r, r').
$$

At this point, we make use of the fact that the causal Lagrangian is of short range in the sense that $A(r, r')$ decays at least cubically in $r' - r$, i.e.

$$
\big| A(r, r') \big| \leq \frac{c}{|r' - r|^3} \qquad \text{for all } r, r' > R
$$

and a suitable constant $c > 0$ (for details see [20] and [8, Appendix A]). As a consequence, the double integrals in (6.15), (6.16) and (6.18) are bounded uniformly in L. Therefore, in the limit $L \to \infty$ only the summand (6.17) contributes, i.e.

(6.19)
$$
\int_{R_{\min}}^{R} dr \int_{R}^{\infty} dr' \, A(r, r') = \lim_{L \to \infty} \frac{1}{L} \int_{R}^{R+L} dr \int_{r}^{R+L} dr' \, (r' - r) \, A(r, r').
$$

Rewriting the boundaries of integration in the last integral with the help of a Heaviside function, we may exchange the integrals and use that the

integrand is symmetric,

$$\int_R^{R+L} dr \int_r^{R+L} dr' \, (r' - r) \, A(r, r')$$

$$= \int_R^{R+L} dr \int_R^{R+L} dr' \, \Theta(r' - r) \, (r' - r) \, A(r, r')$$

$$= \int_R^{R+L} dr' \int_R^{R+L} dr \, \Theta(r' - r) \, (r' - r) \, A(r, r')$$

$$= \int_R^{R+L} dr \int_R^r dr' \, (r' - r) \, A(r, r') \, .$$

Using this relation, we can write (6.19) as

$$\int_{R_{\min}}^R dr \int_R^\infty dr \, A(r, r') = \lim_{L \to \infty} \frac{1}{2L} \int_R^{R+L} dr \int_R^{R+L} dr' \, (r' - r) \, A(r, r') \, .$$

Finally, changing the integration range of the last integral to $(R_{\min}, \infty)$ modifies the integrals only by a contribution which vanishes in the limit $L \to \infty$. This gives the result. $\qquad\square$

In view of this lemma, it suffices to compute the following expression for large R,

$$\mathfrak{M}(R) := \int_{R_{\min}}^\infty (r - R) \, A(R, r) \, dr$$

$$= \frac{R^2}{4\pi} \int_{R_{\min}}^\infty r^2 \, dr \int_{S^2} d\omega \, (r - R) \, \big(D_{1,w} - D_{2,w}\big) \mathcal{L}(\mathbf{x}, \mathbf{y}) \, ,$$

where, choosing polar coordinates, the points $\mathbf{x}$ and $\mathbf{y}$ have the form

(6.20) $$\mathbf{x} = (R, N) \qquad \text{and} \qquad \mathbf{y} = (r, \omega) \, ,$$

and $N \in S^2$ is the north pole. Including the time integral, we obtain

(6.21) $$\mathfrak{M}(R) = \frac{R^2}{4\pi} \int_{\mathcal{M}} (r - R) \, \big(D_{1,w} - D_{2,w}\big) \mathcal{L}\big(x, y\big) \, d\rho(y) \, ,$$

where $x = (0, \mathbf{x})$ and $y = (t, \mathbf{y})$. Our findings are summarized as follows:

Proposition 6.2. *Assume that integral expression $\mathfrak{M}(R)$ in (6.21) exists and converges in the limit $R \to \infty$. Then it coincides with the total mass,*

$$\mathfrak{M} = \lim_{R \to \infty} \mathfrak{M}(R) \, .$$

6.6. Computation of the spacetime integral

The remaining task is to compute the spacetime integral in (6.21). The first order variation of the kernel of the fermionic projector in the presence of a gravitational field was computed in [12, Appendix B]. These formulas also apply in our setting. Since the curvature tensor involves second derivatives of the metric, its components decay at least cubically,

$$\text{Riem} = \mathcal{O}\big(R^{-3}\big) \,.$$

As a consequence, the contributions to (6.21) involving curvature or derivatives of curvature vanish in the limit $R \to \infty$. Therefore, it suffices to consider the contributions by the infinitesimal diffeomorphism as given in (6.9). Moreover, we must also take into account the inner solution u in (6.12).

Proposition 6.3. *The volume constraint* (6.10) *can be taken into account in the computation of the spacetime integral in* (6.21) *by setting the inner solution u to zero and by computing the jet-derivatives instead of* (6.9) *by*

$$D_{1,w} P^\varepsilon(x,y) = 0$$
$$D_{2,w} P^\varepsilon(x,y) = -\frac{1}{2} \int_0^1 d\alpha \, h^i_j|_{\alpha y + (1-\alpha) x} \, \xi^j \, \frac{\partial}{\partial y^i} P^\varepsilon(x,y) \,.$$

Proof. We use a scaling argument which is based on the observation that contributions by the inner solution u to the surface layer integral (6.13) involve a scaling factor $\mathfrak{s}$. This can be seen in various ways: One way is to note that inner solutions describe infinitesimal diffeomorphisms and that the corresponding volume change shows up in (1.11) with a prefactor $\mathfrak{s}$. Another way is to evaluate the conservation law of Lemma 2.3 for an inner solution $\mathfrak{u} = (\text{div } u, u)$,

$$\int_\Omega d\mu(\mathbf{x}) \int_{N\backslash\Omega} d\mu(\mathbf{y}) \left(\nabla_{1,\mathfrak{u}} - \nabla_{2,\mathfrak{u}}\right) \mathcal{L}(\mathbf{x},\mathbf{y}) = \mathfrak{s} \int_\Omega \text{div } u \, d\mu \,.$$

Alternatively, this surface layer integral can be computed using integration by parts (for details see [22, eq. (3.3) in Proposition 3.4]). Such an integration-by-parts method can also be applied in (6.21), showing that the contribution of the inner solution to $\mathfrak{M}(R)$ again involves a factor $\mathfrak{s}$.

Using that the EL equations (2.3) hold for all $x \in N$, we can compute (2.2) asymptotically near infinity to obtain

$$\mathfrak{s} = \int_{\mathcal{M}} \mathcal{L}_\kappa \big(0, (t, \mathbf{y})\big) \, dt \, d^3 \mathbf{y} \, ,$$

where $\mathcal{L}_\kappa$ is the Lagrangian in the Minkowski vacuum. Comparing with the constant c in (1.22), one sees that the total mass is by at least one scaling factor ε *smaller* than the contribution by the inner solution. This will also become clear in the computations leading to (6.23) below, whereas a general scaling argument is given in Appendix A.

Combining these results, we conclude that the inner solutions compensate precisely all the contributions by the jet v in (6.9) to $\mathfrak{M}(R)$ which involve a scaling factor $\mathfrak{s}$. Since these jets describe an infinitesimal diffeomorphism, their contribution to (6.21) can be written as

(6.22)
$$\mathfrak{M}(R) \asymp \frac{R^2}{4\pi} \int_{\mathcal{M}} (r - R) \left(v^j(x, y) \frac{\partial}{\partial x^j} - v^j(y, x) \frac{\partial}{\partial y^j} \right) \mathcal{L}(x, y) \, d\rho(y) \, ,$$

where $v^j(x, y)$ is given by the line integral in (6.9). We now expand $v^j(x, .)$ and $v^j(., x)$ in Taylor series about x for y along a null line through x. A direct computation shows that the zero order term of this expansion gives precisely the contributions to (6.22) which involve a scaling factor $\mathfrak{s}$. Subtracting these contributions gives the result. □

We remark that in Proposition B.1, it is shown by explicit computation that the unbounded line integrals in (6.9) do not contribute to the surface layer integrals, provided that the linear perturbation of the metric does not change the volume form. This gives an alternative, more computational proof of the above proposition for linearized gravity.

In view of this result, we may keep the point x fixed. The change of the metric, however, can be described by the following change of coordinates:

Lemma 6.4. *Choosing in Minkowski space the coordinates*

$$\tilde{t} = t + t \, \frac{M_S}{R} \, , \quad \tilde{r} = r - (r - R) \, \frac{M_S}{R} \, , \quad \tilde{\vartheta} = \vartheta \, , \quad \tilde{\varphi} = \varphi \, ,$$

the new metric coincides with the Schwarzschild metric near x.

Proof. A short computation gives

$$t = \tilde{t} - \tilde{t}\,\frac{M_{\mathrm{S}}}{R} + \mathcal{O}\!\left(M_{\mathrm{S}}^2\right), \qquad\qquad r = \tilde{r} + (\tilde{r} - R)\,\frac{M_{\mathrm{S}}}{R} + \mathcal{O}\!\left(M_{\mathrm{S}}^2\right)$$

$$\frac{\partial t}{\partial \tilde{t}} = 1 - \frac{M_{\mathrm{S}}}{R}\,, \qquad\qquad \frac{\partial r}{\partial \tilde{r}} = 1 + \frac{M_{\mathrm{S}}}{R}$$

$$\tilde{g}_{00} = 1 - \frac{2M_{\mathrm{S}}}{R}\,, \qquad\qquad \tilde{g}_{11} = 1 + \frac{2M_{\mathrm{S}}}{R}\,,$$

giving agreement with the Schwarzschild metric linearly in M_{S}. □

In order to clarify the signs, we write this coordinate transformation as a diffeomorphism,

$$\Phi \ : \ (t, r, \omega) \mapsto (\tilde{t}, \tilde{r}, \omega)\,.$$

Then

$$\tilde{g}_{ij}\,\tilde{u}^i\,\tilde{u}^j = \eta_{ij}\,u^i u^j\,,$$

where $\tilde{g}$ and η are the Schwarzschild and Minkowski metrics, respectively. Therefore, the Lagrangian in the Schwarzschild metric can be written as $\mathcal{L}(x, \Phi^{-1}(y))$, where $\mathcal{L}$ denotes is the Lagrangian of the vacuum. Hence

$$D_{2,w}\mathcal{L}(x, y) = \mathcal{L}\!\left(x, \Phi^{-1}(y)\right) - \mathcal{L}(x, y) + \mathcal{O}\!\left(M_{\mathrm{S}}^2\right),$$

and using that

$$\Phi^{-1}(t, r, \omega) - (t, r, \omega) = \frac{M_S}{R}\left(-t, (r - R), 0\right) + \mathcal{O}\!\left(M_{\mathrm{S}}^2\right),$$

we conclude that

$$D_{2,w}\mathcal{L}(x, y) = \frac{M_{\mathrm{S}}}{R}\left(-t\,\frac{\partial}{\partial t} + (r - R)\,\frac{\partial}{\partial r}\right)\mathcal{L}(x, y) + \mathcal{O}\!\left(M_{\mathrm{S}}^2\right).$$

Hence, to first order in M_{S},

$$\mathfrak{M}(R) = \frac{R^2}{4\pi}\int_{\mathcal{M}} (r - R)\,\left(D_{1,w} - D_{2,w}\right)\mathcal{L}(x, y)\,d\rho(y)$$

$$= -\frac{R^2}{4\pi}\int_{\mathcal{M}} (r - R)\,D_{2,w}\mathcal{L}(x, y)\,d\rho(y)$$

$$
= -\frac{R\,M_{\mathrm{S}}}{4\pi} \int_{-\infty}^{\infty} dt \int_0^{\infty} dr \,(r - R)\, r^2
$$

$$
\times \int_{S^2} d\omega \left(-t\,\frac{\partial}{\partial t} + (r - R)\,\frac{\partial}{\partial r} \right) \mathcal{L}\big(x, (t, r, \omega)\big)
$$

$$
\overset{(*)}{=} -\frac{R\,M_{\mathrm{S}}}{4\pi} \int_{-\infty}^{\infty} dt \int_0^{\infty} dr \int_{S^2} d\omega\, \mathcal{L}\big(x, (t, r, \omega)\big)
$$

$$
\times \left(\frac{\partial}{\partial t}\big(t\,(r - R)\,r^2\big) - \frac{\partial}{\partial r}\big((r - R)^2\,r^2\big) \right)
$$

$$
= \frac{R\,M_{\mathrm{S}}}{4\pi} \int_{-\infty}^{\infty} dt \int_0^{\infty} dr \int_{S^2} d\omega\, \mathcal{L}\big(x, (t, r, \omega)\big)\, r\,(r - R)\,\big(2\,(r - R) + r\big),
$$

where in $(*)$ we integrated by parts. Carrying out the time integration according to (3.2) and again choosing polar coordinates (6.20), we obtain

$$
\mathfrak{M}(R) = \frac{R\,M_{\mathrm{S}}}{4\pi} \int_0^{\infty} dr \int_{S^2} d\omega\, \mathcal{L}(\mathbf{x}, \mathbf{y})\, r\,(r - R)\,\big(2\,(r - R) + r\big).
$$

From now on, we can work in Minkowski space. Assuming spherical symmetry, the Lagrangian depends only on the Euclidean distance, which with the law of cosines is given by

$$
|\mathbf{x} - \mathbf{y}|^2 = R^2 - r^2 - 2Rr\,\cos\vartheta = (r - R)^2 + 2Rr\,(1 - \cos\vartheta).
$$

In order to obtain a clean expansion in powers of $1/R$, it is useful to transform to the new coordinates

$$
\ell(r) = r - R \qquad \text{and} \qquad \sigma(\vartheta) = \sqrt{2Rr\,(1 - \cos\vartheta)}.
$$

Then

$$
d\ell = dr \qquad \text{and} \qquad \sigma\,d\sigma = Rr\,d\cos\vartheta.
$$

Using that $d\omega = 2\pi\,d\cos\vartheta$, the integral transforms to

$$
\mathfrak{M}(R) = \frac{R\,M_{\mathrm{S}}}{4\pi} \int_{-\infty}^{\infty} d\ell \int_0^{\infty} \frac{2\pi\sigma}{Rr}\, d\sigma\, \mathcal{L}(\mathbf{x}, \mathbf{y})\, r\,(r - R)\,\big(2\,(r - R) + r\big)
$$

$$
= \frac{M_{\mathrm{S}}}{4\pi} \int_{-\infty}^{\infty} d\ell \int_0^{\infty} 2\pi\sigma\,d\sigma\, \mathcal{L}\big[\ell^2 + \sigma^2\big]\,\ell\,\big(2\,\ell + (\ell + R)\big),
$$

where the square brackets indicate that the Lagrangian depends only on $\ell^2 + \sigma^2$. The linear term in ℓ drops out by symmetry,

$$
\mathfrak{M}(R) = \frac{M_{\mathrm{S}}}{4\pi} \int_{-\infty}^{\infty} d\ell \int_0^{\infty} 2\pi\sigma\,d\sigma\, \mathcal{L}\big[\ell^2 + \sigma^2\big]\, 3\ell^2.
$$

Now we can regard ℓ and σ as cylindrical coordinates in $\mathbb{R}^3$. Again using spherical symmetry, we obtain

$$\mathfrak{M}(R) = \frac{M_S}{4\pi} \int_{\mathbb{R}^3} |\mathbf{y}|^2 \, \mathcal{L}[\mathbf{y}^2] \, d^3\mathbf{y} \, .$$

Again writing out the time integration and applying Proposition 6.2, we conclude that the total mass is given by

$$(6.23) \qquad \mathfrak{M} = \frac{M_S}{4\pi} \int_{\mathcal{M}} |\mathbf{y}|^2 \, \mathcal{L}\big(0, (t, \mathbf{y})\big) \, d^4y \, .$$

This concludes the proof of Theorem 1.10.

7. Scaling behavior of Dirac systems

In this section we work out the scaling behavior for static Dirac systems and analyze the questions **(b)**–**(e)** posed in the introduction on page 1746. We begin by discussing the length scales in (1.23) in more detail. First of all, these length scales are related to each other by

$$\varepsilon \lesssim \delta \ll \frac{1}{m} \, .$$

At first sight, it might seem natural to identify the regularization length ε with the Planck scale δ. However, as is explained in detail in [18, Chapter 4], it is preferable to regard ε and δ as two different parameters, where typically $\varepsilon \ll \delta$.

We first recall the scalings in the *Minkowski vacuum* as worked out in [20] and [8, Appendix A]. The scaling of the function $\mathfrak{t}$ in (3.11) can be obtained in a straightforward way from a dimensional argument,

$$\mathfrak{t}(x) \simeq \frac{\sigma \, \lambda^4}{\varepsilon^8} \, ,$$

where σ and λ are the parameters describing the rescaling freedom (see (3.12) in Section 3.4). The contributions to the Lagrangian are less obvious because of our limited knowledge about the structure of physical spacetime on the Planck scale. Nevertheless, the Lagrangian can be written as

$$\ell(x) + \mathfrak{s} \simeq \sigma \, \lambda^4 \left(\frac{(\varepsilon m)^p}{\varepsilon^8} + \frac{1}{\delta^8} \left(\frac{\delta}{\varepsilon} \right)^{\hat{s}} \right)$$

with parameters p and $\hat{s}$, which from general considerations are known to be in the range

$$p \geq 5 \qquad \text{and} \qquad \hat{s} \in \{0, 2\} \, .$$

We next explain how to choose the parameters σ and λ used in the rescaling. It is most convenient to work in the units of length determined by the measure $\tilde{\mu}$. In order to keep this length scale fixed, we must not rescale the volume, meaning that σ must be kept fixed. For simplicity, we choose

$$\sigma = 1 \qquad \text{for all } \tau \, .$$

Moreover, the local trace must be kept constant (2.38). As is worked out in detail in [18, Section 2.5] and [8, Appendix A.3], this trace scales like

$$\operatorname{tr}(x) = \operatorname{Tr}\big(P(x, x)\big) \simeq \frac{\lambda m}{\varepsilon^2} \, .$$

This leads us to choose $\lambda \simeq \varepsilon^2 / m$. In this way, the freedom in scaling the measure is exhausted. Moreover, the above formulas for $\mathfrak{t}(x)$ and $\ell(x) + \mathfrak{s}$ simplify to

$$(7.1) \qquad \mathfrak{t}(x) \simeq \frac{1}{m^4} \qquad \text{and} \qquad \ell(x) + \mathfrak{s} \simeq \left(\frac{(\varepsilon m)^p}{m^4} + \frac{1}{m^8} \left(\frac{\varepsilon}{\delta} \right)^{8 - \hat{s}} \right) \, .$$

In *curved spacetime*, the functions $\ell(x)$ and $\mathfrak{t}(x)$ vary in spacetime, but the parameters in (1.23) as well as the Lagrange parameters $\mathfrak{s}$, κ and the parameter c in (2.38) are still constants. Moreover, we always consider critical measures. The resulting weak EL equations (2.9) (again for the κ-Lagrangian; see also (5.2)) read

$$\nabla_{\mathfrak{u}} \big(\ell + \kappa \, \mathfrak{t} \big) |_N = 0 \, .$$

In Definition 1.6 we consider a family $(\mu_\tau)_{\tau \in (-1, 1)}$ of critical static measures for a *decreasing* value of κ. The crucial question is how the parameters m and δ change when we decrease κ for a measure describing the Minkowski vacuum. The parameter κ is the Lagrange multiplier corresponding to the boundedness constraint. Therefore, decreasing κ corresponds to weakening the boundedness constraint by increasing the parameter C in (2.36). This has the effect that the functional in (2.36) becomes larger, whereas the causal action (2.33) becomes smaller. For a translation invariant system like Minkowski space, this means that $\mathfrak{t}(x)$ *increases*, whereas $\ell(x) + \mathfrak{s}$ *decreases*.

In view of the left side of (7.1), this means that the mass of the Dirac particles decreases,

$$(7.2) \qquad \frac{d}{d\tau} \log m \Big|_{\tau=0} < 0 \,.$$

The scaling behavior of δ is less obvious because of the unknown parameters p and $\hat{s}$ in (7.1). In view of the fact that the parameter δ was introduced in [18, Chapter 4] in order to compensate for the fact that the neutrino masses differ from the masses of the charged leptons, it is natural to assume that it scales in the same way as $1/m$, i.e.

$$(7.3) \qquad \frac{d}{d\tau} \log \big(m\,\delta\big) \Big|_{\tau=0} = 0 \,.$$

This *natural scaling of δ* means that when κ is decreased, the Compton length and the Planck length are increased by the same factor. When considering a gravitating system, the natural scaling of δ has the convenient property that the interaction remains unchanged, only the size of the whole system changes. Keeping in mind that the length scale is determined by the measure, one can also say that, assuming natural scaling of δ, only the measure $\tilde{\mu}$ and the regularization length ε are rescaled, but otherwise the system remains unchanged. Although the natural scaling of δ seems reasonable and sensible, it is not compelling, neither for mathematical nor for physical reasons. With our present knowledge, it is conceivable that for families of minimizers of the causal action, the parameter δ might have a different or more complicated scaling behavior.

Having the above scalings in mind, the relation (1.19) has a direct meaning: The vector field v changes the parameters m, δ and ε for the measure describing the Minkowski vacuum. Likewise, the vector field $\tilde{v}$ describes a variation of the gravitating system. Since asymptotically at infinity, the gravitating system goes over to Minkowski space, the parameters m, δ and ε of the gravitating system change just as in Minkowski space. As a consequence, the effect of the change of these parameters drops out when taking the *difference* of v and $\tilde{v}$ in (1.19). What remains is the change of the gravitating system, which in turn changes the strength of the gravitational field at infinity as described by the vector field w in (1.19). This consideration also explains why static Dirac systems are κ-scalable in the sense of Definition 1.7, thus answering question **(b)** on page 1746.

The above scaling analysis also makes it possible to address the questions **(c)**–**(e)** on page 1746. Making these considerations mathematically precise goes beyond the scope of this paper. Instead, we merely discuss these

questions in remarks which explain connections to be explored in more detail in the future.

Remark 7.1. Why is the gravitational force attractive? Having explained how the relation (1.19) comes about, we can now consider the sign of the gravitational coupling constant. As explained above, the jet $\tilde{v}$ describes a change of the gravitating system due to a variation of the parameters m, ε and δ, on the length scale determined by the measure $\tilde{\mu}$. Clearly, in view of (1.16) we only need to be concerned about the behavior near spatial infinity. Thus the question is how the strength of the gravitational field changes near infinity if the parameters m, ε and δ are varied infinitesimally. Clearly, increasing m makes the gravitational field stronger. In general relativity, the strength of the gravitational field is given by Gm (having the dimension of length). Keeping in mind that $G \sim \delta^2$, in the causal fermion system the corresponding quantity is $m\,\delta^2$. Thus for a static spacetime which is asymptotically Schwarzschild, the change of the gravitational field is given by the quantity

$$\frac{d}{d\tau} \log \left(m\,\delta^2 \right) \Big|_{\tau=0} .$$

Consequently, the gravitational constant g in (1.19) has the same sign as this quantity. If δ has the natural scaling, then this sign is *positive* in view of (7.2) and (7.3). More generally, the gravitational constant is positive, provided that the parameter $m\delta$ does not decrease too fast if κ is decreased. This seems a very sensible and natural assumption. However, exactly as explained after (7.3) for the natural scaling, there seems no compelling mathematical argument which explains the sign of g. ◇

Remark 7.2. Why is the local energy condition satisfied? It is sensible to assume that the measure μ describing the Minkowski vacuum is a minimizer of the causal action principle. Writing the action as

$$\mathcal{S} = \int_N d\mu(\mathbf{x}) \int_M d\rho(y)\, \mathcal{L}(\mathbf{x}, y) = \int_N \left(\ell(\mathbf{x}) + \mathfrak{s} \right) d\rho(\mathbf{x}) ,$$

the function $\ell(\mathbf{x}) + \mathfrak{s}$ can be regarded as the action per spatial volume. Likewise, in an asymptotically flat spacetime, the function $\tilde{\ell}(\mathbf{x}) - \ell_\infty$ tells about how the action per volume at $\mathbf{x}$ differs from the action per volume in the vacuum. With this in mind, the positivity of this function as imposed by the local energy condition in Definition 1.8 seems to be a direct consequence of the minimality of the causal action in the vacuum. Although being correct in principle, it seems that this argument cannot be made precise in a simple

way. The basic difficulty is that comparing the actions per volume in a rigorous way makes sense only in the homogeneous setting where these actions per volume are constant. Using a rescaling method where one "zooms into" spacetime on smaller and smaller scales, one can construct a homogeneous measure for which $\tilde{\ell}$ coincides with $\tilde{\ell}(\mathbf{x})$ of our curved spacetime. But in this limiting case, the rest mass m tends to zero. As a consequence, the contribution by matter to the function $\tilde{\ell}$ also tends to zero, which also means that the information on the sign of the energy density gets lost. This consideration shows that, in order to relate the sign of $\tilde{\ell}(\mathbf{x}) - \tilde{\ell}_\infty$ to the minimality of the causal action in the vacuum, one needs to enter a quantitative analysis of the scaling behavior of different contributions to $\tilde{\ell}(\mathbf{x})$.

In view of these difficulties, here we are content with the following weaker statement:

> Assume that a contribution by matter to $\tilde{\ell}(\mathbf{x})$ has the property that the same contribution can be arranged for a *homogeneous* physical system (i.e. for a system where $\tilde{\ell}(\mathbf{x})$ is constant). Then this contribution to $\tilde{\ell}(\mathbf{x}) - \tilde{\ell}_\infty$ is necessarily non-negative.

For the resulting homogeneous systems, the minimality of ℓ for the vacuum immediately gives the result. Typical examples for contributions to $\tilde{\ell}$ which can be "homogenized" in the above sense are the energy-momentum tensor of Dirac particles (where in the corresponding homogeneous system one replaces the Dirac wave functions of matter by a plane wave) or the energy-momentum tensor for a Maxwell field (in which case for the homogeneous system one takes a plane electromagnetic wave). In this formulation, our argument explains why the local energy condition is satisfied for classical matter on large scales. But we cannot exclude the possibility that there might be quantum fluctuations on microscopic scales which violate the local energy condition.   ◇

Remark 7.3. Why do vacuum measures describe flat spacetime?
For static Dirac systems, there is a direct way of understanding why a vacuum measure according to Definition 1.4 describes Minkowski space. Indeed, as is shown explicitly in [12, Appendix A], the curvature tensor enters the kernel of the fermionic projector $P(x, y)$. As a consequence, the curvature tensor has an effect on the eigenvalues λ_i^{xy} of the closed chain which enter the Lagrangian (2.32). The contributions involving the Ricci tensor are worked out in more detail in [18, §4.5.2]. Even without working out the detailed form of the resulting contributions, it is clear from the formulas in [12, Appendix A] that also the components of the Weyl tensor come into

play. Since the Lagrangian is non-negative and vanishes in the continuum limit, the resulting curvature contributions are strictly positive. Therefore, the condition (1.14) is satisfied if and only if the curvature tensor vanishes identically.   ◇

Appendix A. Scaling behavior of the total mass for static Dirac systems

The goal of this section is to show that for static Dirac systems, the total mass scales like

$$\mathfrak{M}(\tilde{\mu}, \mu) \lesssim \frac{\varepsilon}{l_{\mathrm{macro}}} \, \mathfrak{s} \, ,$$

where l_{macro} denotes the length scale on which the gravitational field changes. In order to derive this scaling behavior, we deform spacetime such as to obtain a spacetime with zero total mass. Then we analyze how the total mass changes along the path describing the deformation. Before entering the construction, we point out that the spacetimes along the deformed spacetimes are in general not static. Therefore, this appendix is the only place in this paper where we consider the time-dependent setting as introduced in Section 2.2.

We let $\tilde{\rho} = dt \, d\tilde{\mu}$ be the static spacetime of interest. We assume that the corresponding spacetime is of the form $M = \mathbb{R} \times N$, where N is a surface of constant time. Typically, the measure $\tilde{\rho}$ will be a causal fermion system constructed from a static Lorentzian spacetime as explained in Section 2.3, but this is not essential for the following construction. We first deform the measure $\tilde{\mu}$ such as to obtain a family of measures $(\tilde{\mu}_\tau)_{\tau \in [0,1)}$ with the following properties:

(i) For $\tau = 0$, we get back the static measure describing our spacetime of interest, i.e.

$$\tilde{\mu}_0 = \tilde{\mu} \, .$$

(ii) In the limit $\tau \nearrow 1$ and outside a compact set, the measures $\tilde{\mu}_\tau$ should go over to a measure describing a spatial hyperplane in Minkowski space.

(iii) The deformation should be smooth in the sense that

$$\tilde{N}_\tau = F_\tau(N) \qquad \text{with} \qquad F \in C^\infty([0,1) \times \tilde{N}, \mathcal{F}) \, .$$

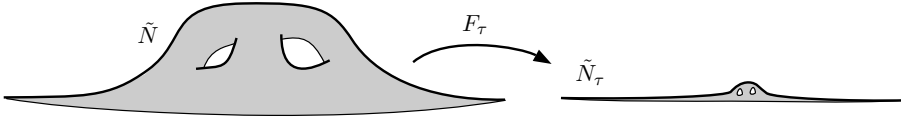This deformation is illustrated in Figure A1. The next step is to extend the

Figure A1: Deformation of the initial data.

measures $\tilde{\mu}_\tau$ to critical measures $\tilde{\rho}_\tau$ in spacetime. To this end, one can proceed in two alternative ways: One method is to solve the classical field equations (Einstein equations coupled to Dirac and possibly other fields) with initial data $\tilde{N}_\tau$, and then to define $\tilde{\rho}_\tau$ as in Section 2.3 as the push-forward of the volume measure under the local correlation map. Alternatively, one can also solve directly the EL equations of the causal action using the results and methods in [9]: Generalizing (2.12) to the time-dependent setting, we make the ansatz

$$\tilde{\rho}_\tau = (F_\tau)_* \big(f_\tau \, \rho \big) \, ,$$

where f and F are smooth,

$$f \in C^\infty \big([0,1) \times M \to \mathbb{R}^+ \big) \qquad \text{and} \qquad F \in C^\infty \big([0,1) \times M \to \mathfrak{F} \big) \, .$$

Next, using that

$$\tilde{\rho}_\tau = \tilde{\rho}_0 + \int_0^\tau \dot{\tilde{\rho}}_s \, ds \, ,$$

it suffices to solve the linearized field equations for each τ for the jet

$$\mathfrak{v}_\tau(x) = \frac{d}{d\tau} \big(f_\tau, F_\tau \big) \, .$$

As shown in [9], this Cauchy problem can be solved with energy methods. Before going on, we point out that the above procedure does *not* work if the solutions develop singularities, as is the case for example if a black hole forms. Thus, our deformation method requires that the deformed initial data admits solutions of the Cauchy problem which exist for a time which is sufficiently large for the surface layer integrals to be well-defined. This requirement seems sensible for most applications in mind.

Our next step is to incorporate the volume constraint by arranging that the weight function f_τ vanishes identically, i.e.

$$\tilde{\rho}_\tau = (F_\tau)_* \tilde{\rho} \qquad \text{with} \qquad F \in C^\infty \big([0,1) \times M, \mathfrak{F} \big) \, .$$

This can be arranged by Moser's theorem (i.e. by integrating the infinitesimal version of Lemma 2.7 or similarly [22, Proposition 3.6] in the hyperbolic setting).

After these preparation, we can analyze how the total mass changes along the curve $(\tilde{\rho}_\tau)_{\tau \in [0,1]}$. At $\tau = 0$, we can compute the total mass via (1.16), which in the spacetime setting we write as

$$\mathfrak{M} = \lim_{V \nearrow M} \int_{V \cap N} d\mu(x) \int_{M \setminus V} d\rho(y) \big(D_{1,w} - D_{2,w}\big) \mathcal{L}(x, y) \,,$$

and V is chosen of the form $\mathbb{R} \times \Omega$ with $\Omega \subset N$. Since the system is static, we can incorporate the time integral equivalently in the first argument. We thus obtain the alternative formula for the total mas

$$\mathfrak{M} = \frac{1}{2} \lim_{V \nearrow M} \left(\int_{V \cap N} d\mu(x) \int_{M \setminus V} d\rho(y) + \int_V d\rho(x) \int_{N \setminus V} d\mu(y) \right) \big(D_{1,w} - D_{2,w}\big) \mathcal{L}(x, y) \,,$$

which is indeed preferable for the following consideration. Applying the fundamental theorem of calculus and using that the total mass vanishes in the limit $\tau \nearrow 1$, we obtain

$$\mathfrak{M} = -\frac{1}{2} \lim_{V \nearrow M} \int_0^1 d\tau \left(\int_{V \cap N} d\mu(x) \int_{M \setminus V} d\rho(y) + \int_V d\rho(x) \int_{N \setminus V} d\mu(y) \right)$$
$$\times \big(D_{1,v_\tau} - D_{2,v_\tau}\big) \mathcal{L}(x, y) \,,$$

where v_τ is the vector field $v_\tau = dw/d\tau$. Since only the asymptotics near spatial infinity enters, this surface layer integral can be computed just as well using the perturbed measures,

$$\mathfrak{M}(V) = -\frac{1}{2} \lim_{\tilde{V} \nearrow \tilde{M}} \int_0^1 d\tau \left(\int_{\tilde{V} \cap \tilde{N}} d\tilde{\mu}_\tau(x) \int_{M \setminus V} d\tilde{\rho}_\tau(y) + \int_{\tilde{V}} d\tilde{\rho}_\tau(x) \int_{\tilde{N} \setminus \tilde{V}} d\tilde{\mu}_\tau(y) \right)$$
$$\times \big(D_{1,v_\tau} - D_{2,v_\tau}\big) \mathcal{L}\big(F_\tau(x), F_\tau(y)\big) \,.$$

This has the advantage that the jets v_τ are defined globally. Using that the integrand is anti-symmetric, we obtain

$$\left(\int_{\tilde{V} \cap \tilde{N}} d\tilde{\mu}_\tau(x) \int_{M \setminus V} d\tilde{\rho}_\tau(y) + \int_{\tilde{V}} d\tilde{\rho}_\tau(x) \int_{\tilde{N} \setminus \tilde{V}} d\tilde{\mu}_\tau(y) \right) \big(D_{1,v_\tau} - D_{2,v_\tau}\big) \mathcal{L}\big(F_\tau(x), F_\tau(y)\big)$$
$$= \left(\int_{\tilde{V} \cap \tilde{N}} d\tilde{\mu}_\tau(x) \int_M d\tilde{\rho}_\tau(y) + \int_{\tilde{V}} d\tilde{\rho}_\tau(x) \int_{\tilde{N}} d\tilde{\mu}_\tau(y) \right)$$
$$\times \big(D_{1,v_\tau} - D_{2,v_\tau}\big) \mathcal{L}\big(F_\tau(x), F_\tau(y)\big)$$

$$= \left(- \int_{\tilde{V} \cap \tilde{N}} d\tilde{\mu}_\tau(x) \int_M d\tilde{\rho}_\tau(y) + \int_{\tilde{V}} d\tilde{\rho}_\tau(x) \int_{\tilde{N}} d\tilde{\mu}_\tau(y) \right)$$
$$\times \left(D_{1,v_\tau} - D_{2,v_\tau} \right) \mathcal{L}\big(F_\tau(x), F_\tau(y)\big) ,$$

where in the last step we used the EL equations and the linearized field equations. The last expression is anti-symmetrized in the time integrals. As a consequence, we only get a contribution if either v_τ or F_τ are differentiated with respect to time. This gives the desired scaling factor $\varepsilon/l_{\mathrm{macro}}$.

Appendix B. Explicit treatment of a linearized gravitational field

We now explain geometrically how the formulas (6.5), (6.6) and (6.9) for the kernel of the fermionic projector in the presence of linearized gravity come about. Moreover, we analyze how the line integrals of the metric perturbation enter the surface layer integral which defines the total mass. Clearly, in curved spacetime the boundary of the light cone is generated by null geodesics. Perturbing the geodesic equation

$$\ddot{\gamma}^i(\alpha) = -\Gamma^i_{jk}\big(\gamma(\alpha)\big)\, \dot{\gamma}^j(\alpha)\, \dot{\gamma}^k(\alpha)$$

to first order about the straight line $\alpha y + (1 - \alpha)\, x$ gives

$$\ddot{\gamma}^i(\alpha) = -\Gamma^i_{jk}|_{\alpha y + (1-\alpha)\, x}\, \xi^j \xi^k ,$$

with the linearized Christoffel symbols given by

$$(B.1) \qquad \Gamma^i_{jk} = \frac{1}{2}\, \eta^{il}\big(\partial_j h_{lk} + \partial_k h_{lj} - \partial_l h_{jk}\big)$$

(where we again set $\xi = y - x$). Integrating by parts twice and keeping the geodesic fixed at $\alpha = -\infty$, one sees that

$$\Delta \gamma^i|_x = -\int_{-\infty}^0 \alpha\, \ddot{\gamma}^i(\alpha)\, d\alpha = \int_{-\infty}^0 \alpha\, \Gamma^i_{jk}\big|_{\alpha y + (1-\alpha)\, x}\, \xi^j \xi^k\, d\alpha$$
$$(B.2) \qquad = \frac{1}{2} \int_{-\infty}^0 \alpha\, \big(2\, \partial_j h^i{}_k - \partial^i h_{jk}\big)\big|_{\alpha y + (1-\alpha)\, x}\, \xi^j \xi^k\, d\alpha ,$$

where in the last step we used (B.1).

The transformation $x \mapsto x + \Delta \gamma|_x$ describes an infinitesimal diffeomorphism. Therefore, in a a suitable gauge and ignoring curvature terms, the

transformation of the kernel of the fermionic projector is described by an infinitesimal coordinate transformation,

$$\Delta P(x, y) = \left(\Delta\gamma^i|_x \frac{\partial}{\partial x^i} + \Delta\gamma^i|_y \frac{\partial}{\partial y^i} \right) P(x, y)$$

(B.3)
$$= -\left(\Delta\gamma^i|_x - \Delta\gamma^i|_y \right) \frac{\partial}{\partial y^i} P(x, y),$$

where in the last step we used that the kernel $P(x, y)$ in the Minkowski vacuum depends only on the difference vector $y - x$.

The above formulas can be further simplified. We begin with the *unregularized kernel*. In this case, we know from Lorentz invariance that the partial derivatives of $P(x, y)$ are proportional to the vector ξ. Contracting (B.2) with ξ, we obtain

(B.4)
$$\Delta\gamma^i|_x \, \xi_i = \frac{1}{2} \int_{-\infty}^{0} \alpha \, \partial_j h^i{}_k|_{\alpha y + (1-\alpha) x} \, \xi^j \xi^k \xi_i \, d\alpha \,.$$

Rewriting the directional derivative $\xi^j \partial_j$ as an α-derivative, we can integrate by parts to obtain

(B.5)
$$\Delta\gamma^i|_x \, \xi_i = -\frac{1}{2} \int_{-\infty}^{0} h^i{}_k|_{\alpha y + (1-\alpha) x} \, \xi^k \, \xi_i \, d\alpha \,.$$

Using this formula in (B.3) gives

$$\Delta P(x, y) = -\frac{1}{2} \left(\int_{-\infty}^{0} - \int_{-\infty}^{1} \right) d\alpha \, h^i{}_k|_{\alpha y + (1-\alpha) x} \, \xi^k \frac{\partial}{\partial y^i} P(x, y)$$

(B.6)
$$= \frac{1}{2} \int_{0}^{1} d\alpha \, h^i{}_k|_{\alpha y + (1-\alpha) x} \, \xi^k \frac{\partial}{\partial y^i} P(x, y) \,.$$

We thus obtain (6.5).

Before going on, we make a few remarks. Clearly, keeping the geodesic fixed at $\alpha = -\infty$ was an arbitrary choice. If instead we would have kept the geodesic fixed at $\alpha = +\infty$, the unbounded line integrals would have to be modified according to the replacement rules

$$\int_{-\infty}^{0} \to -\int_{0}^{\infty} \qquad \text{and} \qquad \int_{-\infty}^{1} \to -\int_{1}^{\infty} \,.$$

However, this has no effect on expressions involving bounded line integrals like (B.6). We also remark that the appearance of unbounded line integrals

in (B.5) motivates the line integrals in (6.9). The only additional ingredient needed in order to derive (6.9) is that the way causality is incorporated in the so-called causal perturbation expansion (see [10] or the more recent paper [35]) has the consequence that one must always take the *arithmetic mean* of the expressions obtained in the cases when the geodesic is fixed at $\alpha = -\infty$ and $\alpha = +\infty$.

We next consider the case *with regularization*. Since the regularized kernel $P^\varepsilon(x, y)$ is not Lorentz invariant, the method used after (B.4) no longer applies. But we can derive similar results in the *static* and *spherically symmetric* situation as follows. The first summand in the integrand in (B.2) contains a derivative $\xi^j \partial_j$ which we can again rewrite as an α-derivative and integrate by parts. The second summand in (B.2), however, is not of this form. In order to study this summand in more detail, we consider its contribution to one of the terms in (B.3),

(B.7)
$$\Delta\gamma^i|_x \frac{\partial}{\partial y^i} P^\varepsilon(x, y) \asymp -\frac{1}{2} \int_{-\infty}^{0} d\alpha \, \alpha \, \partial^i h_{jk}\big|_{\alpha y + (1-\alpha)\, x} \, \xi^j \xi^k \, \frac{\partial}{\partial y^i} P^\varepsilon(x, y) \, .$$

Using spherical symmetry and homogeneity, we can write the unregularized kernel of the vacuum as

(B.8)
$$P^\varepsilon(x, y) = P^\varepsilon\big[\xi^i \xi_i, \xi^0\big] \, .$$

Moreover, since h_{jk} is static, the index i in (B.7) is non-zero. Therefore, the partial derivatives do not act on the second argument of the kernel in (B.8). Consequently, the product rule gives again a factor ξ_i, making it possible to integrate by parts exactly as explained after (B.4). We conclude that

$$\Delta\gamma^i|_x \frac{\partial}{\partial y^i} P^\varepsilon\big[\xi^i \xi_i, \xi^0\big] = -\frac{1}{2} \int_{-\infty}^{0} d\alpha \, h^i{}_k\big|_{\alpha y + (1-\alpha)\, x} \, \xi^k \, \frac{\partial}{\partial y^i} P^\varepsilon\big[\xi^i \xi_i, \xi^0\big] \, .$$

Using this formula in (B.3) gives (6.6).

We finally analyze how the terms (6.9) contribute to the surface layer integral (1.16) describing the total mass. Since the jets in (6.9) describe an infinitesimal diffeomorphism, their contribution to (1.16) can be written in analogy to (6.22) as

(B.9)
$$\mathfrak{M}(\tilde{\mu}, \mu) \asymp \lim_{\Omega \nearrow N} \int_\Omega d\mu(\mathbf{x}) \int_{\mathbb{R} \times (N \setminus \Omega)} d\rho(y) \left(v^j(x, y) \frac{\partial}{\partial x^j} - v^j(y, x) \frac{\partial}{\partial y^j} \right) \mathcal{L}_\kappa(\mathbf{x}, \mathbf{y})$$

with

(B.10)
$$v^j(x, y) := \frac{1}{4} \int_{-\infty}^{\infty} d\alpha \, \epsilon(\alpha) \, h^j_k|_{\alpha y + (1-\alpha) \, x} \, \xi^k \, .$$

Proposition B.1. *The contribution to the mass given by* (B.9) *and* (B.10) *vanishes if the perturbation of the metric* h_{ij} *is compactly supported and trace-free.*

Proof. It suffices to consider the summand involving the x-derivative in (B.9), because the other summand can be treated in the same way. Thus our task is to analyze the integral expression

$$I := \int_{\Omega} d^3\mathbf{x} \int_{N \backslash \Omega} d^3\mathbf{y} \int_{-\infty}^{\infty} d\alpha \, \epsilon(\alpha) \, h^j_k|_{\alpha\mathbf{y} + (1-\alpha)\,\mathbf{x}} \int_{-\infty}^{\infty} d\xi^0 \, \xi^k \, \frac{\partial}{\partial x^j} \, \mathcal{L}_\kappa(x, y) \, .$$

The first step is to transform the integral over $\mathbf{y}$ to an integral over $\mathbf{z} := \alpha\mathbf{y} + (1 - \alpha)\,\mathbf{x}$,

$$
\begin{aligned}
I &= \{ \ \mathbf{z} - \mathbf{x} = \alpha\,(\mathbf{y} - \mathbf{x}), \ \ d^3\mathbf{z} = \alpha^3 \, d^3\mathbf{y} \ \} \\
&= \int_{-\infty}^{\infty} d\alpha \, \epsilon(\alpha) \int_{\mathbb{R}^3} \frac{d^3\mathbf{z}}{\alpha^3} \, h^j_k(\mathbf{z}) \\
&\quad \times \int_{\Omega} d^3\mathbf{x} \, \chi_{N \backslash \Omega}(\mathbf{y}) \int_{-\infty}^{\infty} d\xi^0 \, \xi^k \, \frac{\partial}{\partial x^j} \, \mathcal{L}_\kappa(x, y) \Big|_{\mathbf{y} = \mathbf{x} + \frac{\mathbf{z} - \mathbf{x}}{\alpha}} \, .
\end{aligned}
$$

Next, for the $\mathbf{x}$-integration we choose polar coordinates around $\mathbf{z}$, i.e.

(B.11)
$$\mathbf{x} = \mathbf{z} + r\,\zeta \qquad \text{with} \qquad r \in \mathbb{R}^+ \text{ and } \zeta \in S^2 \, .$$

We thus obtain

(B.12)
$$I = \int_{\mathbb{R}^3} d^3\mathbf{z} \, h^j_k(\mathbf{z}) \int_{S^2} d^2\zeta \, J^k_j(\mathbf{z}, \zeta)$$

with

$$J^k_j(\mathbf{z}, \zeta) := \int_{-\infty}^{\infty} d\alpha \, \frac{\epsilon(\alpha)}{\alpha^3} \int_0^{\infty} r^2 \, dr \, \chi_{\Omega}(\mathbf{x}) \, \chi_{N \backslash \Omega}(\mathbf{y}) \int_{-\infty}^{\infty} d\xi^0 \, \xi^k \, \frac{\partial}{\partial x^j} \, \mathcal{L}_\kappa(x, y) \, ,$$

where $\mathbf{x}$ and $\mathbf{y}$ are given by (B.11) and

$$\mathbf{y} = \mathbf{x} + \frac{\mathbf{z} - \mathbf{x}}{\alpha} = \mathbf{z} + \left(1 - \frac{1}{\alpha}\right)(\mathbf{x} - \mathbf{z}) = \mathbf{z} + \left(1 - \frac{1}{\alpha}\right) r\,\zeta \, .$$
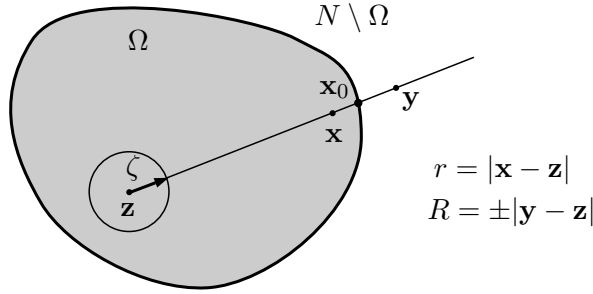
Figure B1: The unbounded line integrals in the surface layer integral.

Finally, we transform from the integration variable α to R defined by

$$\mathbf{y} = \mathbf{z} + R\zeta \quad \text{and thus} \quad R = \left(1 - \frac{1}{\alpha}\right) r \,.$$

Then

$$J_j^k(\mathbf{z}, \zeta) = \left\{ \alpha = \frac{r}{r - R} \,, \quad d\alpha = \frac{\alpha^2}{r} \, dR \right\}$$

$$= \int_0^\infty r^2 \, dr \int_{-\infty}^\infty \frac{dR}{r\,\alpha} \, \epsilon(r - R) \, \chi_\Omega(\mathbf{x}) \, \chi_{N\setminus\Omega}(\mathbf{y}) \int_{-\infty}^\infty d\xi^0 \, \xi^k \, \frac{\partial}{\partial x^j} \, \mathcal{L}_\kappa(x, y)$$

$$= \int_0^\infty dr \int_{-\infty}^\infty dR \, |r - R| \, \chi_\Omega(\mathbf{x}) \, \chi_{N\setminus\Omega}(\mathbf{y}) \int_{-\infty}^\infty d\xi^0 \, \xi^k \, \frac{\partial}{\partial x^j} \, \mathcal{L}_\kappa(x, y) \,.$$

The integrals over r and R are integrals along the straight line $z + \zeta \mathbb{R}$, as shown in Figure B1. Let us assume that the ray $z + \zeta \mathbb{R}^+$ intersects the boundary of Ω only once at a point $\mathbf{x}_0 = z + r_0 \, \zeta$ (this is clearly the case if we chose Ω as a ball or a convex set). Moreover, assume that z lies inside Ω. Then, introducing the new integration variables $\tau = r - r_0$ and $\tau' = R - r_0$, we obtain

$$J_j^k(\mathbf{z}, \zeta) = \int_{-r_0}^0 d\tau \int_0^\infty d\tau' \, (\tau' - \tau) \int_{-\infty}^\infty d\xi^0 \, \xi^k \, \frac{\partial}{\partial x^j} \, \mathcal{L}_\kappa(x, y) \Big|_{\substack{x = x_0 + \tau\zeta \\ y = x_0 + \tau'\zeta}} \,.$$

If the set Ω is chosen as a ball whose radius tends to infinity, this simplifies to

$$J_j^k(\mathbf{z}, \zeta) = \int_{-\infty}^0 d\tau \int_0^\infty d\tau' \, (\tau' - \tau) \int_{-\infty}^\infty d\xi^0 \, \xi^k \, \frac{\partial}{\partial x^j} \, \mathcal{L}_\kappa(x, y) \Big|_{\substack{x = x_0 + \tau\zeta \\ y = x_0 + \tau'\zeta}} \,.$$

Moreover, we can evaluate these integrals for the regularized Lagrangian in Minkowski space. In particular, using that the Lagrangian depends only on the difference vector $\xi = y - x$, one of the line integrals can be carried out,

$$
\begin{aligned}
J_j^k(\mathbf{z}, \zeta) &= -\int_{-\infty}^0 d\tau \int_0^\infty d\tau' \, (\tau' - \tau) \int_{-\infty}^\infty d\xi^0 \, \xi^k \, \frac{\partial}{\partial \xi^j} \, \mathcal{L}_\kappa[\xi] \Big|_{\xi = \left(\xi^0, (\tau' - \tau)\zeta\right)} \\
&= \left\{ \begin{array}{c} \beta = \tau' - \tau \\ d\beta = d\tau' \end{array} \right\} \\
&= -\int_{-\infty}^0 d\beta \int_{-\beta}^0 d\tau \, \beta \int_{-\infty}^\infty d\xi^0 \, \xi^k \, \frac{\partial}{\partial \xi^j} \, \mathcal{L}_\kappa[\xi] \Big|_{\xi = (\xi^0, \beta\zeta)} \\
&= -\int_{-\infty}^0 d\beta \, \beta^2 \int_{-\infty}^\infty d\xi^0 \, \xi^k \, \frac{\partial}{\partial \xi^j} \, \mathcal{L}_\kappa[\xi] \Big|_{\xi = (\xi^0, \beta\zeta)} .
\end{aligned}
$$

Let us consider the different cases for the tensor indices. If $j = 0$, we can integrate by parts in y^0 to obtain

$$
J_0^k(\mathbf{z}, \zeta) = \delta_0^k \int_0^\infty d\beta \, \beta^2 \int_{-\infty}^\infty d\xi^0 \, \mathcal{L}_\kappa[\xi] \Big|_{\xi = (\xi^0, \beta\zeta)} .
$$

We next compute the trace,

$$
\begin{aligned}
J_k^k(\mathbf{z}, \zeta) &= -\int_0^\infty d\beta \, \beta^2 \int_{-\infty}^\infty d\sigma \, \beta \, \xi^k \, \frac{\partial}{\partial \xi^k} \, \mathcal{L}_\kappa[\xi] \Big|_{\xi = \beta \, (\sigma, \zeta)} \\
&= -\int_0^\infty d\beta \, \beta^2 \int_{-\infty}^\infty d\sigma \, \beta^2 \, \frac{\partial}{\partial \beta} \mathcal{L}_\kappa[\xi] \Big|_{\xi = \beta \, (\sigma, \zeta)} \\
&= -\int_0^\infty d\beta \, \beta^4 \, \frac{\partial}{\partial \beta} \int_{-\infty}^\infty d\sigma^0 \, \mathcal{L}_\kappa[\xi] \Big|_{\xi = \beta \, (\sigma, \zeta)} \\
&= 4 \int_0^\infty d\beta \, \beta^3 \int_{-\infty}^\infty d\sigma^0 \, \mathcal{L}_\kappa[\xi] \Big|_{\xi = \beta \, (\sigma, \zeta)} .
\end{aligned}
$$

Comparing the above formulas, we conclude that

$$
J_k^k(\mathbf{z}, \zeta) = 4 \, J_0^0(\mathbf{z}, \zeta) .
$$

Using spherical symmetry, it follows that

$$
J_j^k(\mathbf{z}, \zeta) = c(\mathbf{z}) \left(\delta_0^k \, \delta_j^0 + 3 \, \zeta^k \zeta_j \right)
$$

with a scalar function $c(\mathbf{z})$. As a consequence, integrating over ζ in (B.12) gives

$$I = 4\pi \int_{\mathbb{R}^3} d^3\mathbf{z}\, h_k^j(\mathbf{z})\, c(\mathbf{z})\, \delta_j^k = 0\,,$$

because $h_k^j(\mathbf{z})$ is assumed to be trace-free. $\qquad\square$

References

[1] R. Arnowitt, S. Deser, and C.W. Misner, *Energy and the criteria for radiation in general relativity*, Phys. Rev. (2) **118** (1960), 1100–1104.

[2] C. Bär, N. Ginoux, and F. Pfäffle, *Wave Equations on Lorentzian Manifolds and Quantization*, `arXiv:0806.1036 [math.DG]`, ESI Lectures in Mathematics and Physics, European Mathematical Society (EMS), Zürich, 2007.

[3] H. Baum, *Spinor structures and Dirac operators on pseudo-Riemannian manifolds*, Bull. Polish Acad. Sci. Math. **33** (1985), no. 3-4, 165–171.

[4] A.N. Bernal and M. Sánchez, *On smooth Cauchy hypersurfaces and Geroch's splitting theorem*, `arXiv:gr-qc/0306108`, Commun. Math. Phys. **243** (2003), no. 3, 461–470.

[5] Y. Bernard and F. Finster, *On the structure of minimizers of causal variational principles in the non-compact and equivariant settings*, `arXiv:1205.0403 [math-ph]`, Adv. Calc. Var. **7** (2014), no. 1, 27–57.

[6] V.I. Bogachev, *Measure Theory. Vol. I*, Springer-Verlag, Berlin, 2007.

[7] P.R. Chernoff, *Essential self-adjointness of powers of generators of hyperbolic equations*, J. Functional Analysis **12** (1973), 401–414.

[8] E. Curiel, F. Finster, and J.M. Isidro, *Two-dimensional area and matter flux in the theory of causal fermion systems*, `arXiv:1910.06161 [math-ph]`, Internat. J. Modern Phys. D **29** (2020), 2050098.

[9] C. Dappiaggi and F. Finster, *Linearized fields for causal variational principles: Existence theory and causal structure*, `arXiv:1811.10587` `[math-ph]`, Methods Appl. Anal. **27** (2020), no. 1, 1–56.

[10] F. Finster, *Definition of the Dirac sea in the presence of external fields*, `arXiv:hep-th/9705006`, Adv. Theor. Math. Phys. **2** (1998), no. 5, 963–985.

[11] _____, *Local* U$(2, 2)$ *symmetry in relativistic quantum mechanics*, `arXiv:hep-th/9703083`, J. Math. Phys. **39** (1998), no. 12, 6276–6290.

[12] _____, *Light-cone expansion of the Dirac sea to first order in the external potential*, `arXiv:hep-th/9707128`, Michigan Math. J. **46** (1999), no. 2, 377–408.

[13] _____, *Light-cone expansion of the Dirac sea in the presence of chiral and scalar potentials*, `arXiv:hep-th/9809019`, J. Math. Phys. **41** (2000), no. 10, 6689–6746.

[14] _____, *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.

[15] _____, *A variational principle in discrete space-time: Existence of minimizers*, `arXiv:math-ph/0503069`, Calc. Var. Partial Differential Equations **29** (2007), no. 4, 431–453.

[16] _____, *Causal variational principles on measure spaces*, `arXiv:0811.2666` `[math-ph]`, J. Reine Angew. Math. **646** (2010), 141–194.

[17] _____, *A formulation of quantum field theory realizing a sea of interacting Dirac particles*, `arXiv:0911.2102` `[hep-th]`, Lett. Math. Phys. **97** (2011), no. 2, 165–183.

[18] _____, *The Continuum Limit of Causal Fermion Systems*, `arXiv:1605.04742` `[math-ph]`, Fundamental Theories of Physics, vol. 186, Springer, 2016.

[19] _____, *Causal fermion systems: A primer for Lorentzian geometers*, `arXiv:1709.04781` `[math-ph]`, J. Phys.: Conf. Ser. **968** (2018), 012004.

[20] _____, *The causal action in Minkowski space and surface layer integrals*, `arXiv:1711.07058` `[math-ph]`, SIGMA Symmetry Integrability Geom. Methods Appl. **16** (2020), no. 091.

[21] _____ , *Perturbation theory for critical points of causal variational principles*, `arXiv:1703.05059 [math-ph]`, Adv. Theor. Math. Phys. **24** (2020), no. 3, 563–619.

[22] F. Finster and N. Kamran, *Complex structures on jet spaces and bosonic Fock space dynamics for causal variational principles*, `arXiv:1808.03177 [math-ph]`, Pure Appl. Math. Q. **17** (2021), no. 1, 55–140.

[23] _____ , *Fermionic Fock spaces and quantum states for causal fermion systems*, `arXiv:2101.10793 [math-ph]`, Ann. Henri Poincaré **23** (2022), no. 4, 1359–1398.

[24] F. Finster and S. Kindermann, *A gauge fixing procedure for causal fermion systems*, `arXiv:1908.08445 [math-ph]`, J. Math. Phys. **61** (2020), no. 8, 082301.

[25] F. Finster and J. Kleiner, *Noether-like theorems for causal variational principles*, `arXiv:1506.09076 [math-ph]`, Calc. Var. Partial Differential Equations **55:35** (2016), no. 2, 41.

[26] _____ , *A Hamiltonian formulation of causal variational principles*, `arXiv:1612.07192 [math-ph]`, Calc. Var. Partial Differential Equations **56:73** (2017), no. 3, 33.

[27] _____ , *A class of conserved surface layer integrals for causal variational principles*, `arXiv:1801.08715 [math-ph]`, Calc. Var. Partial Differential Equations **58:38** (2019), no. 1, 34.

[28] F. Finster, J. Kleiner, and J.-H. Treude, *An Introductory Course on Causal Fermion Systems*, in preparation, `https://www.causal-fermion-system.com/intro-public.pdf`.

[29] F. Finster and M. Kraus, *The regularized Hadamard expansion*, `arXiv:1708.04447 [math-ph]`, J. Math. Anal. Appl. **491** (2020), no. 2, 124340.

[30] F. Finster and C. Langer, *Causal variational principles in the σ-locally compact setting: Existence of minimizers*, `arXiv:2002.04412 [math-ph]`, to appear in Adv. Calc. Var. (2022).

[31] F. Finster and M. Lottner, *Banach manifold structure and infinite-dimensional analysis for causal fermion systems*, `arXiv:2101.11908 [math-ph]`, Ann. Global Anal. Geom. **60** (2021), no. 2, 313–354.

[32] F. Finster and M. Reintjes, *A non-perturbative construction of the fermionic projector on globally hyperbolic manifolds I – Space-times of finite lifetime*, `arXiv:1301.5420 [math-ph]`, Adv. Theor. Math. Phys. **19** (2015), no. 4, 761–803.

[33] F. Finster and C. Röken, *Self-adjointness of the Dirac Hamiltonian for a class of non-uniformly elliptic boundary value problems*, `arXiv:1512.00761 [math-ph]`, Annals of Mathematical Sciences and Applications **1** (2016), no. 2, 301—-320.

[34] F. Finster, J. Smoller, and S.-T. Yau, *Non-existence of time-periodic solutions of the Dirac equation in a Reissner-Nordström black hole background*, `arXiv:gr-qc/9805050`, J. Math. Phys. **41** (2000), no. 4, 2173–2194.

[35] F. Finster and J. Tolksdorf, *Perturbative description of the fermionic projector: Normalization, causality and Furry's theorem*, `arXiv:1401.4353 [math-ph]`, J. Math. Phys. **55** (2014), no. 5, 052301.

[36] P.R. Halmos, *Measure Theory*, Springer, New York, 1974.

[37] S. Helgason, *Groups and Geometric Analysis*, Mathematical Surveys and Monographs, vol. 83, American Mathematical Society, Providence, RI, 2000, Integral geometry, invariant differential operators, and spherical functions, Corrected reprint of the 1984 original.

[38] M. Herzlich, *Universal positive mass theorems*, `arXiv:1401.6009 [math.DG]`, Commun. Math. Phys. **351** (2017), no. 3, 973–992.

[39] T. Kato, *Perturbation Theory for Linear Operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995.

[40] S. Lang, *Fundamentals of Differential Geometry*, Graduate Texts in Mathematics, vol. 191, Springer-Verlag, New York, 1999.

[41] H.B. Lawson, Jr. and M.-L. Michelsohn, *Spin Geometry*, Princeton Mathematical Series, vol. 38, Princeton University Press, Princeton, NJ, 1989.

[42] T. Parker and C.H. Taubes, *On Witten's proof of the positive energy theorem*, Commun. Math. Phys. **84** (1982), no. 2, 223–238.

[43] A. Platzer, *The Concept of Mass for Asymptotically Flat and Static Causal Fermion Systems*, Dissertation, Universität Regensburg, 2019.

[44] R. Schoen and S.-T. Yau, *On the proof of the positive mass conjecture in general relativity*, Commun. Math. Phys. **65** (1979), no. 1, 45–76.

[45] _____, *Positive scalar curvature and minimal hypersurface singularities*, `arXiv:1704.05490` `[math.DG]` (2017).

[46] E. Witten, *A new proof of the positive energy theorem*, Commun. Math. Phys. **80** (1981), no. 3, 381–402.

Fakultät für Mathematik, Universität Regensburg
D-93040 Regensburg, Germany
E-mail address: `finster@ur.de`
E-mail address: `andy.platzer@web.de`