

Unimodular gravity and general relativity from graviton self-interactionsCarlos Barceló,^{1,*} Raúl Carballo-Rubio,^{1,†} and Luis J. Garay^{2,3,‡}¹*Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía, 18008 Granada, Spain*²*Departamento de Física Teórica II, Universidad Complutense de Madrid, 28040 Madrid, Spain*³*Instituto de Estructura de la Materia (IEM-CSIC), Serrano 121, 28006 Madrid, Spain*

(Received 5 March 2014; published 16 June 2014)

It is commonly accepted that general relativity is the only solution to the consistency problem that appears when trying to build a theory of interacting gravitons (massless spin-2 particles). Padmanabhan's 2008 thought-provoking analysis raised some concerns that are having resonance in the community. In this paper we present the self-coupling problem in detail and explicitly solve the infinite-iterations scheme associated with it for the simplest theory of a graviton field, which corresponds to an irreducible spin-2 representation of the Poincaré group. We make explicit the nonuniqueness problem by finding an entire family of solutions to the self-coupling problem. Then we show that the only resulting theory which implements a deformation of the original gauge symmetry happens to have essentially the structure of unimodular gravity. This makes plausible the possibility of a natural solution to the first cosmological constant problem in theories of emergent gravity. Later on, we change for the sake of completeness the starting free-field theory to Fierz-Pauli theory, an equivalent theory but with a larger gauge symmetry. We indicate how to carry out the infinite summation procedure in a similar way. Overall, we conclude that as long as one requires the (deformed) preservation of internal gauge invariance, one naturally recovers the structure of unimodular gravity or general relativity but in a version that explicitly shows the underlying Minkowski spacetime, in the spirit of Rosen's flat-background bimetric theory.

DOI: [10.1103/PhysRevD.89.124019](https://doi.org/10.1103/PhysRevD.89.124019)

PACS numbers: 04.20.-q, 04.60.-m, 04.62.+v, 11.30.Cp

I. INTRODUCTION

The tight connection between general relativity and geometry is what makes this theory conceptually beautiful, but also very different from the formalism developed to describe the other fundamental forces we know about: the standard model of particle physics. While the latter is formulated as a quantum field theory in Minkowski spacetime, in general relativity there is no such notion of preferred, immutable arena in which physics takes place. Instead this environment (spacetime) is also a dynamical object in its own right. This is arguably the root of the conceptual problems concerning the reconciliation between general relativity and quantum mechanics.

Trying to bridge this gap Rosen [1,2] showed that general relativity can be reinterpreted as a nonlinear field theory over Minkowski spacetime. Later Gupta [3] proposed that a consistent theory of self-interacting gravitons should have precisely the structure of general relativity. In brief, Gupta's idea is to start with a free massless spin-2 field in Minkowski spacetime, and then make it interact with the rest of the fields. General considerations show that the fact that this field has spin 2 implies that this can be done only if the graviton field interacts with itself, making the overall theory nonlinear. Since one can always express

general relativity plus matter as a nonlinear theory for the deviations of the metric with respect to some flat reference metric, there is wiggle room to reconcile both visions.

However, to make this program come to fruition, one should be able to determine the nature of the resulting nonlinear theory which arises from the self-coupling of the graviton field. The first subtle point is that the Lagrangian density of such a theory contains, in principle, infinite interaction terms which are obtained consecutively by an iterative process, so one should devise a way to manage them and show that this infinite series converges to the Lagrangian density of general relativity. This question was indirectly addressed in the work of Kraichnan [4] and Feynman [5], but was finally settled by Deser [6]. To do that, he used specific variables which make the series finite, thus avoiding to perform the sum of an infinite series.

The second source of concern is the nonuniqueness of the construction as there are many and, in principle, inequivalent ways to make the graviton field self-interact. This was first raised by Huggins in his 1962 thesis [7]. The central point of his argument is that one needs more information to uniquely fix the stress-energy tensor of the graviton field to which it couples itself. Thus there are potentially many self-interacting theories and, as there is no control of those theories, it is not easy to conclude whether or not they are equivalent to general relativity. Recently, Padmanabhan has raised equivalent arguments [8]. In fact, this work has been motivated by Padmanabhan's paper, a subsequent follow-up by Butcher *et al.* [9], and the reply by Deser [10].

* carlos@iaa.es

† raulc@iaa.es

‡ luisj.garay@ucm.es

In this paper we start by considering the simplest theory of a graviton field, corresponding to an irreducible spin-2 representation of the Poincaré group. After describing the linear spin-2 theory, we develop in detail the self-interacting scheme and find the (formal) sum of the series for the action. To the best of our knowledge, it is the first time that an infinite series arising in the graviton self-coupling problem is summed. This construction is not unique. We present the nonuniqueness problem and show that, indeed, there is a one-parameter family of solutions to the self-coupling problem which were not found in previous approaches. We analyze these resulting theories and show that requiring the preservation of gauge symmetry can be used to single out one of them, which turns out to be equivalent to unimodular gravity. After the detailed analysis of the spin-2 theory, we move to compare this approach with the more standard that starts with Fierz-Pauli theory. We sketch how to solve in an equivalent way the self-coupling problem in this case. We study in a similar way the issue of uniqueness and then, at the light of our results, we present a discussion aimed at reconciling what seem disparate results in the recent literature [8–10].

On the one hand, our results confirm the concerns of [8] in that the self-coupling problem by itself does not uniquely lead to unimodular gravity (or general relativity, depending on the starting linear theory) unless further conditions are imposed along the process. Specifically, one needs to require that the gauge structure of the initial linear theory is preserved, although deformed, in the final outcome. This condition singles out unimodular gravity (or general relativity) but in a version that explicitly shows the underlying Minkowski spacetime, in the spirit of Rosen’s flat-background bimetric theory. On the other hand, once the gauge preservation condition is applied, the entire construction can be taken to completion in a natural way using only flat spacetime notions, a position which is defended in [10]. In fact, the presence of the Minkowski background structure permits one to clearly separate the internal gauge transformations from invariance under changes of coordinates. Within this bimetric construction, one obtains a quadratic Lagrangian density, invariant under changes of coordinates. However, this quadratic Lagrangian density is not invariant under internal gauge transformations, but the action is. One could take a further leap and adopt a geometrical (single-metric) interpretation [6,10]. From this perspective the quadratic Lagrangian density is not diffeomorphism invariant: the surface term of the Einstein-Hilbert action cannot be recovered and would have to be added by hand, as pointed out in [8]. Although certainly appealing, within the self-coupling problem we do not find a compelling reason to take this geometrization leap.

The structure of the paper is the following. Sections II and III are intended to be a recapitulation of the knowledge which can be found in different sources, concerning the field-theoretical description of gravitons and the consistency

problem which appears when one tries to make them interact with matter. This problem leads to the consideration of a self-interacting scheme for the gravitons, which is developed in the next two sections, the core of the paper. In Sec. IV we consider the simplest theory of a graviton field, corresponding to an irreducible spin-2 representation of the Poincaré group. We develop the self-interacting scheme and find the (formal) sum of the series for the action. In Sec. V we discuss how the condition of preservation of the maximal amount of internal gauge invariance can eliminate the ambiguities inherent to the self-coupling procedure. We also show that this outcome of the self-coupling problem is equivalent to unimodular gravity. Section VI is devoted to show how to apply the same program to Fierz-Pauli theory and how general relativity comes into play. We also include a discussion on gravitational energy which is particularly interesting in the case of unimodular gravity. We end with a brief summary and some conclusions.

Notation and writing style.—We use the metric convention $(-, +, +, +)$ and we will always avoid making explicit the spatiotemporal dependence of the different fields considered in the text. No distinction is made between Greek and Latin indices. By gravitons we mean massless spin-2 particles, though this notion still has wiggle room to permit different implementations. Here we are considering two types of graviton fields: the spin-2 and Fierz-Pauli fields. When the Minkowski metric η_{ab} is used, it is understood as written in a generic coordinate system. The d’Alembert operator in the flat metric is \square . Similarly, the covariant derivative ∇ is always related to the flat metric, while ∇' corresponds to a curved metric. Curvature-related quantities will be defined by following Misner-Thorne-Wheeler’s convention [11]. This paper was initially motivated by the paper of Padmanabhan [8] and, thus, we have decided to partially maintain his notation to facilitate the translation of the results. Our intention throughout the paper has been to use an “aseptic” writing style that avoids contamination from geometric notions motivated by previous knowledge of general relativity. In this way, one can see more clearly the different steps which are necessary to obtain general relativity (or unimodular gravity) as solutions to the self-coupling problem.

II. FREE GRAVITON FIELDS

The unitary representations of the Poincaré group as first classified by Wigner are determined by the value of the mass m and the eigenvalues of the so-called little group [12,13]. For a particle with mass $m \neq 0$ the little group is $SO(3)$, so the corresponding label is the angular momentum j and one has $2j + 1$ states in each representation, corresponding to polarizations which range from $\sigma = -j$ to $\sigma = +j$ jumping in units. However, for a massless particle the little group is $ISO(2)$ (the two-dimensional Euclidean group) and only the states with polarizations $\sigma = \pm j$ are left. This means that massless particles with integer spin carry

only two independent degrees of freedom. As linear representation space, one would like to construct a tensor-field space using exclusively these degrees of freedom, but it is in this construction where gauge invariance appears inevitably intertwined with Lorentz invariance. In the following, we are going to work in a covariant fashion with respect to changes of coordinates in Minkowski spacetime. This will be useful to distinguish between covariance and internal gauge invariance in the resulting nonlinear theory.

The natural way to define such a spacetime tensor field is to construct an object \mathfrak{h}^{ab} exclusively made of the two physical polarizations, which can be done for example in momentum space [14]. The problem is that Lorentz transformations do not leave this space invariant, not even the transformations that belong to the little group. Assuming a tensorial character for \mathfrak{h}^{ab} we have that, under an infinitesimal Lorentz transformation whose generators are ξ_ω^a , it transforms as

$$h_\omega^{ab} = \mathfrak{h}^{ab} + \xi_\omega^c \nabla_c \mathfrak{h}^{ab} + \mathfrak{h}^{ac} \nabla_c \xi_\omega^b + \mathfrak{h}^{bc} \nabla_c \xi_\omega^a. \quad (1)$$

The Lorentz generators ξ_ω^a verify the conditions

$$\nabla_a \xi_\omega^a = 0, \quad \square \xi_\omega^a = 0. \quad (2)$$

The problem is that the last terms in this transformation law do not belong to the linear space of objects of \mathfrak{h}^{ab} type [15], i.e.,

$$\xi_\omega^c \nabla_c \mathfrak{h}^{ab} + \mathfrak{h}^{ac} \nabla_c \xi_\omega^b + \mathfrak{h}^{bc} \nabla_c \xi_\omega^a \neq \mathfrak{h}^{ab}. \quad (3)$$

Thus, the transformed object h_ω^{ab} does not belong to the linear representation space we started with (the same problem appears in electrodynamics, where the gauge fixing condition $\epsilon^0 = 0$ on the polarization vector e^μ is not Lorentz invariant). One can prove instead that the terms driving one outside the representation space are of the form

$$h_\omega^{ab} - \mathfrak{h}^{ab} = \eta^{ac} \nabla_c \xi_\omega^b + \eta^{bc} \nabla_c \xi_\omega^a. \quad (4)$$

To circumvent the representation problem we have two options: instead of using a tensor space as representation space, one could define \mathfrak{h}^{ab} as a nontensorial object; the other possibility is to maintain a subsidiary tensor-field space but consider as representation space not its individual elements but equivalence classes of them related by gauge transformations. We shall proceed using this second approach.

The minimal realization of the gauge approach is to take as representation space tensorial objects h^{ab} such that they are traceless and transverse,

$$\eta_{ab} h^{ab} = 0, \quad \nabla_b h^{ab} = 0, \quad (5)$$

and with equation of motion

$$\square h^{ab} = 0. \quad (6)$$

Moreover, any h^{ab} and h'^{ab} related by an internal gauge transformation,

$$h'^{ab}_\xi = h^{ab} + \eta^{ac} \nabla_c \xi^b + \eta^{bc} \nabla_c \xi^a, \quad (7)$$

will represent the same physical configuration, with generators verifying the conditions

$$\nabla_a \xi^a = 0, \quad \square \xi^a = 0. \quad (8)$$

Notice that the internal gauge transformation (7) and the external transformation associated with a general change of coordinates with generators $\tilde{\xi}^a$,

$$h'^{ab}_{\tilde{\xi}} = h^{ab} + \tilde{\xi}^c \nabla_c h^{ab} + h^{ac} \nabla_c \tilde{\xi}^b + h^{bc} \nabla_c \tilde{\xi}^a, \quad (9)$$

are completely different from each other: the space of generators is different and so is their implementation in the symmetry transformation. Moreover, the last transformation affects the coordinates and the rest of the fields.

It is easy to check that the traceless and transverse conditions are preserved by these internal gauge transformations. These constraints in the definition of the field h^{ab} can be thought as the elimination of the scalar and vector representations of the Poincaré group (see Appendix I in [16]). A detailed analysis shows that one can always find a vector ξ^a such that the states with helicities $\sigma = 0, \pm 1$ are gauged away or, in other words, the corresponding components h^{00} and h^{0i} , $i = 1, 2, 3$ are set to zero while the remaining components are constrained so that there are two independent degrees of freedom. Another common choice to show this is the light-cone gauge [17].

Notice that there are only two contractions of $(\nabla h)^2$ with metric objects (η_{ab} , η^{ab} and δ_b^a) which are not zero by the traceless and transverse conditions (5). By virtue of this, the Lagrangian density should have the form

$$\mathcal{L}_{G,0} := c_1 \eta^{ai} \eta_{bj} \eta_{ck} \nabla_a h^{bc} \nabla_i h^{jk} + c_2 \delta_k^a \delta_c^i \eta_{bj} \nabla_a h^{bc} \nabla_i h^{jk}. \quad (10)$$

Here c_1 and c_2 are real constants. The second term is equivalent to a total divergence because of the transverse condition in (5), so it does not affect the form of the equations of motion. However, its presence can affect the definition of the source of the self-interacting equations as we will see. We can set the normalization to $c_1 = -1/4$ and introduce a bit of notation to conveniently write the free action as

$$\mathcal{A}_{G,0} := \frac{1}{4} \int d\mathcal{V}_\eta M_{\mathfrak{s}}^{ai}{}_{bcjk}(\eta) \nabla_a h^{bc} \nabla_i h^{jk}, \quad (11)$$

where $d\mathcal{V}_\eta := d^4x \sqrt{-\eta}$ is the Minkowski volume element and the Lorentz tensor $M_{\mathfrak{s}}^{ai}{}_{bcjk}(\eta)$ is given by

$$M_{\mathfrak{s}}^{ai}{}_{bcjk}(\eta) := \mathfrak{s}[\eta_{b(j} \delta_k^a \delta_c^i + \eta_{c(j} \delta_k^a \delta_b^i] - \eta^{ai} \eta_{b(j} \eta_{k)c}. \quad (12)$$

The tensorial quantity $M_{\mathfrak{z}}^{ai}{}_{bcjk}(\eta)$ is symmetric under $b \leftrightarrow c$, $j \leftrightarrow k$ and $(a, b, c) \leftrightarrow (i, j, k)$, as one can check from its definition. When used in the action we do not need to worry about these symmetries because it is contracted with an object, $(\nabla h)^2$, which already has these symmetries. However, to solve the iterative equations of the self-coupling problem it will be necessary to use its symmetric form as it appears in (12). The parameter \mathfrak{z} , directly proportional to c_2 , controls the surface terms we are considering in the free action. One can easily check that this action is invariant up to a surface term under the gauge transformations (7) with generators satisfying (8). In fact, the case $\mathfrak{z} = 1$ is special in the sense that one could drop the second condition in (8) and these transformations are still a symmetry. For this reason we will always assume that this is the case i.e., only the first condition in (8) applies when considering $\mathfrak{z} = 1$, thus recovering the minimal theory of gravitons which was considered in [18]. The reader will notice that, through the calculations in Sec. IV, we always keep using the object h^{ab} and never its covariant counterpart $h_{ab} := \eta_{ac}\eta_{bd}h^{cd}$. This simplifies some steps which involve taking variational derivatives with respect to an auxiliary metric γ_{ab} after the replacement $\eta_{ab} \rightarrow \gamma_{ab}$ in the action. In the following, this traceless field will be called spin-2 field.

An alternative way of constructing a spacetime field h^{ab} is to drop the traceless and transverse conditions in (5) while enlarging the gauge symmetry. This is the well-known Fierz-Pauli theory [19], in which the fundamental field is just a symmetric Lorentz tensor. We will call it Fierz-Pauli field. We are not going to enter into details here about the construction of the Fierz-Pauli action as the procedure is equivalent (but a little more involved) to the one we have followed for the spin-2 field. The details can be found e.g., in [8], and we will mention part of them in Sec. VI. Notice that the transverse and traceless conditions can be imposed on the Fierz-Pauli field only within the space of solutions of the free theory. That is, the so-called transverse-traceless gauge can be applied only for fields h^{ab} verifying the condition

$$\nabla_a \nabla_b h^{ab} = \eta_{ab} \square h^{ab}, \quad (13)$$

which is precisely the trace of the Fierz-Pauli dynamical equations [20].

To develop the self-coupling scheme, we are going to treat these theories as classical field theories. Our conclusions are applicable then to the long wavelength limit of theories in which a graviton propagates over Minkowski space in interaction with matter, independently of the ultraviolet completion of the theory.¹ From the point of view of the

¹In particular, the notion of a Minkowski preferred background could be emergent in the sense of being applicable only below some characteristic energy scale, instead of a fundamental structure present in all regimes [21].

classical equations of motion, which are what we are interested in, surface terms in the resulting action of the self-interacting theory are irrelevant. Notice that, if there is a regime in the theory in which gravity functions classically but the matter fields behave quantum mechanically (semiclassical gravity), it is reasonable to expect that our conclusions would also apply to it as the self-coupling only occurs in the gravitational sector, still described by c -numbers.

III. COUPLING TO MATTER

In this section we address the question: could we make an interacting theory of the spin-2 and matter fields? We have included this review section to facilitate the reading of the paper, but its contents are well known in the literature (see [20] and references therein).

To couple a matter field to the spin-2 field we need to define a quantity \tilde{T}_{ab} that is symmetric, traceless and transverse on solutions. Then, we could write

$$\square h_{ab} = \lambda \tilde{T}_{ab}. \quad (14)$$

A natural consistency condition is to impose that this equation can be obtained from an action. In fact, it can be obtained as the corresponding Euler-Lagrange equation with respect to restricted variations of h^{ab} (such that $\eta_{ab}\delta h^{ab} = 0$) as long as we add to the Lagrangian density a term

$$\Delta \mathcal{L} := \lambda h^{ab} T_{ab}, \quad (15)$$

with T_{ab} symmetric, transverse on solutions, and with constant trace. An object with these characteristics is precisely the stress-energy momentum tensor, where the transverse condition amounts to its conservation. We comment on the condition of constancy of the trace at the end of the section; for now let us just require conservation of T_{ab} .

One immediately realizes that one cannot use the stress-energy tensor of the free matter theory: for consistency one must use the total stress-energy tensor of the interacting theory. Had we started tentatively by adding a term $\lambda h^{ab} T_{ab}^M$ to the Lagrangian density of the matter sector, with T_{ab}^M the free matter stress-energy tensor, this very term would have changed the matter stress-energy tensor making it necessary to add new energetic contributions to the Lagrangian density. This iterative process is the one we are going to follow in the next section.

This is a general property of the coupling with matter: as the coupling is done through the stress-energy tensor, the transverse condition for the spin-2 field implies that the total matter traceless stress-energy tensor should be divergenceless.² However, this would not be the case when interaction is switched on, as the matter fields no longer

²When the transverse condition is relaxed, it is the resulting gauge invariance of the theory the responsible for this feature [20].

behave as an isolated system, being the energy transferred between them and the spin-2 field. The natural way to remedy this is to realize that the spin-2 field must also act as a source of itself (the charge/source of the graviton field is the energy and gravitons should possess energy), which leads us to the issue of the spin-2 field self-coupling. Therefore, the iterative procedure has to act also in the spin-2 sector.

An important problem shows up when thinking about the stress-energy tensor of the spin-2 sector: there is no way of constructing a nontrivial conserved stress-energy tensor for the spin-2 field that is invariant under the gauge transformations (7) [22,23]. By nontrivial we mean that it is not exactly zero for any solution. An indirect way of realizing it could be the Weinberg-Witten theorem [13,24], which explicitly forbids this possibility. Thus, one cannot associate a local notion of energy with the physical configurations in the free theory.³ One can live with this fact if the theory is noninteracting, so that there is no operational way to define what it is energy and momentum. Within an interaction scheme this is untenable.

As mentioned before, strictly speaking the consistency of a complete nonlinear spin-2 theory will need that the transverse condition applies to the traceless object \tilde{T}_{ab} , independently of that of T_{ab} . When the trace of the stress-energy tensor is a constant the conservation of T_{ab} implies the conservation of \tilde{T}_{ab} . In this respect, notice that the trace of the stress-energy tensor of a single linear matter sector (e.g., a single scalar field) is always a constant as there are no sources that can cause an inhomogeneity in the system, and hence in the trace. When putting together the spin-2 and matter sectors with all its nonlinear interactions, we do not know *a priori* what could happen. All in all one is left with the expectation that when applying a consistent self-interacting scheme some meaningful result would show up for all the standing problems. We will come back to this issue after finding the solutions to the self-coupling problem.

IV. SELF-COUPLING

The first question to answer when considering a self-interacting scheme is: what is the object to which we are going to couple the spin-2 field in a first stage? A natural candidate to consider is the canonical stress-energy tensor, that is, a conserved quantity of any field theory which is Poincaré invariant, associated with invariance under translations:

$$\Theta^a_b := \mathcal{L}_0 \delta_b^a - \frac{\delta \mathcal{L}_0}{\delta(\nabla_a \psi^\mu)} \nabla_b \psi^\mu, \quad \nabla_a \Theta^a_b = 0. \quad (16)$$

³Something equivalent happens in non-Abelian Yang-Mills theory: one cannot find a Lorentz covariant conserved current which is at the same time gauge invariant.

Here \mathcal{L}_0 is the free Lagrangian density of both spin-2 and matter fields, collectively denoted by ψ^μ . If we manage to use this quantity as the right-hand side of our equations of motion for h^{ab} , the theory will naturally verify the condition obtained by Weinberg [25,26] as a necessary one if we want to have a Lorentz-invariant theory: the coupling between the spin-2 field to matter and to itself must be governed by the same coupling constant.

However, direct use of this quantity is not possible: in general, the fully covariant or contravariant counterparts of (16) are not symmetric. But we can exploit the ambiguity in the addition of identically conserved tensors, the so-called Belinfante-Rosenfeld terms, to obtain a symmetric stress-energy tensor which leads to the same conserved quantities. This symmetric stress-energy tensor is not unique: one can still add identically conserved tensors keeping the symmetric character. All the manipulations that follow can be performed by directly using the symmetrized versions of the canonical stress-energy tensor. Therefore, these manipulations in no way involve any curved spacetime notion. However, as shown by Belinfante and Rosenfeld [27,28], these symmetric stress-energy tensors can be equivalently obtained by the simple Hilbert prescription,

$$T_{ab} := - \lim_{\gamma \rightarrow \eta} \frac{1}{\sqrt{-\gamma}} \frac{\delta \mathcal{A}_0}{\delta \gamma^{ab}}, \quad (17)$$

where the flat metric η_{ab} in \mathcal{A}_0 has been replaced by an auxiliary (generally curved) metric γ_{ab} , being γ^{ab} its inverse. Recall that the two steps one needs to follow are: write the action in curvilinear coordinates in flat space, and then generalize it to curved space. It is in this second step where the ambiguities show up in Hilbert's prescription. In practice, the ambiguities in the stress-energy tensor appear now as the addition of nonminimal couplings of the physical fields to the auxiliary metric γ_{ab} , and they are added to the one-parameter family of surface terms we are considering in the free action (11). In fact, as we will see later these nonminimal couplings can be understood as surface terms in the original free action, though different from the one-parameter family we have been considering up to now. We will show that these different choices of stress-energy tensor as the source lead to different solutions to the self-coupling problem. Let us stress again that here we use Hilbert's prescription as a mere calculational device and insist that no curved spacetime notion is used throughout the calculations.

Now that we have discussed the relevant properties of the stress-energy tensor, we would like to derive the self-interacting equations of motion from an action. The coupling constant is denoted by λ . This can be done if we add a term $\lambda \mathcal{A}_1$ of order $\mathcal{O}(\lambda)$ in the action, such that

$$\frac{\delta \mathcal{A}_1}{\delta h^{ab}} = \lim_{\gamma \rightarrow \eta} \frac{\delta \mathcal{A}_0}{\delta \gamma^{ab}}. \quad (18)$$

Here \mathcal{A}_0 has two terms: the free action for the spin-2 field given by (11) and the matter content one wants to consider, $\mathcal{A}_0 = \mathcal{A}_{G,0} + \mathcal{A}_{M,0}$.

As noticed by Gupta [3], this additional term of order $\mathcal{O}(\lambda)$ in the action would modify the definition of the source by a term of order $\mathcal{O}(\lambda^2)$, which implies that we need to contemplate a term $\lambda^2 \mathcal{A}_2$ in the action. This is the iterative procedure we would want to solve for. It will generate an action of the form⁴

$$\mathcal{A} = \sum_{n=0}^{\infty} \lambda^n \mathcal{A}_n, \quad (19)$$

where the set of partial actions $\{\mathcal{A}_n\}_{n=1}^{\infty}$ must verify the iterative equations

$$\frac{\delta \mathcal{A}_n}{\delta h^{ab}} = \lim_{\gamma \rightarrow \eta} \frac{\delta \mathcal{A}_{n-1}}{\delta \gamma^{ab}}, \quad n \geq 1. \quad (20)$$

In more detail, let us write the resulting action of the self-coupling procedure as

$$\mathcal{A} := \mathcal{A}_0 + \mathcal{A}_I, \quad (21)$$

where the free part \mathcal{A}_0 is already defined and \mathcal{A}_I is the self-interacting part we are going to solve for. Given this action, one would be able to obtain its stress-energy tensor. \mathcal{A}_I is then fixed by the requirement of leading to this stress-energy tensor as the source of the equations of motion:

$$\frac{\delta \mathcal{A}_I}{\delta h^{ab}} = \lambda \lim_{\gamma \rightarrow \eta} \frac{\delta(\mathcal{A}_0 + \mathcal{A}_I)}{\delta \gamma^{ab}}. \quad (22)$$

One just needs to expand $\mathcal{A}_I = \sum_{n=1}^{\infty} \lambda^n \mathcal{A}_n$ and compare different orders in the coupling constant λ to obtain the set of iterative equations (20).

Recall that we started with traceless equations of motion. This means that, to keep the same number of equations of motion in the iterative procedure, the field h^{ab} must be constrained in some way. One option is to maintain the traceless condition with respect to the original Minkowski metric. In any case, the set of equations (20) is general enough to permit the imposition of this condition, as well as other possible scalar constraints over the field h^{ab} , after finding its solutions. So we postpone this discussion to Sec. V, although we will keep in mind the existence of this constraint which, at least at the lowest order, must be equivalent to the traceless condition.

⁴Notice that there is no reason to expect, in principle, this series to be infinite. There are two examples in the literature of this kind of series: the first one is the trivial one, in the sense that one does not consider self-interactions of the graviton field (see paragraph below). This series is infinite. The only example of a self-interacting series is the one constructed by Deser [6] which is finite, with only $\mathcal{A}_1 \neq 0$. In this paper we are going to consider always infinite series.

The integration of the iterative equations for the matter part is straightforward, as the corresponding part of the stress-energy tensor does not contain the spin-2 field explicitly. That is, the matter part of the right-hand side of (18) is independent of h^{ab} at the lowest order, linear at first order, and so on, making the integration of this part of the equation trivial. The resulting action is obtained as a Taylor series which can be summed. The formal result of this sum is the free matter action expressed in terms of a curved metric, $\mathcal{A}_{M,0}[g]$ with $g^{ab} := \eta^{ab} + \lambda h^{ab}$ [6,8]. Notice that nonminimal couplings to matter are not ruled out by any consistency condition, so minimal coupling to the physical metric in the resulting matter action is not a necessity in this approach. It is the gravitational self-interacting part of the iterative procedure which has to be handled carefully, and this is the part we are going to work with in the rest of the text.

Concerning this self-interacting part, one could expect that the resulting theory exhibits a nonlinear deformation of the original gauge invariance, which is broken at each stage of the iterative procedure. The search for this symmetry has been a commonly used route to argue that general relativity should be the only consistent self-interacting theory of the Fierz-Pauli field, as the only nonlinear deformation of such linear symmetry is diffeomorphism invariance [16,29,30] (see also the related discussion in [20]). However, here we would like to understand the interplay between the preservation of this symmetry and the iterative self-coupling procedure, instead of taking its existence as an assumption from the beginning.

In the rest of the section we are going to solve the iterative equations (20) step by step. First we are going to solve these equations in the simplest case in which there are no non-minimal couplings. We show then that there is a unique solution of these iterative equations, which corresponds to a selection of a certain value of the parameter \mathfrak{z} in (11). Then we devote the next subsection to understand the role of nonminimal couplings. Their inclusion will permit us to obtain the general solution to the self-coupling problem.

A. Explicit integration and summation of the series

In this section we are going to see how to manage the infinite set of iterative equations (20) for the spin-2 field. Let us start with the first-order iterative equation (18). To do that, we are going to evaluate the right-hand side of this equation and then integrate its functional form to obtain the corresponding left-hand side.

The first step is to apply Hilbert's prescription to obtain the source of the equations of motion. To do that we have to extend the action (11) to a general curved metric.⁵ Adopting a minimal prescription we can write

⁵It is in this step where the ambiguities in the addition of nonminimal coupling terms can arise. We will deal with this ambiguity in the following section, thus making here the simplest choice.

$$\mathcal{A}_0[\gamma] = \frac{1}{4} \int d\mathcal{V}_\gamma M_{\mathfrak{s}}^{ai}{}_{bcjk}(\gamma) \nabla'_a h^{bc} \nabla'_i h^{jk}, \quad (23)$$

where we have dropped the G subindex. Here ∇' is the covariant derivative with respect to γ_{ab} and $d\mathcal{V}_\gamma := d^4x \sqrt{-\gamma}$ the corresponding volume element. Notice that we have changed the metric in the argument of the tensor $M_{\mathfrak{s}}^{ai}{}_{bcjk}(\eta)$ defined in (12).

We obtain the stress-energy tensor by performing variations on this metric, and then taking the limit back to flat space. Under such a variation, the action (23) changes as

$$\begin{aligned} \delta\mathcal{A}_0[\gamma] &= \frac{1}{4} \int d^4x \delta[\sqrt{-\gamma} M_{\mathfrak{s}}^{ai}{}_{bcjk}(\gamma)] \nabla'_a h^{bc} \nabla'_i h^{jk} \\ &+ \frac{1}{2} \int d^4x \sqrt{-\gamma} M_{\mathfrak{s}}^{ai}{}_{bcjk}(\gamma) \delta[\nabla'_a h^{bc}] \nabla'_i h^{jk}. \end{aligned} \quad (24)$$

The first term gives two contributions, one coming from the variation of the determinant and the other from the variation of $M_{\mathfrak{s}}^{ai}{}_{bcjk}(\gamma)$. There are two possible ways of dealing with the former. The first one is to notice that the first-order equation must be traceless so the corresponding term is not going to contribute. This observation can be extended to all orders with the following recipe: do not change the measure in the partial actions \mathcal{A}_n when writing them in terms of the auxiliary metric γ_{ab} . Although a departure from Hilbert's prescription, this alternative procedure leads to a sensible source to be used in the self-coupling procedure when the constraints on the field h^{ab} are taken into account. We shall follow this approach in this section. A second option is to proceed with no previous knowledge of the restrictions on h^{ab} and integrate the contribution coming from the variation of the determinant. The iterative equations (20) are linear, so we only need to add the corresponding contribution obtained this way to the result of the calculations of this section. We will show in the next section the result of this procedure. Of course, this is only an operational choice which does not affect the physical results at the end of the day, when the constraints on the field h^{ab} are considered.

Let us now deal with the second term of (24). There we have the difference of two Levi-Civita connections associated with γ_{ab} and $\gamma_{ab} + \delta\gamma_{ab}$, respectively. This difference is characterized by the tensor

$$C'^b{}_{ad} := \frac{1}{2} (\gamma^{be} + \delta\gamma^{be}) \nabla'_\mu (\gamma_{\nu\rho} + \delta\gamma_{\nu\rho}) D^{\mu\nu\rho}{}_{aed}, \quad (25)$$

where

$$D^{\mu\nu\rho}{}_{aed} := \delta_a^\mu \delta_d^\nu \delta_e^\rho + \delta_a^\mu \delta_a^\nu \delta_e^\rho - \delta_e^\mu \delta_a^\nu \delta_d^\rho, \quad (26)$$

(see e.g., [31]). Then one can see that the variation $\delta[\nabla'_a h^{bc}]$ is given, at first order, by

$$\delta C'^b{}_{ad} h^{cd} = \gamma^{e(b} h^{c)d} \nabla'_\mu \delta\gamma_{\nu\rho} D^{\mu\nu\rho}{}_{aed}. \quad (27)$$

The notation $\delta C'^b{}_{ad}$ here means that we only take the terms in (25) which are linear in the variations $\delta\gamma_{ab}$. If we integrate by parts, the contribution of these terms equals

$$\begin{aligned} & - \frac{1}{2} \int d^4x \sqrt{-\gamma} \delta\gamma_{\nu\rho} \gamma^{e(b} D^{\mu\nu\rho}{}_{aed} M_{\mathfrak{s}}^{ai}{}_{bcjk} \nabla'_\mu [h^{c)d} \nabla'_i h^{jk}] \\ &= \frac{1}{2} \int d^4x \sqrt{-\gamma} \delta\gamma^{\rho q} \gamma_{\rho\nu} \gamma_{q\rho} \gamma^{de} D^{a\nu\rho}{}_{\mu e(b} M_{\mathfrak{s}}^{\mu i}{}_{c)djk} \\ &\quad \times \nabla'_a (h^{bc} \nabla'_i h^{jk}). \end{aligned} \quad (28)$$

The corresponding source then takes the form

$$\begin{aligned} T_{\rho q} &:= - \frac{1}{4} \frac{\delta M_{\mathfrak{s}}^{ai}{}_{bcjk}}{\delta\gamma^{\rho q}} \Big|_{\gamma \rightarrow \eta} \nabla'_a h^{bc} \nabla'_i h^{jk} \\ &\quad - \frac{1}{2} \eta_{\rho\nu} \eta_{q\rho} \eta^{de} D^{a\nu\rho}{}_{\mu e(b} M_{\mathfrak{s}}^{\mu i}{}_{c)djk} \nabla'_a (h^{bc} \nabla'_i h^{jk}). \end{aligned} \quad (29)$$

Notice that this expression contains second derivatives of the spin-2 field. It is important to notice also that it naturally splits into two kinds of terms, proportional to $(\nabla h)^2$ and $\nabla(h\nabla h)$, respectively. As it stands, it is symmetric under the exchanges $p \leftrightarrow q$ and $b \leftrightarrow c$.

The objective now is to find a term in the action $\lambda\mathcal{A}_1$ whose variation with respect to h^{ab} gives the desired source term. The most general expression which contains no more than two derivatives of the spin-2 field can be always written as

$$\frac{1}{4} \int d\mathcal{V}_\eta N^{ai}{}_{bcjkpq}(\eta) h^{pq} \nabla'_a h^{bc} \nabla'_i h^{jk}. \quad (30)$$

Then taking the functional derivative with respect to h^{pq} we obtain

$$\begin{aligned} & \frac{1}{4} \int d\mathcal{V}_\eta [N^{ai}{}_{bcjkpq}(\eta) \nabla'_a h^{bc} \nabla'_i h^{jk} \\ &\quad - 2N^{ai}{}_{pqjkb}(\eta) \nabla'_a (h^{bc} \nabla'_i h^{jk})] \delta h^{pq}. \end{aligned} \quad (31)$$

We get two equations coming from the comparison of the coefficients accompanying the two independent combinations $(\nabla h)^2$ and $\nabla(h\nabla h)$ in both Eqs. (29) and (31):

$$N^{ai}{}_{bcjkpq}(\eta) = \frac{\delta M_{\mathfrak{s}}^{ai}{}_{bcjk}(\gamma)}{\delta\gamma^{\rho q}} \Big|_{\gamma \rightarrow \eta}, \quad (32)$$

and

$$-N^{ai}{}_{bcjkpq}(\eta) = \eta_{\rho\nu} \eta_{q\rho} \eta^{de} D^{a\nu\rho}{}_{\mu e(b} M_{\mathfrak{s}}^{\mu i}{}_{c)djk}(\eta). \quad (33)$$

The first equation provides the form of the first-order action (30). The second equation then becomes a consistency condition that must be satisfied for the whole procedure to be well defined:

$$-\left. \frac{\delta M_{\mathfrak{z}}^{ai}{}_{pqjk}(\gamma)}{\delta \gamma^{bc}} \right|_{\gamma \rightarrow \eta} = \eta_{p\nu} \eta_{q\rho} \eta^{de} D^{a\nu\rho}{}_{\mu e(b} M_{\mathfrak{z}}^{\mu i}{}_{c) djk}(\eta). \quad (34)$$

It is this equation which imposes restrictions to the solutions of the iterative equations which, in fact, select $\mathfrak{z} = 1$. To obtain this condition on \mathfrak{z} , let us notice that the right-hand side of (34) can be written as

$$\eta_{pb} M_{\mathfrak{z}}^{ai}{}_{qcjk}(\eta) + \delta_b^a \eta_{p\mu} M_{\mathfrak{z}}^{\mu i}{}_{qcjk}(\eta) - \eta_{pb} \eta_{q\mu} \eta^{ad} M_{\mathfrak{z}}^{\mu i}{}_{dcjk}(\eta), \quad (35)$$

where we must impose a symmetrization under the exchange of indices $p \leftrightarrow q$ and $b \leftrightarrow c$. It is useful to write this expression explicitly by using (12),

$$\begin{aligned} & \mathfrak{z} \eta_{pb} [\eta_{q(j} \delta_k^a) \delta_c^i + \eta_{c(j} \delta_k^a) \delta_q^i] - \eta_{pb} \eta^{ai} \eta_{q(j} \eta_{k)c} \\ & + \mathfrak{z} \delta_b^a [\eta_{p(j} \eta_{k)q} \delta_c^i + \eta_{p(j} \eta_{k)c} \delta_q^i] - \delta_b^a \eta_{q(j} \eta_{k)c} \delta_p^i \\ & - \mathfrak{z} [\eta^{ai} \eta_{pb} \eta_{q(j} \eta_{k)c} + \eta_{pb} \eta_{q(j} \delta_k^a) \delta_c^i] + \delta_q^i \delta_{(j} \eta_{p)b} \eta_{k)c}, \end{aligned} \quad (36)$$

and symmetrize this equation with respect to $p \leftrightarrow q$, so it can be simplified to

$$\begin{aligned} & \mathfrak{z} [\eta_{pb} \eta_{c(j} \delta_k^a) \delta_q^i + \eta_{qb} \eta_{c(j} \delta_k^a) \delta_p^i] + \delta_b^a \delta_c^i \eta_{p(j} \eta_{k)q} \\ & - \frac{\mathfrak{z} + 1}{2} \eta^{ai} [\eta_{pb} \eta_{c(j} \eta_{k)q} + \eta_{qb} \eta_{p(j} \eta_{k)c}] \\ & + \frac{\mathfrak{z} - 1}{2} \delta_b^a [\eta_{p(j} \eta_{k)c} \delta_q^i + \eta_{q(j} \eta_{k)c} \delta_p^i]. \end{aligned} \quad (37)$$

This equation must be compared with the left-hand side of (34) i.e., with

$$\begin{aligned} -\left. \frac{\delta M_{\mathfrak{z}}^{ai}{}_{pqjk}(\gamma)}{\delta \gamma^{bc}} \right|_{\gamma \rightarrow \eta} & = \mathfrak{z} [\eta_{pb} \eta_{c(j} \delta_k^a) \delta_q^i + \eta_{qb} \eta_{c(j} \delta_k^a) \delta_p^i] \\ & + \delta_b^a \delta_c^i \eta_{p(j} \eta_{k)q} \\ & - \eta^{ai} [\eta_{pb} \eta_{c(j} \eta_{k)q} + \eta_{qb} \eta_{p(j} \eta_{k)c}], \end{aligned} \quad (38)$$

which must be still symmetrized under the exchange $b \leftrightarrow c$. A direct comparison of these equations tells us that the only solution of (34) is given by $\mathfrak{z} = 1$.

In this way, we have shown how to integrate the first-order iterative equation (18). The result is

$$\begin{aligned} & \mathcal{A}_0 + \lambda \mathcal{A}_1 + \mathcal{O}(\lambda^2) \\ & = \frac{1}{4} \int d\mathcal{V}_\eta M_{1bcjk}^{ai}(\eta + \lambda h) \nabla_a h^{bc} \nabla_i h^{jk} + \mathcal{O}(\lambda^2). \end{aligned} \quad (39)$$

Now that we have worked out the first order in detail, the objective is to show that the result which can be anticipated

from this order is in fact the correct result. That is, that the total action is

$$\mathcal{A} = \frac{1}{4} \int d\mathcal{V}_\eta M_{1bcjk}^{ai}(\eta + \lambda h) \nabla_a h^{bc} \nabla_i h^{jk}. \quad (40)$$

So we decompose this ansatz (40) in partial actions, $\mathcal{A} = \sum_{n=0}^{\infty} \lambda^n \mathcal{A}_n$, with

$$\mathcal{A}_n = \frac{1}{4n!} \int d\mathcal{V}_\eta \left. \frac{\delta^n M_{1bcjk}^{ai}(\gamma)}{\delta \gamma^{pq} \delta \gamma^{st} \dots} \right|_{\gamma \rightarrow \eta} \nabla_a h^{bc} \nabla_i h^{jk} h^{pq} h^{st} \dots, \quad (41)$$

and apply the iterative equations (20) to this sequence, finding the consistency condition:

$$\begin{aligned} & n \eta_{p\nu} \eta_{q\rho} \eta^{de} D^{a\nu\rho}{}_{\mu e(b} \left. \frac{\delta^{n-1} M_{1c) djk}^{\mu i}(\gamma)}{\delta \gamma^{st} \dots} \right|_{\gamma \rightarrow \eta} \\ & = - \left. \frac{\delta^n M_{1pqjk}^{ai}(\gamma)}{\delta \gamma^{bc} \delta \gamma^{st} \dots} \right|_{\gamma \rightarrow \eta}. \end{aligned} \quad (42)$$

Notice the symmetrization of the pair (b, c) . To work better with this expression, we can avoid at first to evaluate it in the limit $\gamma \rightarrow \eta$, working thus with the equation

$$n \gamma_{p\nu} \gamma_{q\rho} \gamma^{de} D^{a\nu\rho}{}_{\mu e(b} \frac{\delta^{n-1} M_{1c) djk}^{\mu i}(\gamma)}{\delta \gamma^{st} \dots} = - \frac{\delta^n M_{1pqjk}^{ai}(\gamma)}{\delta \gamma^{bc} \delta \gamma^{st} \dots}, \quad (43)$$

which can be viewed as a differential equation with an initial condition imposed in flat space. In fact, if we drop the indices we can write it schematically as

$$n \Theta(\gamma) \frac{\partial^{n-1} M(\gamma)}{\partial \gamma^{n-1}} = - \frac{\partial^n M(\gamma)}{\partial \gamma^n}, \quad (44)$$

with $\Theta \sim (\gamma)^{-1}$. Up to now, we have shown that (34) is valid, which in this simplified notation becomes

$$\Theta(\gamma) M(\gamma) = - \frac{\partial M(\gamma)}{\partial \gamma}. \quad (45)$$

Thus to show by induction that (40) is in fact the solution to the iterative problem we only need, as we have already proved that it holds for $n = 1$, to show that $\Theta(\gamma)$ verifies the differential equation,

$$\frac{\partial \Theta(\gamma)}{\partial \gamma} = -\Theta^2(\gamma), \quad (46)$$

as it is indeed the case. Coming back to the full equations, one has

$$\begin{aligned}
 & -\frac{\delta^{n+1}M_1^{ai}{}_{pqjk}(\gamma)}{\delta\gamma^{uv}\delta\gamma^{bc}\delta\gamma^{st}\dots} \\
 & = n\gamma_{pv}\gamma_{qp}\gamma^{de}D^{av\rho}{}_{\mu e(b)}\frac{\delta^n M_1^{\mu i}{}_{c)dk}(\gamma)}{\delta\gamma^{uv}\delta\gamma^{st}\dots} \\
 & \quad + nD^{av\rho}{}_{\mu e(b)}\frac{\delta^{n-1}M_1^{\mu i}{}_{c)dk}(\gamma)}{\delta\gamma^{st}\dots}\frac{\delta}{\delta\gamma^{uv}}(\gamma_{pv}\gamma_{qp}\gamma^{de}) \\
 & \quad + \{(u, v) \leftrightarrow (b, c)\}. \tag{47}
 \end{aligned}$$

Then the consistency condition with the induction can be read as

$$\begin{aligned}
 & D^{av\rho}{}_{\mu e(b)}\frac{\delta^{n-1}M_1^{\mu i}{}_{c)dk}(\gamma)}{\delta\gamma^{st}\dots}\frac{\delta}{\delta\gamma^{uv}}(\gamma_{pv}\gamma_{qp}\gamma^{de}) \\
 & \quad + D^{av\rho}{}_{\mu e(u)}\frac{\delta^{n-1}M_1^{\mu i}{}_{v)dk}(\gamma)}{\delta\gamma^{st}\dots}\frac{\delta}{\delta\gamma^{bc}}(\gamma_{pv}\gamma_{qp}\gamma^{de}) \\
 & = \frac{1}{n}\gamma_{pv}\gamma_{qp}\gamma^{de}D^{av\rho}{}_{\mu e(b)}\frac{\delta^n M_1^{\mu i}{}_{c)dk}(\gamma)}{\delta\gamma^{uv}\delta\gamma^{st}\dots} \\
 & \quad + \frac{1}{n}\gamma_{pv}\gamma_{qp}\gamma^{de}D^{av\rho}{}_{\mu e(u)}\frac{\delta^n M_1^{\mu i}{}_{v)dk}(\gamma)}{\delta\gamma^{bc}\delta\gamma^{st}\dots}. \tag{48}
 \end{aligned}$$

Because of the symmetrization, we can take only one of the terms in each side of the last equation, thus obtaining the equation

$$\begin{aligned}
 & D^{av\rho}{}_{\mu e(b)}\frac{\delta^{n-1}M_1^{\mu i}{}_{c)dk}(\gamma)}{\delta\gamma^{st}\dots}\frac{\delta}{\delta\gamma^{uv}}(\gamma_{pv}\gamma_{qp}\gamma^{de}) \\
 & = \frac{1}{n}\gamma_{pv}\gamma_{qp}\gamma^{de}D^{av\rho}{}_{\mu e(u)}\frac{\delta^n M_1^{\mu i}{}_{v)dk}(\gamma)}{\delta\gamma^{bc}\delta\gamma^{st}\dots} \\
 & = -\gamma_{pv}\gamma_{qp}\gamma^{de}D^{av\rho}{}_{\mu e(u)\gamma\alpha\gamma\iota)\beta\gamma^{\gamma\delta}}D^{\mu\alpha\beta}{}_{\theta\delta(b)}\frac{\delta^{n-1}M_1^{\theta i}{}_{c)\gamma jk}(\gamma)}{\delta\gamma^{st}}. \tag{49}
 \end{aligned}$$

In the last line we have used (43). So we arrive at the equation

$$\begin{aligned}
 & D^{av\rho}{}_{\mu e b}\frac{\delta}{\delta\gamma^{st}}(\gamma_{pv}\gamma_{qp}\gamma^{de}) \\
 & = -\gamma_{pv}\gamma_{qp}\gamma^{\gamma e}D^{av\rho}{}_{\theta e(s)\gamma\alpha\gamma\iota)\beta\gamma^{d\delta}}D^{\theta\alpha\beta}{}_{\mu\delta b}, \tag{50}
 \end{aligned}$$

where we have changed the free indices to avoid potential confusions. This is the equation represented schematically by (46). The reader can find in the Appendix the demonstration that this algebraic relation is indeed true and, therefore, the induction proof is finished.

Notice that, as the construction of the iterative series relies ultimately in the solution of a system of ordinary differential equations schematically represented by (45),

with an initial condition posed in flat space, the solution is unique. The solution \mathcal{A} is then

$$\begin{aligned}
 & \frac{1}{4}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\int d\mathcal{V}_\eta\frac{\delta^n M_1^{ai}{}_{bcjk}}{\delta\gamma^{pq}\delta\gamma^{st}\dots}\Big|_{\gamma\rightarrow\eta}\nabla_a h^{bc}\nabla_i h^{jk}h^{pq}h^{st}\dots \\
 & = \frac{1}{4}\int d\mathcal{V}_\eta M_1^{ai}{}_{bcjk}(\eta + \lambda h)\nabla_a h^{bc}\nabla_i h^{jk} \\
 & = \frac{1}{4\lambda^2}\int d\mathcal{V}_\eta M_1^{ai}{}_{bcjk}(g)\nabla_a g^{bc}\nabla_i g^{jk} = \mathcal{A}, \tag{51}
 \end{aligned}$$

where we have defined the field

$$g^{ab} := \eta^{ab} + \lambda h^{ab}. \tag{52}$$

Remember that ∇ is the covariant derivative compatible with η_{ab} .

B. Nonminimal couplings and surface terms

In this section we shall deal with the effect of allowing contributions to the stress-energy tensor coming from nonminimal couplings or, what is equivalent in this case, covariant surface terms. These terms fully parametrize the ambiguity inherent to the definition of the source in the equations of motion. They must be considered for the sake of completeness when the free action in flat space is generalized to a general metric space in terms of the auxiliary metric γ_{ab} , remember e.g., (23).

Nonminimal couplings are defined as scalar quantities which can be written in terms of the auxiliary metric γ_{ab} in the procedure above and the spin-2 field, and which vanish in the flat-space limit. The most general form of these terms, as they would be added to (23), is given by

$$\int d\mathcal{V}_\gamma [A^i{}_{bcjk}(\gamma, \nabla\gamma)h^{bc}\nabla_i h^{jk} + B_{bcjk}(\gamma, \nabla\gamma)h^{bc}h^{jk}]. \tag{53}$$

The first function $A^i{}_{bcjk}(\gamma, \nabla\gamma)$ must be proportional to $\nabla\gamma$, while the second one $B_{bcjk}(\gamma, \nabla\gamma)$ to $(\nabla\gamma)^2$ or $\nabla^2\gamma$. Using the flat covariant derivative ∇ guarantees that these terms vanish in the flat-space limit. We have also restricted them with the condition of leading to contributions to the stress-energy tensor which are quadratic in the derivatives of the spin-2 field. These contributions are obtained by varying this expression with respect to γ^{ab} (after integrating by parts) and then taking the flat-space limit.

The reader could find strange the form (53) we associate with nonminimal couplings. While the usual representation uses curvature-related tensor quantities, as the Riemann tensor, constructed from specific combinations of the auxiliary metric and its ordinary derivatives $\partial\gamma$, in (53) we are using arbitrary scalar combinations of the metric and its covariant derivatives $\nabla\gamma$. To do that we are exploiting the fact that we have a Minkowski reference metric, which

permits us to easily construct scalar quantities which contain the (η -covariant) derivatives of the auxiliary metric. Let us consider as an example the Riemann tensor: given a generic decomposition of a metric γ_{ab} in the form $\gamma_{ab} = q_{ab} + \epsilon_{ab}$, one can always write its Riemann tensor, $R^i{}_{bcj}(\gamma)$, as

$$R^i{}_{bcj}(\gamma) = R^i{}_{bcj}(q) + 2\bar{\nabla}_{[c}\bar{C}^i{}_{j]b} + 2\bar{C}^i{}_{d[c}\bar{C}^d{}_{j]b}. \quad (54)$$

In this expression, $\bar{C}^a{}_{bc}$ is the tensor which characterizes the difference between covariant derivatives with respect to the two metrics γ_{ab} and q_{ab} , respectively denoted by ∇' and $\bar{\nabla}$ [see for example [31], Eq. (D7) adapted to our sign conventions]. Now one can consider the special situation in which $q_{ab} = \eta_{ab}$ to realize that contractions with the Riemann tensor of γ_{ab} can be written as a particular case of the integrand in the expression (53).

With this definition of the possible nonminimal couplings, it is not difficult to realize that the same effect can be reproduced by adding a general covariant surface term instead. This term would have the following form, after writing the original action in terms of the auxiliary metric [i.e., it would be added to (23)]:

$$\int d\mathcal{V}_\gamma \nabla'_a (S^{ia}{}_{bcjk}(\gamma) h^{bc} \nabla'_i h^{jk}). \quad (55)$$

As in the case of nonminimal couplings, this is the most general possible expression containing two covariant derivatives of the spin-2 field. Recall that ∇' is the covariant derivative associated with γ_{ab} .

Whether surface terms contribute or not to the stress-energy tensor is a matter of choice. If we first substitute the bulk integral by an integral in the boundary, and then perform variations of γ^{ab} but maintaining it fixed on this boundary, one would obtain nothing from this variation. If instead one first performs this same variation, one obtains a local contribution to the stress-energy tensor inside the boundary, which does not change the values of global charges. Which stress-energy tensor is the appropriate one can only be distinguished precisely by gravitational experiments. Without further knowledge this is an ambiguity in the definition of the stress-energy tensor (it is equivalent to the ambiguity exploited in the Belinfante-Rosenfeld prescription).

The question now is whether the results we have obtained in the previous section could change because of the introduction of nonminimal couplings. In other words, we want to know whether there exists a different functional

$$\mathcal{A}' := \frac{1}{4} \int d\mathcal{V}_\eta O^{ai}{}_{bcjkpq}(\eta) h^{pq} \nabla_a h^{bc} \nabla_i h^{jk} \quad (56)$$

solution up to order $\mathcal{O}(\lambda)$ of the iterative procedure when certain additional terms in the stress-energy tensor are taken into account.

The effect of these surface terms or nonminimal couplings would be to add some terms of the form $\nabla(h\nabla h)$ to the stress-energy tensor. Thus, the iterative equations give us two conditions, analogous to (32) and (33): the first one is directly

$$O^{ai}{}_{bcjkpq} = \left. \frac{\delta M_{\mathfrak{z}}^{ai}{}_{bcjk}}{\delta \gamma^{pq}} \right|_{\gamma \rightarrow \eta}, \quad (57)$$

as in the minimal coupling case, while the second one will notice the effect of the surface terms, as it is changed to

$$-O^{ai}{}_{bcjkpq} = \eta_{pv} \eta_{qp} \eta^{de} D^{av\rho}{}_{\mu e(b} M_{\mathfrak{z}}^{\mu i}{}_{c) djk} + \Delta^{ai}{}_{bcjkpq}, \quad (58)$$

where the term $\Delta^{ai}{}_{bcjkpq}$ is the contribution coming from the surface term.

The second equation must be now understood as the condition which permits us to know what surface term we need in order to make the self-coupling procedure consistent for different values of the parameter \mathfrak{z} . That is, the addition of surface terms allows us to find solutions to the problem for $\mathfrak{z} \neq 1$. Now the first equation (57) implies that the solution, if it exists, will be expressible as the first term of a Taylor expansion in λ of the free action displaced to $\eta^{ab} + \lambda h^{ab}$, for any value of \mathfrak{z} . That is,

$$\begin{aligned} \mathcal{A}_0 + \lambda \mathcal{A}_1 + \mathcal{O}(\lambda^2) \\ = \frac{1}{4} \int d\mathcal{V}_\eta M_{\mathfrak{z}}^{ai}{}_{bcjk}(\eta + \lambda h) \nabla_a h^{bc} \nabla_i h^{jk} + \mathcal{O}(\lambda^2). \end{aligned} \quad (59)$$

Then, the complete iterative procedure will give place to the complete Taylor series in complete analogy with the minimal coupling case (51).

Let us consider now the issue of the variation of the volume element $d\mathcal{V}_\gamma$ or, in other words, of the factor $\sqrt{-\gamma}$ in the partial actions $\mathcal{A}_n[\gamma]$. The only difference in the integration of the first-order iterative equation is that the variation of the determinant $\delta\sqrt{-\gamma}$ must be taken into account in (24). This implies that the measure in (59) as well as in the final action would be given by $d\mathcal{V}_g := d\mathcal{V}_\eta \kappa$ with $\kappa := \sqrt{-g}/\sqrt{-\eta}$ instead of $d\mathcal{V}_\eta$. But the effect of this volume element at each order can be absorbed by non-minimal couplings as we did to obtain the solutions (59) with values of the parameter $\mathfrak{z} \neq 1$. This means that the general solution to the self-coupling problem is given by

$$\begin{aligned} \mathcal{A} &= \frac{1}{4} \int d\mathcal{V}_\eta \kappa' M_{\mathfrak{z}}^{ai}{}_{bcjk}(\eta + \lambda h) \nabla_a h^{bc} \nabla_i h^{jk} \\ &= \frac{1}{4\lambda^2} \int d\mathcal{V}_\eta \kappa' M_{\mathfrak{z}}^{ai}{}_{bcjk}(g) \nabla_a g^{bc} \nabla_i g^{jk}. \end{aligned} \quad (60)$$

This form (especially the fact that only the combination $g^{ab} = \eta^{ab} + \lambda h^{ab}$ appears) was not a logical necessity from

the beginning, but the analysis shows that it is actually the result. The factor κ' is either $\kappa' = 1$ or $\kappa' = \kappa$ depending on the prescription we follow to obtain the source at different orders. By expanding this action with respect to $g^{ab} = \eta^{ab} + \lambda h^{ab}$ in the formalism of [9], one can alternatively see that it indeed leads to a solution to the self-coupling problem with the appropriate quadratic (zeroth order) form for each value of the parameter \mathfrak{z} , and evaluate the necessary surface terms in a different way. What at the linear level is a surface term, in the final theory is no longer reducible to a surface term, giving place to a complete s -parameter family of solutions to the problem. Before ending this section, let us recall that all the solutions we have found are constructed as bimetric theories. Although the final theories all exhibit the tensor g^{ab} in their coupling to matter, the flat metric η^{ab} forms also part of the construction. We will discuss later what happens if one decides to eliminate η^{ab} to make contact with the standard general relativistic formulations.

V. THE FULL ACTION: RELATION WITH UNIMODULAR GRAVITY

In this section we briefly analyze the resulting nonlinear theories which we have obtained as solutions to the self-coupling problem. In particular, we investigate the internal gauge symmetry of these theories, and argue that this feature can be used to distinguish between the different possibilities.

Summarizing, we have found a family of theories (60) which depend on a real parameter \mathfrak{z} and are expressed in terms of the variable $g^{ab} = \eta^{ab} + \lambda h^{ab}$. Moreover, this field g^{ab} is constrained by a finite version of a (possibly) nonlinear equation of the form $f_{ab}\delta g^{ab} = 0$. This constraint guarantees that the resulting theory has the same degrees of freedom as the original linear construction of the spin-2 field. Since this is a scalar constraint, two options arise: $\eta_{ab}h^{ab} = 0$ or $\sqrt{-g} = \sqrt{-\eta}$. The first one is the original constraint imposed in the free-field functional space. However, when considering the self-interacting theory it is natural to expect that a modified nonlinear condition unfolds instead of maintaining the original traceless condition. This is the second case which reduces to the first one at the lowest nontrivial order in the coupling constant λ . These different selections of the parametric and functional freedom lead to different theories. Each theory has its own peculiarities which one can like or dislike. Let us now discuss these peculiarities case by case. Notice that all of these theories are by construction invariant under general changes of coordinates. However, the amount of internal gauge symmetry that they present can be different.

A. Nonlinear trace theories

In this section we will consider that the deformation $\sqrt{-g} = \sqrt{-\eta}$ holds. Under this condition, $\kappa' = 1$ always.

The first useful thing to do is try to express the action of the theory, (60), in an alternative form. To do that, let us introduce the derivative operator $\tilde{\nabla}$ associated with the field g^{ab} interpreted as a spacetime metric, such that

$$\tilde{\nabla}_a g^{bc} = 0. \quad (61)$$

Now we can define a tensor field C_{ab}^c relating the two derivative operators $\tilde{\nabla}$ and ∇ .⁶ We are going to show that the entire action can be written in terms of this tensor field.

If we expand the compatibility condition (61) and multiply it by the field g^{ab} , one has

$$g^{ak}\nabla_k g^{bc} = -C_{kl}^b g^{ak} g^{cl} - C_{kl}^c g^{ak} g^{bl}. \quad (62)$$

By performing permutations of the free indices of this equation, we can write

$$-2C_{kl}^c g^{ak} g^{bl} = g^{ak}\nabla_k g^{bc} + g^{bk}\nabla_k g^{ac} - g^{ck}\nabla_k g^{ab}. \quad (63)$$

Now we can multiply this equation by the inverse matrix g_{ab} (which has nothing to do with the contravariant version of g^{ab} obtained by acting twice with the flat metric η_{ab}) to solve for C_{ab}^c :

$$\begin{aligned} C_{ab}^c &= -\frac{1}{2}g_{al}g_{bm}(g^{lk}\nabla_k g^{mc} + g^{mk}\nabla_k g^{lc} - g^{ck}\nabla_k g^{lm}) \\ &= -\frac{1}{2}(g_{bm}\nabla_a g^{mc} + g_{al}\nabla_b g^{lc} - g_{al}g_{bm}g^{ck}\nabla_k g^{lm}). \end{aligned} \quad (64)$$

Then one has

$$\begin{aligned} g^{ab}C_{ja}^i C_{ib}^j &= \frac{g^{ab}}{4}(g_{jl}\nabla_a g^{li} + g_{al}\nabla_j g^{li} - g_{al}g_{jm}g^{ik}\nabla_k g^{lm}) \\ &\quad \times (g_{bn}\nabla_i g^{nj} + g_{in}\nabla_b g^{nj} - g_{ir}g_{bs}g^{jn}\nabla_n g^{rs}) \\ &= \frac{1}{4}(2g_{ln}\nabla_j g^{li}\nabla_i g^{nj} - g_{ln}g_{jm}g^{ik}\nabla_k g^{lm}\nabla_i g^{nj}) \\ &= \frac{1}{4}M_{1\ bcjk}^{ai}(g)\nabla_a g^{bc}\nabla_i g^{jk}. \end{aligned} \quad (65)$$

This means that we can write the action, at least for the special case $\mathfrak{z} = 1$, as

$$\frac{1}{4\lambda^2} \int d\mathcal{V}_\eta M_{1\ bcjk}^{ai}(g)\nabla_a g^{bc}\nabla_i g^{jk} = \frac{1}{\lambda^2} \int d\mathcal{V}_\eta g^{ab} C_{ja}^i C_{ib}^j. \quad (66)$$

What is interesting about this expression is that it permits us to connect with the usual geometrical language of general

⁶The most consistent notation with previous definitions would be C_{ab}^c instead of C_{ab}^c . However, here we have chosen the latter notation which simplifies the appearance of the subsequent equations.

relativity, with g_{ab} playing the role of the spacetime metric. To see that, let us consider the Einstein-Hilbert action which contains the curvature scalar R of a metric g_{ab} . As we have already discussed in IV B, if the metric is split as $g^{ab} = \eta^{ab} + \lambda h^{ab}$, the curvature scalar can be written in terms of η -compatible derivatives ∇ of the field C_{ab}^c . Then, we can eliminate a surface term by just realizing [32] that

$$\begin{aligned} & \frac{2}{\lambda^2} \int d\mathcal{V}_g g^{bj} (\nabla_{[c} C_{j]b}^c + C_{d[c}^c C_{j]b}^d) \\ &= \frac{2}{\lambda^2} \int d\mathcal{V}_\eta \nabla_c (\sqrt{-g} \delta_a^{[c} g^{d]b} C_{bd}^a) \\ & \quad - \frac{2}{\lambda^2} \int d\mathcal{V}_g g^{bc} C_{d[c}^c C_{j]b}^d. \end{aligned} \quad (67)$$

The surface term is

$$\frac{2}{\lambda^2} \int d\mathcal{V}_\eta \nabla_c (\sqrt{-g} \delta_a^{[c} g^{d]b} C_{bd}^a), \quad (68)$$

and the remaining action is precisely

$$-\frac{2}{\lambda^2} \int d\mathcal{V}_g g^{bc} C_{d[c}^c C_{j]b}^d. \quad (69)$$

This action was first written by Rosen in the context of a gravitational theory with a preferred flat background [1]. Notice that the expression under the integral sign in this last equation is a scalar under general coordinate transformations, so one does not need to complement it with a surface term to ensure this invariance. On the other hand, it is not a scalar under internal gauge transformations as one would need to add the surface term (68) to guarantee that. In this paper we stick to a bimetric formulation (in the sense of Rosen), as our analysis shows that it is really one of the most important consequences of giving full credit to the self-coupling problem. It is not possible to obtain unimodular gravity (or general relativity, see next section) in a strict sense but Rosen's reformulation of it, when starting from a free theory in flat spacetime and using only flat-spacetime notions.

The only thing we need to do to make full contact with our action (66) is to impose the condition on the determinant $\sqrt{-g} = \sqrt{-\eta}$. Under this condition, $d\mathcal{V}_g = d\mathcal{V}_\eta$ and $C_{bc}^b = 0$ as it can be shown by using a particular Minkowski reference frame:

$$C_{bc}^b|_{\mathcal{M}} = -\frac{1}{2} g_{ab} \partial_c g^{ab} = \frac{1}{\sqrt{-g}} \partial_c \sqrt{-g} = 0. \quad (70)$$

So the first term in (69) can be dropped and the volume element is replaced by $d\mathcal{V}_\eta$, making this action completely equivalent to (66). Therefore, we have recovered the general relativity action, subject to the determinant restriction. Now it is easy to analyze the internal gauge symmetry

of this theory, whose generators will be denoted by ξ^a . Its infinitesimal counterpart has exactly the same form as a diffeomorphism,

$$\delta_\xi h^{ab} = \mathcal{L}_\xi h^{ab} = -\xi^c \nabla_c g^{ab} + g^{ac} \nabla_c \xi^b + g^{bc} \nabla_c \xi^a, \quad (71)$$

with the additional condition of preserving the Minkowski volume element:

$$\nabla_a \xi^a = 0. \quad (72)$$

Here \mathcal{L}_ξ is the Lie derivative operator. This condition is nothing but the first one in (8). As we have the same number of generators subjected to the same number of restrictions there is no reduction of gauge symmetry, just a deformation.

However, the contrary happens when $\mathfrak{s} \neq 1$. One would need to impose additional conditions on the generators to make sure that (71) is a symmetry of the theory for these values of $\mathfrak{s} \neq 1$, which means that the gauge symmetry is reduced (see the next section for additional comments).

Independently of this we would have, when $\mathfrak{s} \neq 1$, additional restrictions coming from the transverse condition:

$$\nabla_a g^{ab} = \nabla_a h^{ab} = 0, \quad (73)$$

which can be alternatively written as

$$C_{bc}^a g^{bc} = 0. \quad (74)$$

This means that we will also need to impose a deformation of the second condition in the same equation whose expression can be obtained by imposing $\nabla_b \delta_\xi h^{ab} = 0$:

$$\square \xi^a + \lambda h^{bc} \nabla_b \nabla_c \xi^a = 0. \quad (75)$$

Remember that what singles out the case $\mathfrak{s} = 1$ from the other values from the point of view of the internal symmetry is that the transverse condition can be dropped even at the level of the free spin-2 theory, which means that the conditions (73) can be relaxed from the beginning. This corresponds to the situation analyzed by symmetry arguments in [18]; then our discussion is compatible with the content of this paper (notice that the transverse condition plays no role in the solution of the iterative equations of the self-coupling problem). This is different from $\mathfrak{s} \neq 1$, as in those cases we will always need to impose the transverse condition even in the free theory, which implies that the resulting theory will always present the deformation (75) in order to ensure this condition.

Even if we consider g^{ab} as a metric, we still use the natural volume form in Minkowski spacetime to write the action of the resulting theory, (66). As we started from a theory formulated in a fixed background, the Minkowski spacetime, there is no logical reason to expect a perfect decoupling of this background structure in the sense of

emergent gravity scenarios [33]. In fact, we have seen that the presence of this background is only partially camouflaged in the self-interacting theory when all orders in the interaction are taken into account, but not entirely, as this preferred notion of volume survives. However, in this case this nonperfect decoupling is not an undesirable feature: the existence of a background volume element is a definitory characteristic of unimodular gravity and the ultimate reason which makes this theory avoid the first cosmological constant problem [34,35]. The consideration of unimodular gravity as the self-interacting theory of a spin-2 field in Minkowski spacetime thus offers a natural way of understanding this feature which, on the other hand, is rather unnatural from the geometrical point of view.

Concerning the purely geometrical point of view, at this stage one could take it one step further. Imagine that instead of writing the action in coordinate independent manner we select a particular Cartesian coordinate system. Then, all the covariant derivatives in the action would be substituted by partial derivatives. If someone gives us this action without informing where does it come from we will not have any way of noticing the existence of an external background. The restricted variation conditions will become

$$\sqrt{-g} = 1, \quad \Gamma_{bc}^a g^{bc} = 0, \quad (76)$$

i.e., the unit determinant and the so-called harmonic gauge condition. In this way any reference to a Minkowski background is completely erased. The background is so camouflaged that one can even forget it exists. We could switch to a completely geometrical interpretation of the theory. In this form the theory would have to be interpreted similarly to general relativity but only selecting certain coordinate systems as special. This happens because of the mixing of the external and internal symmetries, both are now one and the same symmetry. One can forget about the harmonic condition and allow for almost generic coordinates in the geometrical description. Then only the determinant of the metric is restricted to be minus one. This is precisely what is standardly considered as unimodular gravity. To make this geometrization conceptual jump or not is however optional, not required by the iterative construction, and most importantly, it is not inconsequential.

B. Linear trace theories

Now let us consider what happens if the functional space is constrained by the condition $\eta_{ab} h^{ab} = 0$. As in the previous case it is better to start with the particular value $\mathfrak{s} = 1$. The invariance of the traceless condition, $\eta_{ab} \delta_\xi h^{ab} = 0$, assumed in this particular construction implies the additional requirement for the generators ξ^a :

$$\eta_{ab} g^{ac} \nabla_c \xi^b = 0. \quad (77)$$

The invariance of the transverse condition, $\nabla_b \delta_\xi h^{ab} = 0$ implies in this case

$$\square \xi^a + \lambda h^{bc} \nabla_b \nabla_c \xi^a + g^{ab} \nabla_b \nabla_c \xi^c = 0. \quad (78)$$

Therefore, the gauge symmetry of the original linear spin-2 theory has been deformed but also reduced because the generators ξ^a are subjected to three conditions (the transverse condition $\nabla_a \xi^a = 0$ if $\kappa' = 1$ or the deformation $\nabla_a \xi^a = 0$ if $\kappa' = \kappa$ must also hold) instead of two. A similar thing occurs with the transverse condition and the theories with $\mathfrak{s} \neq 1$ independently. We see then that the constraints (5) which define the original functional space of the free theory must be suitably relaxed to guarantee that the conditions which their preservation impose on the generators of the transformation (71) match with the conditions which guarantee that this transformation is in fact a symmetry of the theory, thus making compatible these kinematical and dynamical aspects of the resulting nonlinear theory. If this is not the case, the gauge symmetry of the theory would be reduced. Whether this reduction of symmetry can be acceptable or not seems to be a matter of taste from the perspective of the self-coupling problem only, although a deeper analysis of these theories probably would reveal potentially observable consequences of this reduction. Notice that in this theory the Minkowski metric appears explicitly in the field equations (something similar occurs in the theory which was found by self-coupling in [36]).

VI. RELATION TO PREVIOUS WORK

In this section we are going to schematically describe how the techniques developed in previous sections can be applied to Fierz-Pauli theory. We also discuss some of the seemingly contradictory conclusions in the recent literature. From the perspective of the spin-2 theory, Fierz-Pauli theory constitutes an enlargement of the available functional space in which the field h^{ab} is defined, with a parallel enlargement of the internal gauge symmetry. Fierz-Pauli theory [19] is defined by the action

$$\mathcal{F}_0 := \frac{1}{4} \int dV_\eta F_{\mathfrak{s} \ bcjk}^{ai}(\eta) \nabla_a h^{bc} \nabla_i h^{jk}, \quad (79)$$

with

$$\begin{aligned} F_{\mathfrak{s} \ bcjk}^{ai}(\eta) := & M_{1 \ bcjk}^{ai}(\eta) - 2\delta_{(b}^a \delta_{c)}^i \eta_{jk} + \eta^{ai} \eta_{bc} \eta_{jk} \\ & + \frac{1 - \mathfrak{s}}{2} [\eta_{j(b} \delta_{k)}^a \delta_c^i + \eta_{k(b} \delta_j^a \delta_c^i - \delta_{(b}^a \delta_j^i \eta_{c)k} \\ & - \delta_{(b}^a \delta_k^i \eta_{c)j}]. \end{aligned}$$

The parameter \mathfrak{s} can acquire any value leaving the theory unchanged because the combination

$$(\delta_j^a \delta_b^i \eta_{ck} - \delta_b^a \delta_j^i \eta_{ck}) \nabla_a h^{bc} \nabla_i h^{jk} \quad (80)$$

is just a surface term and this is why in many places Fierz-Pauli theory is presented as the previous one with \mathfrak{z} set to one. In our discussion we are leaving this term explicit, since we have learned from the spin-2 case that it could be of importance. This action has the internal gauge symmetry

$$\delta_\xi h^{ab} = \eta^{ac} \nabla_c \xi^b + \eta^{bc} \nabla_c \xi^a, \quad (81)$$

where now the generators are unrestricted. This action is the only quadratic action invariant under these gauge transformations [8]. We are not going to need here the explicit form of the tensor $F_{\mathfrak{z}}^{ai}{}_{bcjk}(\eta)$ though. The only thing we need to keep in mind is that it can be written in terms of $M_1^{ai}{}_{bcjk}(\eta)$ plus additional terms, which are now present because the conditions (5) no longer hold.

We can try to apply now the same self-interacting scheme, but with $F_{\mathfrak{z}}^{ai}{}_{bcjk}(\eta)$ instead of $M_{\mathfrak{z}}^{ai}{}_{cbjk}(\eta)$, to see whether we are able to obtain general relativity as the outcome. However, this procedure does not work out so straightforwardly in this case. The first evidence of this is that there does not exist any \mathfrak{z} for which the analogue of (34) is true, i.e.,

$$\sqrt{-\gamma} \gamma_{\rho\nu} \gamma_{\rho\sigma} \gamma^{de} D^{av\rho}{}_{\mu e(b} F_{\mathfrak{z}}^{\mu i}{}_{c) djk} \neq - \frac{\delta \sqrt{-\gamma} F_{\mathfrak{z}}^{ai}{}_{pqjk}}{\delta \gamma^{bc}}. \quad (82)$$

A way of realizing this is the following: the right-hand side of this equation contains terms which are proportional to γ_{bc} , not contracted with $F_{\mathfrak{z}}^{ai}{}_{pqjk}$, because of the variation of the determinant. However, the left-hand side of this equation does not contain this kind of terms. Independently of the form of $F_{\mathfrak{z}}^{ai}{}_{bcjk}$, for the first term in the left-hand side of the previous equation one has e.g.,

$$\gamma_{\rho\nu} \gamma_{\rho\sigma} \gamma^{de} D^{av\rho}{}_{\mu e b} = \gamma_{b(p} \delta_{\mu}^a \delta_{q)}^d + \delta_b^a \gamma_{\mu(p} \delta_{q)}^d - \gamma_{b(p} \gamma_{q)\mu} \gamma^{ad}. \quad (83)$$

The index b never appears in combination with the free index c . The same happens with the second term in (82). This means that not all the sources are valid for the self-coupling procedure, if we want the final theory to be describable by a Lagrangian theory. In particular, we have seen that the application of the Hilbert prescription to the free Fierz-Pauli action (79) leads to a stress-energy tensor which cannot be derived from an action by performing variations of h^{ab} [20]. The requirement that the equations of motion be derivable from an action cannot be ignored, since it is in the heart of the definition of the iterative self-coupling problem.

The natural thing to check next is whether it is possible to find a solution of the iterative equations by introducing nonminimal couplings or, what is equivalent, covariant surface terms. In particular, one can ask whether general

relativity can appear as a result of such self-coupling scheme. This question is clearly and affirmatively answered in the work by Butcher *et al.* [9]. To do that, these authors performed a reverse engineering exercise. Let us discuss it briefly here. The Einstein-Hilbert action can be expanded as a series in λ using the decomposition $g^{ab} = \eta^{ab} + \lambda h^{ab}$. By construction, this series is a solution of the iterative equations (20).⁷ These authors explicitly show that, to guarantee that the overall procedure makes sense, one needs to accept the following condition. When writing the lowest order \mathcal{A}_0 in terms of an auxiliary metric γ_{ab} to obtain $\mathcal{A}_0[\gamma]$, this quantity must contain nonminimal couplings as they are necessary to obtain the stress-energy tensor appearing in the lowest-order iterative equation (this happens also for higher orders). The quadratic action $\mathcal{A}_0[\gamma]$, when particularized to Minkowski space, leads precisely to \mathcal{F}_0 in (79) with $\mathfrak{z} = 1$.

Thus there exists a certain source, obtained through the addition of nonminimal couplings, which permits one to recover general relativity as a self-interacting theory of the Fierz-Pauli field. As it happened before with the spin-2 theory, here the nonminimal couplings can be understood as surface terms in the free Lagrangian density. In fact, with the appropriate addition of surface terms all of the actions (79) with an arbitrary value of \mathfrak{z} can be uplifted to nonlinear theories that are solutions of the iterative equations. Another matter is whether these final theories have some internal gauge symmetry or not. What is clear is that only the value $\mathfrak{z} = 1$ leads to a theory with an internal symmetry of the form of the usual diffeomorphism invariance.

If we consider the nontensorial general-relativity action

$$\frac{1}{\lambda^2} \int dV_g g^{ab} (\Gamma_{da}^c \Gamma_{cb}^d - \Gamma_{ab}^c \Gamma_{cd}^d), \quad (84)$$

and perform an expansion in the parameter λ with $g^{ab} = \eta^{ab} + \lambda h^{ab}$, we will see that it precisely exhibits a coupling term of the form $h^{ab} S_{ab}$ at first order in λ . Padmanabhan pointed out the role of this object S_{ab} in any coupling scheme leading to general relativity [8]. He showed for instance that this object S_{ab} can be obtained from the quadratic term (zeroth order in λ) by applying only a half-covariantization scheme which might be regarded at least as unnatural (see Appendix A in [9] for additional comments on this quantity). Somewhat surprisingly, whereas the quadratic action is tensorial, the first order correction should already be nontensorial. The variation of this new action with respect to γ_{ab} might lead in principle to a nontensorial stress-energy object (though finally this is not the case). Therefore, one could argue, as Padmanabhan, that the construction of general relativity from a self-coupling

⁷It can be checked that, as long as we are performing a Taylor expansion of a function of the form $F(\eta + \lambda h)$, these conditions are verified.

scheme is somewhat unphysical (only at the end of the iterative procedure one would realize the existence of a surface term allowing the construction of a diffeomorphism-invariant Lagrangian density).

However, in our formulation we always keep track of the flat reference metric. This allows us to construct the tensorial action

$$\frac{1}{\lambda^2} \int dV_g g^{ab} (C_{da}^c C_{cb}^d - C_{ab}^c C_{cd}^d), \quad (85)$$

instead of the nontensorial action (84). The cubic term has a form $h^{ab} \bar{S}_{ab}$ where now \bar{S}_{ab} is a proper tensor. Moreover, this object is not and must not be the stress-energy tensor. We have seen that there is a natural definition of \bar{S}_{ab} within the iterative procedure, as the result of the integration of the first-order iterative equation analogue to (18). In fact one of the main differences between the work of Padmanabhan [8] and that in here is that we have explicitly performed the integration of the iterative equations. In other words, from the point of view of the self-coupling consistency problem, \bar{S}_{ab} is just a derived quantity and not a fundamental one. One will be led to it by following the equations carefully [recall for example the discussion around (30)]. Concerning this last point, Deser makes a similar comment in his reply to Padmanabhan [10]: the only role of \bar{S}_{ab} is to lead to the required source when the variations which respect to h^{ab} are performed, and this is precisely the definition of this quantity.

Concerning the surface term: the Einstein-Hilbert action can be partitioned in a first-derivative action plus a surface term in several ways. If one does not introduce a fiducial background metric, this partition has to be nontensorial. Instead, by introducing a flat background metric, one discovers a tensorial partition. In our view what is unnatural from the self-coupling program is precisely to forget about the background metric, making an identification of the coordinate invariance and the invariance under gauge transformations. Once one obtains the action (85), which is a scalar, one would not look for complementing this action with additional surface terms to build the scalar curvature. Only when taking the nontrivial conceptual jump of forgetting about the background structure and taking a complete geometrical description in terms of a single metric, one would start worrying about the significance of the surface term and its nontensorial character. In this stage we agree with Padmanabhan's [8] that the surface term of the Einstein-Hilbert action is not naturally obtainable by the self-coupling procedure, one has to add geometrical information.

Butcher *et al.* [9] say that "general relativity cannot be derived from energy-momentum self-coupling the Fierz-Pauli Lagrangian." More precisely what they mean is that one cannot use the stress-energy tensor obtained straightforwardly from the Fierz-Pauli Lagrangian density by using

a minimal coupling prescription. One has to add specific nonminimal couplings. From reading this paper and Padmanabhan's, one ends up with the impression that to obtain general relativity from self-interaction one needs to know somehow the final result, as one needs to make use of curved-spacetime notions. However, here we have shown that nonminimal couplings are encompassed by covariant surface terms, which form part of the standard arbitrariness in defining the stress-energy tensor even in flat spacetime. Allowing surface terms one finds a one-parameter family of solutions to the self-coupling problem. From them, general relativity is selected by requiring the final theory to have the largest possible amount of gauge invariance. Thus the construction only uses concepts based on Poincaré-invariant field theory and gauge invariance (although some of the mathematical tools used can be geometrical, as the Hilbert prescription to obtain the stress-energy tensor). It is instructive to notice that the necessity of considering the addition of identically conserved terms to the source one would obtain directly from the free action is not exclusive of gravity, but the same thing happens when considering the case of Yang-Mills theory in the second-order formalism, as it is explicitly written (but at some extent ignored) in [6].

All these comments apply to the classic work of Deser [6,10], e.g., the fact that the resulting action will be written in terms of the covariant derivative ∇ with respect to the flat reference metric. The clever choice of independent variables in that work allowed him to lead to completion the iterative procedure in a single step. Precisely, this selection of variables hides the fact that the stress-energy tensor obtained by varying η^{ab} is not the one that one would directly obtain from the minimally coupled Fierz-Pauli theory. That is, Deser's first order formalism naturally selects the specific surface term (or nonminimal coupling) that leads to Einstein equations. The reader should not confuse this surface term at the level of the free theory with the surface term in the Einstein-Hilbert action, which is put by hand. Notice that all the other potential solutions to the self-coupling problem are absent. One could recover them by using additional surface terms in his direct construction. One cannot simply exclude these possibilities from the perspective of logic, but now the arguments in [6] do not apply as the resulting iterative series would be now infinite. A more general treatment such as the one presented here is needed, in which the use of specific variables is avoided and which permits one to handle infinite series to know the nature of these solutions.

In [10] it is argued that the nonuniqueness inherent to the use of Noether currents in the very definition of the self-coupling problem is harmless. The argument is that these identically conserved terms which appear in the definition of the source can be absorbed in a redefinition of the Fierz-Pauli field h^{ab} . If we want to keep us in the linear level, there is only one possible redefinition: shifting h^{ab} by its

trace $\eta_{ab}h^{ab}$. This means that one could absorb only certain types of such identically conserved terms. Even if we forget about this, it is difficult to see how this procedure could work as we argue in the following. Let us start with the first-order self-interacting equation,

$$O_{abcd}h^{cd} = \lambda T_{ab}(h^{\rho\sigma}) + \lambda \Theta_{ab}(h^{\rho\sigma}), \quad (86)$$

where O_{abcd} is a given differential operator [whose form can be obtained from the action (79)], $T_{ab}(h^{\rho\sigma})$ is the source we want to consider, and $\Theta_{ab}(h^{\rho\sigma})$ an identically conserved tensor constructed from h^{ab} . When $\lambda = 0$ we recover the free field equations. Now let us construct a different field,

$$h'^{ab} = h^{ab} + \lambda f^{ab}(h^{\rho\sigma}), \quad (87)$$

with $f^{ab}(h^{\rho\sigma})$ an arbitrary function of h^{ab} (which, if we want to keep at the linear level, should be proportional to $\eta^{ab}\eta_{cd}h^{cd}$). In [10] it is argued that there always exists a choice of $f^{ab}(h^{\rho\sigma})$ such that the field equations (86) can be written as

$$O_{abcd}h'^{cd} = \lambda T'_{ab}(h^{\rho\sigma}), \quad (88)$$

thus absorbing the identically conserved term $\Theta_{ab}(h^{\rho\sigma})$. The function $f^{ab}(h^{\rho\sigma})$ is determined by the following equation:

$$O_{abcd}f^{cd}(h^{\rho\sigma}) = -\Theta_{ab}(h^{\rho\sigma}). \quad (89)$$

As the operator O_{abcd} satisfies $\nabla^a O_{abcd} = 0$ [10], this equation is well posed so it seems that one can conclude that the identically conserved terms can be shifted away. However, one should not forget about the stress-energy tensor, which is not a mere spectator here but explicitly depends in the Fierz-Pauli field h^{ab} . Thus, at best, one can get instead of (88) an equation of the form

$$O_{abcd}h'^{cd} = \lambda T'_{ab}(h^{\rho\sigma}), \quad (90)$$

such as

$$T'_{ab}(h^{\rho\sigma}) = T_{ab}(h^{\rho\sigma}). \quad (91)$$

This means that the effect of the nonminimal couplings cannot be simply shifted away at all orders. One must take them into account, so the nonuniqueness problem remains open until an additional assumption is imposed.

In summary, in the Fierz-Pauli case the self-coupling problem naturally leads to a one-parameter set of solutions that includes general relativity. From this point of view general relativity naturally emerges from the self-coupling of a Poincaré-invariant field theory. However, to select general relativity from the other theories, one has to require the existence of a maximal gauge symmetry. There seems

to be no alternative guiding principle to directly obtain general relativity. Moreover, what is obtained is closer to the bimetric theory of Rosen [1].⁸ Thus, although both theories are observationally equivalent in standard situations, there can be conceptual differences when considering extreme situations such as spacetime singularities.

A. On gravitational energy

One of the aspects which is expected to be addressed by such a reinterpretation in terms of a flat reference metric of the gravitational theories is the issue of the gravitational energy. Also we have some loose ends in our construction (remember the discussion in III concerning the conservation of the traceless source \tilde{T}_{ab}) which demand to investigate this issue. Let us first consider in this section the example of general relativity. As we have been discussing, there are two equivalent descriptions of the same system and, within each one, distinct ways of stating conservation principles: on the one hand, the purely geometrical vision in which quantities are covariantly conserved (i.e., with respect to the covariant derivative $\tilde{\nabla}$ associated with the metric g_{ab}). On the other hand, we expect the system to possess conserved quantities associated with Poincaré invariance. In particular, there will be a stress-energy tensor associated with translation invariance, which will be conserved with respect to the flat derivative operator ∇ (within the space of solutions). In this section we explicitly recall how these two pictures fit (see [1] and e.g., [37]).

If we include matter in our considerations, then the resulting action has two parts. It is easy to realize that, by construction, taking the variational derivative with respect to the contravariant version of the auxiliary metric γ_{ab} in the matter part is equivalent to performing the same operation but with respect to g_{ab} . However, this is not true within the gravitational part. This means that the total canonical stress-energy tensor is given by

$$\kappa \Theta_{\text{M}}^a{}_b + t^a{}_b := \mathcal{L}_{\text{T}} \delta_b^a - \frac{\delta \mathcal{L}_{\text{T}}}{\delta(\nabla_a \psi^\mu)} \nabla_b \psi^\mu. \quad (92)$$

Here $\mathcal{L}_{\text{T}} := \mathcal{L}_{\text{G}} + \mathcal{L}_{\text{M}}$ is the total Lagrangian density, ψ^μ again represents all the fields, and $\kappa = \sqrt{-\tilde{g}}/\sqrt{-\tilde{\eta}}$ as defined before. In this equation, $\Theta_{\text{M}}^a{}_b$ gives the same conserved charges as the Hilbert stress-energy tensor for the matter part, while

$$t^a{}_b := \mathcal{L}_{\text{G}} \delta_b^a - \frac{\delta \mathcal{L}_{\text{G}}}{\delta(\nabla_a h^{cd})} \nabla_b h^{cd} \quad (93)$$

is a new object which provides the notion of gravitational energy. The factor κ in the matter part arises because of the

⁸The reader should not confuse the resulting theory with what is usually considered a bimetric theory, as here one of the metrics (the flat reference metric) is not a dynamical entity.

occurrence of $\sqrt{-g}$ instead of $\sqrt{-\eta}$ as the natural measure in the action. We are using the canonical stress-energy tensor instead of the Hilbert prescription just to make better contact with the literature: the only important thing in the following arguments is the divergence of this quantity on solutions, thus using the Hilbert stress-energy tensor would lead exactly to the same calculations and conclusions.

As we have already mentioned, this overall quantity is conserved with respect to ∇ within the space of solutions, that is,

$$\nabla_a(\kappa\Theta_M^a{}_b + t^a{}_b)|_S = 0. \quad (94)$$

The subscript S means that the Euler-Lagrange equations are used. Now it can be shown that this last equation is equivalent to the covariant conservation of the Einstein tensor,

$$\tilde{\nabla}_a G^a{}_b = 0. \quad (95)$$

The first step to show this is to use the gravitational equations of motion which, in suitable units, permit us to write (94) as

$$\nabla_a(\kappa G^a{}_b + t^a{}_b) = 0. \quad (96)$$

One only needs to realize that this equation is now a purely geometrical statement (that is, it is solely written in terms of the metric g_{ab}) which, in fact, is equivalent to (95) as it was shown in [1].

This argument can be extended also to unimodular gravity. In this case, the equations of motion would be

$$R^a{}_b - \frac{1}{4}R\delta^a{}_b = T_M^a{}_b - \frac{1}{4}T_M\delta^a{}_b, \quad (97)$$

where now $T_M^a{}_b$ is the Hilbert stress-energy tensor of the matter part. But (94) still holds (with $\kappa = 1$), in which we can certainly replace $\Theta_M^a{}_b$ with $T_M^a{}_b$. If we rewrite the left-hand side of (97) as $G^a{}_b + R\delta^a{}_b/4$, then instead of the identity (95) we have the condition

$$\nabla_a(R + T_M) = 0. \quad (98)$$

One can see then that the self-coupling of the spin-2 field solves the problem associated with the conservation of the traceless source \tilde{T}_{ab} in (14). Moreover, one can remember this equation as the condition which permits the equations of motion of unimodular gravity to be formally equivalent to those of general relativity, but with a cosmological constant unrelated to the parameters in the action [38]. This means that in both cases (general relativity and unimodular gravity), the conservation of the matter stress-energy tensor within the space of solutions,

$$\tilde{\nabla}_a T_M^a{}_b|_S = 0, \quad (99)$$

is obtained as a consequence of the Poincaré invariance of the theory and the resulting nonlinear equations of motion. Put differently, the conservation of the total energy (written in a perfectly tensorial way) in flat spacetime on the one hand (94), and the conservation of the gravitational and matter parts independently in a curved dynamical space on the other hand, (95) and (99) respectively, are two ways of writing the same thing. It is important to notice that in previous constructions of unimodular gravity, (99) can only be taken as an assumption additional to the field equations [38], but here we have seen that from the perspective of the self-coupling problem it is in fact a necessary consequence of Poincaré invariance.

The tensor $t^a{}_b$ could be interpreted as the gravitational stress-energy tensor, in the sense that it can be used to evaluate the corresponding conserved charges. However, its interpretation as providing a local notion of energy is still problematic. The stress-energy tensor of the linear theory is not gauge invariant and this fact remains in the final stress-energy tensor: it is not a problem due to self-coupling. In other words, this feature does not appear when one performs the summation of the series but is present at each order, even in the free theory, so it should not be considered as a solid argument against the self-coupling program as Padmanabhan claims [8]. In fact, for us the vision is quite the opposite: if one insists in the preservation of the original gauge symmetry, then even the free theory is pushing you towards some kind of geometrical (nonlocal) interpretation from the beginning and, thus, the lack of this local notion of energy is really natural from the perspective of the self-coupling problem. Notice that even the matter stress-energy tensor T_{ab}^M becomes a gauge noninvariant quantity as a result of the coupling. However, the gauge transformation of T_{ab}^M is of tensorial form while that of $t^a{}_b$ is of nontensorial form: it can always be gauged to zero in a point. For this reason, as opposed to the case in the matter sector, it is not possible to extract any local meaning of energy from $t^a{}_b$. The single-metric geometrical interpretation offers a clear explanation of this issue, associating the gauge to zero of the stress-energy tensor to free-fall observers, which do not detect gravity.

The geometrical interpretation offers another way to define a gravitational stress-energy tensor by varying g^{ab} in the action. The gravitational stress-energy tensor is then directly G_{ab} (a similar proposal is argued for in [39,40]), that is a tensor with respect to changes of coordinates in Minkowski spacetime as well as gauge transformations. In this interpretation and the previous one, one could say that outside matter there is no gravitational energy. However, notice that this does not mean that gravitational waves do not carry energy. Gravitational wave solutions exist in the theory. They are solutions of the equations of motion. Then, the covariant conservation with respect to the metric g_{ab}

of the stress-energy tensor (99) tells us that these waves can imprint energy in the fields acting as detectors. The situation will be reminiscent of action-at-a-distance theories.

The theories of gravitons considered here as the starting points of the self-coupling procedure do not make use of a local meaning of gravitational energy, as the graviton stress-energy tensor is not a gauge-invariant observable. When the original notion of gauge invariance is preserved, this is also reflected in the resulting theories (unimodular gravity and general relativity). A proper local meaning would appear however if gauge invariance is broken. Then t^a_b will be a perfectly defined stress-energy tensor. In a sense, we could say that the self-interaction procedure shows us two possible alternative routes: (a) gauge invariance is preserved and one is pushed towards a complete geometrical interpretation, or (b) self-interaction breaks gauge invariance so that the final theory has a proper notion of local energy over the Minkowski background.

VII. CONCLUSIONS

In this paper we have discussed Gupta's original program in detail, concerning the possible theories which arise as self-interacting theories of gravitons propagating in Minkowski spacetime. The discussion applies to quantum theories whose low-energy spectrum contains gravitons interacting with matter in a flat background, as long as one accepts that the long wavelength limit is described by a classical, second-order Lagrangian field theory.

We have explicitly solved the infinite set of iterative equations that appears when using a standard formalism based on the tensor field variable h^{ab} for the graviton, thus complementing previous work in the subject concerning finite series which appear when specific variables are considered. To do that we have constructed a proof by induction and found the formal sum of the resulting series, starting from a free field theory with a minimal gauge invariance motivated by the irreducible spin-2 representation of the Poincaré group. Finally, we have extended and contrasted our approach with previous discussions in the literature which start instead from Fierz-Pauli theory, which has a larger gauge symmetry. The formalism we have used has permitted us to explicitly show the interplay between internal gauge invariance (a notion which is clearly separated from changes of coordinates) and the self-coupling procedure.

The main conclusions of our analysis are the following:

- (i) One obtains field equations which are equivalent to those of unimodular gravity and general relativity as the only consistent results of the self-coupling of gravitons as long as one requires that the amount of internal gauge symmetry of the linear theory is preserved, although in a deformed version, in the self-interacting theory (considering always theories with up to second derivatives of the fields; beyond
- (ii) Even if the resulting theories admit a geometrical interpretation (it is within this interpretation that we strictly speak of unimodular gravity and general relativity), they are naturally some kind of bimetric theory, similar to the construction of Rosen [1], and not directly unimodular gravity and general relativity, which from this perspective have forgotten the existence of a flat reference metric. In any case, no observational difference can be extracted while considering weak field situations; when dealing with extreme situations (geometries with horizons, cosmological solutions, etc.) there still might be some way to distinguish the two interpretations because some solutions might be forbidden. Therefore, the structures of unimodular gravity and general relativity do appear naturally without recouring to curved spacetime notions, but precisely because of this they appear in a form that does not demand a geometrical interpretation in terms of a unique metric. Their form does not demand either the supplementation of the action with additional surface terms to build the Einstein-Hilbert action. The geometrical interpretation is certainly appealing as, on the one hand, it provides a natural interpretation of the absence of a local meaning for the gravitational energy (a gauge dependent quantity) and, on the other hand, it makes the theory self-contained, with no externally fixed elements. However, here we adhere to Rosen's comment more than 60 years ago

[2]: “Perhaps this (flat spacetime interpretation) may be regarded by some as a step backward. It should be noted, however, that this geometrization referred to has never been extended satisfactorily to other branches of physics, so that gravitation is treated differently from other phenomena. It is therefore not unreasonable to wonder whether it may not be better to give up the geometrical approach to gravitation for the sake of obtaining a more uniform treatment for all the various fields of force that are to be found in nature.” Then, one would postpone to a later stage the inquiry about the very nature of the background vacuum and its interconnections with the rest of the physical system. Let us stress that this does not need to lead to a bimetric theory in the sense of having two dynamical Lorentzian metrics, as even the existence of a preferred flat background can be an effective feature.

- (iii) The problem of not having a well-defined local notion of energy for the spin-2 field is already present in the linear theory, so it is not something that emerges due to the self-interactions. On the contrary, the presence of self-interactions solves the problem. The solution presents itself as two mutually exclusive mechanisms. In mechanism (a) self-interactions preserve the amount of gauge invariance and also the gauge noninvariance of the graviton stress-energy tensor. This gauge noninvariance finds a satisfactory explanation within the geometrical interpretation. In mechanism (b) self-interactions break gauge invariance so that the graviton stress-energy tensor acquires a well-defined meaning. These two alternatives conform with in principle distinct theories.
- (iv) From the point of view of self-coupling, there is no compulsory reason for the global vacuum energy to gravitate. We have recovered that the minimal gauge construction of the graviton field leads to the structure of unimodular gravity, a theory that differs from general relativity in that only the traceless part of the total stress-energy tensor enters the field equations. Any vacuum-energy contribution in the form of a cosmological constant becomes decoupled from geometry, although the structure of the theory still permits the addition of a cosmological constant as a phenomenological integration constant unrelated to the physical vacuum energy (see e.g., [42] for a discussion). In particular, this means that one does not need to consider curved backgrounds as the starting point of the iterative procedure, instead of Minkowski spacetime, to directly obtain a nonzero cosmological constant as it was done in [43]. However, when using the Fierz-Pauli gauge-extended version, the global vacuum energy does appear in the geometrical equations. Thus, although

it is interesting that the minimal field-theoretical realization of the concept of graviton leads through the self-coupling procedure to a degravitation of the vacuum energy (an idea which was first considered in [18]), there is no definite answer in this formalism concerning this problem. One would need to consider specific models to find a definite answer; for instance, it would be interesting to study in detail the situation in string theory [44]. For us, the important lesson we can draw from the discussion presented here of the self-coupling problem is that there exists room for a natural solution to the first cosmological constant problem in theories of emergent gravity, along the lines of what is proposed in [38,45].

Notice that, as claimed by Ellis, one can safely admit that observations prioritize unimodular gravity rather than general relativity [42]. Whereas in a geometrical interpretation unimodular gravity seems rather unnatural in contrast to general relativity, it is perfectly natural in the field-theoretical approach and, what is probably more interesting, its occurrence is tied up to a nonperfect decoupling from the background structure in the sense of emergent gravity scenarios [33]. This may be taken as a hint in favor of the field-theoretical approach to quantum gravity or, in a broader sense, to emergent gravity proposals incorporating in some low-energy regime a nonperfect decoupling from an underlying flat spacetime [21].

ACKNOWLEDGMENTS

Financial support was provided by the Spanish MICINN through Projects No. FIS2011-30145-C03-01 and No. FIS2011-30145-C03-02 (with FEDER contribution), and by the Junta de Andalucía through Project No. FQM219. R. C-R. acknowledges support from CSIC through the JAE-predoc program, cofunded by FSE.

APPENDIX: AN ALGEBRAIC IDENTITY

If we evaluate the derivative with respect to the auxiliary metric and forget momentarily about the symmetrization in the pair (s, t) , we can write (50) as

$$\begin{aligned} D^{a\nu\rho}{}_{\mu e b}(\gamma_{ps}\gamma_{vi}\gamma_{q\rho}\gamma^{de} + \gamma_{pv}\gamma_{qs}\gamma_{ri}\gamma^{de} - \gamma_{pv}\gamma_{q\rho}\delta_s^d\delta_i^e) \\ = \gamma_{pv}\gamma_{q\rho}\gamma_{i\beta}\gamma^{d\delta}D^{a\nu\rho}{}_{\theta se}D^{\theta\beta e}{}_{\mu\delta b}. \end{aligned} \quad (\text{A1})$$

Of course, this equation would only be valid when the terms obtained under the exchange $s \leftrightarrow t$ are added. In the following we are going to show that this equation holds. The left-hand side is easier to evaluate; it is composed by three terms:

$$\begin{aligned}
& \frac{1}{2} \gamma_{ps} \gamma_{\nu i} \gamma_{q\rho} \gamma^{de} (\delta_{\mu}^a \delta_b^{\nu} \delta_e^{\rho} + \delta_{\mu}^a \delta_b^{\rho} \delta_e^{\nu} + \delta_b^a \delta_{\mu}^{\nu} \delta_e^{\rho} + \delta_b^a \delta_{\mu}^{\rho} \delta_e^{\nu} - \delta_e^a \delta_b^{\nu} \delta_{\mu}^{\rho} - \delta_e^a \delta_b^{\rho} \delta_{\mu}^{\nu}) \\
& + \frac{1}{2} \gamma_{p\nu} \gamma_{qs} \gamma_{\rho t} \gamma^{de} (\delta_{\mu}^a \delta_b^{\nu} \delta_e^{\rho} + \delta_{\mu}^a \delta_b^{\rho} \delta_e^{\nu} + \delta_b^a \delta_{\mu}^{\nu} \delta_e^{\rho} + \delta_b^a \delta_{\mu}^{\rho} \delta_e^{\nu} - \delta_e^a \delta_b^{\nu} \delta_{\mu}^{\rho} - \delta_e^a \delta_b^{\rho} \delta_{\mu}^{\nu}) \\
& - \frac{1}{2} \gamma_{p\nu} \gamma_{q\rho} \delta_s^d \delta_t^e (\delta_{\mu}^a \delta_b^{\nu} \delta_e^{\rho} + \delta_{\mu}^a \delta_b^{\rho} \delta_e^{\nu} + \delta_b^a \delta_{\mu}^{\nu} \delta_e^{\rho} + \delta_b^a \delta_{\mu}^{\rho} \delta_e^{\nu} - \delta_e^a \delta_b^{\nu} \delta_{\mu}^{\rho} - \delta_e^a \delta_b^{\rho} \delta_{\mu}^{\nu}).
\end{aligned} \tag{A2}$$

The first six terms are

$$\gamma_{ps} \gamma_{bt} \delta_{\mu}^a \delta_q^d + \gamma_{ps} \gamma_{qb} \delta_{\mu}^a \delta_t^d + \gamma_{ps} \gamma_{t\mu} \delta_b^a \delta_q^d + \gamma_{ps} \gamma_{q\mu} \delta_b^a \delta_t^d - \gamma_{ps} \gamma_{q\mu} \gamma_{bt} \gamma^{ad} - \gamma_{ps} \gamma_{qb} \gamma_{t\mu} \gamma^{ad}. \tag{A3}$$

These are followed by the following six terms:

$$\gamma_{pb} \gamma_{qs} \delta_{\mu}^a \delta_t^d + \gamma_{qs} \gamma_{bt} \delta_{\mu}^a \delta_p^d + \gamma_{p\mu} \gamma_{qs} \delta_b^a \delta_t^d + \gamma_{qs} \gamma_{t\mu} \delta_b^a \delta_p^d - \gamma_{p\mu} \gamma_{qs} \gamma_{bt} \gamma^{ad} - \gamma_{pb} \gamma_{qs} \gamma_{t\mu} \gamma^{ad}. \tag{A4}$$

The last six terms are

$$-\gamma_{pb} \gamma_{qt} \delta_{\mu}^a \delta_s^d - \gamma_{pt} \gamma_{qb} \delta_{\mu}^a \delta_s^d - \gamma_{p\mu} \gamma_{qt} \delta_s^a \delta_b^d - \gamma_{pt} \gamma_{q\mu} \delta_b^a \delta_s^d + \gamma_{p\mu} \gamma_{qb} \delta_s^a \delta_t^d + \gamma_{pb} \gamma_{q\mu} \delta_t^a \delta_s^d. \tag{A5}$$

So we finish with the following expression symmetric in $p \leftrightarrow q$:

$$\begin{aligned}
& \frac{1}{2} (\gamma_{ps} \gamma_{bt} \delta_{\mu}^a \delta_q^d + \gamma_{qs} \gamma_{bt} \delta_{\mu}^a \delta_p^d + \gamma_{ps} \gamma_{qb} \delta_{\mu}^a \delta_t^d + \gamma_{qs} \gamma_{pb} \delta_{\mu}^a \delta_t^d + \gamma_{ps} \gamma_{t\mu} \delta_b^a \delta_q^d + \gamma_{qs} \gamma_{t\mu} \delta_b^a \delta_p^d + \gamma_{ps} \gamma_{q\mu} \delta_b^a \delta_t^d \\
& + \gamma_{qs} \gamma_{p\mu} \delta_b^a \delta_t^d - \gamma_{ps} \gamma_{q\mu} \gamma_{bt} \gamma^{ad} - \gamma_{qs} \gamma_{p\mu} \gamma_{bt} \gamma^{ad} - \gamma_{ps} \gamma_{qb} \gamma_{t\mu} \gamma^{ad} - \gamma_{qs} \gamma_{pb} \gamma_{t\mu} \gamma^{ad} + \gamma_{p\mu} \gamma_{qb} \delta_s^a \delta_t^d \\
& + \gamma_{q\mu} \gamma_{pb} \delta_t^a \delta_s^d - \gamma_{pb} \gamma_{qt} \delta_{\mu}^a \delta_s^d - \gamma_{pt} \gamma_{qb} \delta_{\mu}^a \delta_s^d - \gamma_{p\mu} \gamma_{qt} \delta_s^a \delta_b^d - \gamma_{q\mu} \gamma_{pt} \delta_b^a \delta_s^d).
\end{aligned} \tag{A6}$$

This expression must be still symmetrized under $s \leftrightarrow t$. When one does this some of the terms cancel,

$$3 \leftrightarrow 16, \quad 4 \leftrightarrow 15, \quad 7 \leftrightarrow 18, \quad 8 \leftrightarrow 17, \tag{A7}$$

leaving the simplified result:

$$\begin{aligned}
& \frac{1}{2} (\gamma_{ps} \gamma_{bt} \delta_{\mu}^a \delta_q^d + \gamma_{qs} \gamma_{bt} \delta_{\mu}^a \delta_p^d + \gamma_{ps} \gamma_{t\mu} \delta_b^a \delta_q^d + \gamma_{qs} \gamma_{t\mu} \delta_b^a \delta_p^d - \gamma_{ps} \gamma_{q\mu} \gamma_{bt} \gamma^{ad} - \gamma_{qs} \gamma_{p\mu} \gamma_{bt} \gamma^{ad} \\
& - \gamma_{ps} \gamma_{qb} \gamma_{t\mu} \gamma^{ad} - \gamma_{qs} \gamma_{pb} \gamma_{t\mu} \gamma^{ad} + \gamma_{p\mu} \gamma_{qb} \delta_s^a \delta_t^d + \gamma_{q\mu} \gamma_{pb} \delta_t^a \delta_s^d)_{s \leftrightarrow t}.
\end{aligned} \tag{A8}$$

This is the equation which we must compare with the right-hand side of (A1). In this side, there are 36 terms in total,

$$\begin{aligned}
& \frac{1}{4} \gamma_{p\nu} \gamma_{q\rho} \gamma_{t\beta} \gamma^{d\delta} (\delta_{\mu}^a \delta_s^{\nu} \delta_e^{\rho} + \delta_{\mu}^a \delta_s^{\rho} \delta_e^{\nu} + \delta_s^a \delta_{\mu}^{\nu} \delta_e^{\rho} + \delta_s^a \delta_{\mu}^{\rho} \delta_e^{\nu} - \delta_e^a \delta_{\mu}^{\nu} \delta_s^{\rho} - \delta_e^a \delta_{\mu}^{\rho} \delta_s^{\nu}) \\
& \times (\delta_{\mu}^{\theta} \delta_b^e \delta_{\delta}^{\beta} + \delta_{\mu}^{\theta} \delta_b^{\beta} \delta_{\delta}^e + \delta_b^{\theta} \delta_{\mu}^e \delta_{\delta}^{\beta} + \delta_b^{\theta} \delta_{\mu}^{\beta} \delta_{\delta}^e - \delta_{\delta}^{\theta} \delta_{\mu}^e \delta_b^{\beta} - \delta_{\delta}^{\theta} \delta_{\mu}^{\beta} \delta_b^e).
\end{aligned} \tag{A9}$$

The 36 terms are given by (the following expression is multiplied by 1/4)

$$\begin{aligned}
& \gamma_{ps} \gamma_{qb} \delta_{\mu}^a \delta_t^d + \gamma_{ps} \gamma_{bt} \delta_{\mu}^a \delta_q^d + \gamma_{ps} \gamma_{q\mu} \delta_b^a \delta_t^d + \gamma_{ps} \gamma_{t\mu} \delta_b^a \delta_q^d - \gamma_{ps} \gamma_{q\mu} \gamma_{bt} \gamma^{ad} - \gamma_{ps} \gamma_{qb} \gamma_{t\mu} \gamma^{ad} + \gamma_{pb} \gamma_{qs} \delta_{\mu}^a \delta_t^d + \gamma_{qs} \gamma_{bt} \delta_{\mu}^a \delta_p^d \\
& + \gamma_{p\mu} \gamma_{qs} \delta_b^a \delta_t^d + \gamma_{qs} \gamma_{t\mu} \delta_b^a \delta_p^d - \gamma_{p\mu} \gamma_{qs} \gamma_{bt} \gamma^{ad} - \gamma_{pb} \gamma_{qs} \gamma_{t\mu} \gamma^{ad} + \gamma_{p\mu} \gamma_{qb} \delta_s^a \delta_t^d + \gamma_{p\mu} \gamma_{bt} \delta_s^a \delta_q^d + \gamma_{pb} \gamma_{q\mu} \delta_s^a \delta_t^d \\
& + \gamma_{pb} \gamma_{t\mu} \delta_s^a \delta_q^d - \gamma_{q\mu} \gamma_{tb} \delta_s^a \delta_p^d - \gamma_{qb} \gamma_{t\mu} \delta_s^a \delta_p^d + \gamma_{pb} \gamma_{q\mu} \delta_s^a \delta_t^d + \gamma_{q\mu} \gamma_{bt} \delta_s^a \delta_p^d + \gamma_{p\mu} \gamma_{qb} \delta_s^a \delta_t^d + \gamma_{qb} \gamma_{t\mu} \delta_s^a \delta_p^d \\
& - \gamma_{p\mu} \gamma_{bt} \delta_s^a \delta_q^d - \gamma_{pb} \gamma_{t\mu} \delta_s^a \delta_q^d - \gamma_{p\mu} \gamma_{qs} \delta_b^a \delta_t^d - \gamma_{p\mu} \gamma_{qs} \gamma_{bt} \gamma^{ad} - \gamma_{pb} \gamma_{qs} \delta_{\mu}^a \delta_t^d - \gamma_{pb} \gamma_{qs} \gamma_{t\mu} \gamma^{ad} + \gamma_{qs} \gamma_{bt} \delta_{\mu}^a \gamma_p^d \\
& + \gamma_{qs} \gamma_{t\mu} \delta_b^a \delta_p^d - \gamma_{ps} \gamma_{q\mu} \delta_b^a \delta_t^d - \gamma_{ps} \gamma_{q\mu} \gamma_{bt} \gamma^{ad} - \gamma_{ps} \gamma_{qb} \delta_{\mu}^a \delta_t^d - \gamma_{ps} \gamma_{qb} \gamma_{t\mu} \gamma^{ad} + \gamma_{ps} \gamma_{bt} \delta_{\mu}^a \delta_q^d + \gamma_{ps} \gamma_{t\mu} \delta_b^a \delta_q^d.
\end{aligned} \tag{A10}$$

There are several terms which cancel:

$$1 - 33, \quad 3 - 31, \quad 7 - 27, \quad 9 - 25, \quad 14 - 23, \quad 16 - 24, \quad 17 - 20, \quad 18 - 22. \tag{A11}$$

The 20 remaining terms are paired, and they correspond to the 10 terms in (A8):

$$\begin{aligned} 2 + 35 \sim 1, & \quad 4 + 36 \sim 3, & \quad 5 + 32 \sim 5, & \quad 6 + 34 \sim 7, & \quad 8 + 29 \sim 2, \\ 10 + 30 \sim 4, & \quad 11 + 26 \sim 6, & \quad 12 + 28 \sim 8, & \quad 13 + 21 \sim 9, & \quad 15 + 19 \sim 10. \end{aligned} \quad (\text{A12})$$

-
- [1] N. Rosen, *Phys. Rev.* **57**, 147 (1940).
 [2] N. Rosen, *Phys. Rev.* **57**, 150 (1940).
 [3] S. N. Gupta, *Phys. Rev.* **96**, 1683 (1954).
 [4] R. H. Kraichnan, *Phys. Rev.* **98**, 1118 (1955).
 [5] R. P. Feynman, F. B. Morinigo, W. G. Wagner, and B. Hatfield, *Feynman Lectures on Gravitation*, Advanced Book Program (Westview Press, Boulder, CO, 2002).
 [6] S. Deser, *Gen. Relativ. Gravit.* **1**, 9 (1970).
 [7] E. R. Huggins, Ph. D. thesis, California Institute of Technology, 1962; <http://resolver.caltech.edu/CaltechTHESIS:08182011-085110085>.
 [8] T. Padmanabhan, *Int. J. Mod. Phys. D* **17**, 367 (2008).
 [9] L. M. Butcher, M. Hobson, and A. Lasenby, *Phys. Rev. D* **80**, 084014 (2009).
 [10] S. Deser, *Gen. Relativ. Gravit.* **42**, 641 (2010).
 [11] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Part 3.
 [12] E. Wigner, *Ann. Math.* **40**, 149 (1939).
 [13] F. Loebbert, *Ann. Phys. (Amsterdam)* **17**, 803 (2008).
 [14] S. Weinberg, *The Quantum Theory of Fields*, in The Quantum Theory of Fields 3 Vol. 2 (Cambridge University Press, Cambridge, England, 1996).
 [15] A. Jenkins, Ph. D. thesis, California Institute of Technology, 2006; <http://resolver.caltech.edu/CaltechETD:etd-06022006-145211>.
 [16] V. I. Ogievetsky and I. V. Polubarinov, *Ann. Phys. (N.Y.)* **35**, 167 (1965).
 [17] B. Zwiebach, *A First Course in String Theory* (Cambridge University Press, Cambridge, England, 2004).
 [18] J. J. van der Bij, H. van Dam, and Y. J. Ng, *Physica (Amsterdam)* **116A**, 307 (1982).
 [19] M. Fierz and W. Pauli, *Proc. R. Soc. A* **173**, 211 (1939).
 [20] T. Ortín, *Gravity and Strings*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2007).
 [21] C. Barceló, L. J. Garay, and G. Jannes, *Found. Phys.* **41**, 1532 (2011).
 [22] C. Aragone and S. Deser, *Nuovo Cimento Soc. Ital. Fis. A* **3**, 709 (1971).
 [23] G. Magnano and L. M. Sokolowski, *Classical Quantum Gravity* **19**, 223 (2002).
 [24] S. Weinberg and E. Witten, *Phys. Lett.* **96B**, 59 (1980).
 [25] S. Weinberg, *Phys. Lett.* **9**, 357 (1964).
 [26] S. Weinberg, *Phys. Rev.* **135**, B1049 (1964).
 [27] F. J. Belinfante, *Physica (Utrecht)* **7**, 449 (1940).
 [28] L. Rosenfeld, *Mem. Acad. R. Belg. Sci.* **18**, 1 (1940).
 [29] J. Fang and C. Fronsdal, *J. Math. Phys. (N.Y.)* **20**, 2264 (1979).
 [30] R. M. Wald, *Phys. Rev. D* **33**, 3613 (1986).
 [31] R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, IL, 2010).
 [32] T. Padmanabhan, *Gravitation: Foundations and Frontiers* (Cambridge University Press, Cambridge, England, 2010).
 [33] S. Carlip, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **46**, 200 (2014).
 [34] L. Bombelli, W. E. Couch, and R. J. Torrence, *Phys. Rev. D* **44**, 2589 (1991).
 [35] L. Smolin, *Phys. Rev. D* **80**, 084003 (2009).
 [36] D. Blas, *J. Phys. A* **40**, 6965 (2007).
 [37] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972).
 [38] G. F. R. Ellis, H. van Elst, J. Murugan, and J.-P. Uzan, *Classical Quantum Gravity* **28**, 225007 (2011).
 [39] F. I. Cooperstock and M. J. Dupre, *Ann. Phys. (Amsterdam)* **339**, 531 (2013).
 [40] M. J. Dupré, [arXiv:0903.5225](https://arxiv.org/abs/0903.5225).
 [41] C. Barceló, S. Liberati, and M. Visser, *Living Rev. Relativity* **8**, 12 (2005).
 [42] G. F. R. Ellis, *Gen. Relativ. Gravit.* **46**, 1619 (2013).
 [43] S. Deser, *Classical Quantum Gravity* **4**, L99 (1987).
 [44] E. Alvarez, *J. High Energy Phys.* **03** (2005) 002.
 [45] T. Padmanabhan, *Gen. Relativ. Gravit.* **46**, 1673 (2014).