The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2354-4

Summer School on Cosmology

16 - 27 July 2012

Infrared modifcations of gravity

A. Nicolis
Columbia, New York

Infrared modifications of gravity

Alberto Nicolis

Department of Physics and ISCAP,
Columbia University, New York, NY 10027, USA

nicolis@phys.columbia.edu

Introduction and motivation

These lectures are about theories that modify the long-distance behavior of general relativity (GR). Given the successes of GR as a theory of gravity, one wonders why we should try to modify it at all. One possible answer is that observations tell us that something unexpected happens at large scales: galactic rotation curves do not match what one infers from applying Newtonian gravity to the visible mass distribution, and the acceleration of the universe's expansion rate does not match the basic attractive nature of Newtonian gravity. Both puzzles admit fairly conservative solutions within GR—dark matter for the former, and a cosmological constant (cc) or some other form of 'dark energy' for the latter. Yet, since so far we have detected these hypothetical dark substances only indirectly via their gravitational effects, it is fair to ask whether the correct resolution for the above puzzles lies instead in a different behavior of gravitational forces at scales much larger than solar system ones.

By now, the amount of indirect evidence for the existence of dark matter is so compelling that it is unlikely that dark matter will be replaced by modified gravity. So, the subject of long-distance ('infrared', or IR, in the following) modifications of gravity has been relegated to cosmological scales, as an alternative to the cc or dark energy for explaining the recent acceleration of the Hubble expansion. Which brings us to a second, perhaps more relevant, motivation for trying to modify the IR behavior of gravity: the cosmological constant problem, which can be phrased as "We should not be surprised that the universe is accelerating—we should be surprised that it is accelerating so incredibly slowly". The cc can be thought of as the energy density of the vacuum. In quantum mechanics, there are zero-point energies that contribute to such an energy density. Quantum field theory (QFT)—our most precise understanding of particle physics—associates each Fourier mode of each field with a quantum mechanical oscillator, which contributes to the cc with its zero-point energy. Summing up all these contributions yields a conservative estimate for the quantum corrections to the cc of order $\Delta\Lambda \sim \text{TeV}^4$. [1] This should be added to the 'bare' cc Λ_{bare}—

[1] Such an estimate is conservative because it cuts off all contributions at energies of order TeV, which is roughly the maximum energy scale at which we have directly tested particle physics. With (reasonable) assumptions about the spectrum of beyond-the-standard-model particles, one can go much higher in energy, and obtain a correspondingly higher estimate for $\Delta\Lambda$.

that appearing in the Lagrangian before taking into account quantum corrections—to yield the physical, measurable value of the cc. Although technically the bare cc can be adjusted at will, there is no known physical reason why its value should be so precisely tuned to cancel the bulk of $\Delta\Lambda$ almost perfectly, and leave us with the incredibly small

$$\Lambda_{\rm phys} = \Lambda_{\rm bare} + \Delta\Lambda \sim (10^{-3}\ {\rm eV})^4\ , \tag{0.1}$$

which is the observed value, and is of course infinitesimal ($\sim 10^{-60}$) compared to the quantum contribution $\Delta\Lambda$. The cc problem is so extreme and so robust within GR coupled to quantum field theory, that no solution has been found so far apart from 'environmental' (or 'anthropic') ones. Can IR modifications of gravity shed some light on it? For instance, is it possible that the real cc is indeed huge as estimated from QFT considerations, but that it does not gravitate? Although from a theoretical perspective the cc problem probably remains the best motivation to study IR modifications of gravity, it is fair to say no substantial progress has been made so far in this direction. So, for most studies the working assumption has been that the cc vanishes exactly because of some unknown physical reason—like for instance a very powerful symmetry we have not been able to formulate yet—and the job left for modified gravity is to explain the observed acceleration of the universe in the absence of a cc.

In hindsight, by trying to make sense of modified gravity theories in the last decade or so, we have learned a number of properties of certain effective field theories that are relevant per se, regardless of whether they have anything to do with the acceleration of the universe. (For instance, some of these properties turned out to be useful in a recent proof of the a-theorem for generic renormalization-group flows in QFT [1].) This is an a posteriori motivation that perhaps tells us that we should not be too concerned about finding convincing a priori ones to study what we find interesting. . . And in a sense this is the most natural motivation for trying to modify gravity in the IR: it is a difficult theoretical problem, with many interesting physical aspects.

In these lectures I will make no attempt to be exhaustive about all the attempts that have been made. Instead, I will consider a few proposals that have been put forward. For these, I will focus mainly on their theoretical aspects, both because they are quite peculiar and remarkable, and because being fastidious about theoretical subtleties can actually be *useful*, as the cautionary tale of sect. 1 illustrates.

Definition of the problem, or lack thereof

Roughly speaking, the problem we want to address can be phrased as: *Can one change the IR behavior of gravity in order to have corrections to cosmology of order one, while passing solar system tests?* And if one succeeds, can these "corrections to cosmology" include acceleration in the absence of a cc or dark energy?

Before proceeding, we have to qualify what we mean by "changing gravity". One popular cartoon definition is that dark energy is a correction to the r.h.s. of Einstein's equations, while modified gravity changes the l.h.s. Of course this is quite ambiguous, since one can always interpret a correction to l.h.s. as minus the same corrections to the r.h.s., and vice versa. This is particularly evident in the case of the cc, since its manifestation in Einstein's equations,

$$\Lambda \, g_{\mu\nu} \; , \tag{0.2}$$

involves no degrees of freedom but the metric: it is not clear why it should be interpreted as a correction to $T_{\mu\nu}$ rather than to $G_{\mu\nu}$.

A more precise definition focuses on the propagating degrees of freedom—gravitational waves, or gravitons, in the quantum theory. We say that we have an IR modification of gravity if the propagation of gravitational waves is affected at distances and time-scales that can be parametrically smaller than the curvature scale. To appreciate why this is a sensible definition, we should analyze what happens in systems that usually we would not characterize as "modified gravity". Consider for instance an homogeneous, non-relativistic fluid coupled to GR. It features Jeans instability, which can be thought of as a modification in the propagation of perturbations. Which perturbations? We have two tensor modes, which we can think of as the usual gravitational waves, and one scalar mode, which is a coupled oscillation that involves both density perturbations in the fluid and perturbations in a gravitational degree of freedom—the Newtonian potential—and which is the mode associated with the Jeans instability. So, the propagation of gravitational perturbations *is* affected by the presence of the fluid. But the critical frequency at which this modification becomes relevant is of order of the Hubble scale—flat space is not a solution in the presence of an homogeneous fluid, and the universe gets curved into an FRW solution. So, in this particular example, one cannot disentangle the modifications to the propagation of gravitational perturbations from the effects due to curvature (and related to this, Jeans instability does not develop exponentially, but only in a power-law fashion). One can define a genuine "modification of gravity" as a system in which such a disentanglement is possible. For instance, if gravitational waves were massive in Minkowski space, one would see interesting modifications in their propagation at frequencies of order of their mass, and no modifications due to curvature, which vanishes for Minkowski.

Yet another possible definition—more in line with the viewpoint we will adopt below—says that gravity is modified, if the deviations from GR can be attributed, at least in some distance range, to a new degree of freedom whose gravitational backreaction is negligible with respect to the new force it mediates. In other words: we want to actually modify the gravitational *force* between particles, rather than simply infuse yet another substance into the universe in order to curve it on cosmological scales. Implicit in this definition there is the assumption that, to be part of 'gravity', this new force should be universal, i.e., it should obey the equivalence principle. So, for instance, the electromagnetic force does not count as a modification of gravity.

I hope it is clear that the definitions given above are totally arbitrary. They are useful, in that they focus on some of the desiderata of the theories we are after, but they are not sacrosant. They just mirror what we think an interesting modification of gravity should look like once we find it.

Notation, and a disclaimer

Throughout these notes, I have tried to be consistent with the following notation and conventions:

- Metric signature: $\eta_{\mu\nu} = \mathrm{diag}\,(-,+,+,+)$. Same as Wald's [2], Weinberg's [3], and Misner-Thorne-Wheeler's [4]. In particular, the correct, positive-energy sign for a scalar field's kinetic Lagrangian is *minus*:

$$\mathcal{L} = -\tfrac{1}{2}(\partial\varphi)^2 = \tfrac{1}{2}\,\dot{\varphi}^2 - \tfrac{1}{2}\,(\vec{\nabla}\varphi)^2\ .$$

- Riemann tensor:

$$R^{\alpha}{}_{\beta\gamma\delta} = \partial_\gamma \Gamma^\alpha_{\beta\delta} - \partial_\delta \Gamma^\alpha_{\beta\gamma} + \Gamma^\alpha_{\mu\gamma}\Gamma^\mu_{\beta\delta} - \Gamma^\alpha_{\mu\delta}\Gamma^\mu_{\beta\gamma}\ .$$

  Same as Misner-Thorne-Wheeler's; opposite of Weinberg's; same as Wald's with *opposite ordering* of indices, that is

$$\big(R^{\alpha}{}_{\beta\gamma\delta}\big)_{\rm here} = \big(R_{\delta\gamma\beta}{}^{\alpha}\big)_{\rm Wald}\ .$$

  The definition for the Christoffel symbols is

$$\Gamma^\mu_{\alpha\beta} = \tfrac{1}{2} g^{\mu\nu}\big(\partial_\alpha g_{\nu\beta} + \partial_\beta g_{\nu\alpha} - \partial_\nu g_{\alpha\beta}\big)\ ,$$

  which is pretty much universal.

- Einstein equations:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} = (8\pi G) T_{\mu\nu}\ .$$

  Same as Wald's and Misner-Thorne-Wheeler's. Opposite of Weinberg's.

- Energy density: $T_{00} = +\rho$. Universal.

- Natural units: $\hbar = 1$, $c = 1$.

- Planck scale: $M_{\rm Pl}^2 \equiv (8\pi G)^{-1}$.

However, I have not been particularly careful with factors of 2, π, etc. Nor have I been particularly careful about signs, apart from a few, crucial ones. Please let me know of any inconsistencies you may find.

Contents

1 GR from particle physics

Before trying to modify it, let us rephrase GR in particle physics terms. Actually, let us abandon everything we know about GR, starting with its celebrated geometric interpretation, and let us re-derive GR from scratch, by applying the tenets of relativistic QFT. This experiment was carried out in the 1960's, initiated by Weinberg and completed by Boulware and Deser. The result is simply stated: *GR arises as the only consistent, Poincaré-invariant, low-energy theory of interacting massless spin-2 particles.* The importance of such a statement is hardly overstated. The only assumptions are that gravity is mediated by massless spin-2 particles (gravitons), and that these obey special relativity and quantum mechanics. Then, GR follows. In particular, the equivalence principle and its beautiful (and invaluable) geometric interpretation are forced upon us. I will sketch here the essential logical steps.

1.1 Spin versus helicity

In a Lorentz-invariant theory of particle physics, particles are characterized by their mass and their spin. For massive particles, it is fairly intuitive what we mean by the spin degrees of freedom: for any given state of the particle, we can go to its rest frame, where we can use standard non-relativistic QM know-how to characterize spin. In particular, for a particle of total spin s, we have $(2s + 1)$ states, corresponding to different eigenvalues of J_z in the particle's rest frame. Then, relativity tells us how to 'boost' these spin degrees of freedom to any other reference frame. For massless particles the situation is much less intuitive. We cannot go to their rest frame—they always move at the speed of light, and there is no reference frame in which we can use our non-relativistic QM intuition. Related to this, there is a quantity—the *helicity* $h = \vec{J} \cdot \hat{p}$—which is invariant under all Lorentz transformations (besides the total spin J^2, there is no analogous quantity for a massive particle). It turns out that a massless particle of total spin s can have $h = \pm s$, and no other spin degrees of freedom. So, for $s = 0$, there is one state; for $s = \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots$ there are two. Already for $s = 1$, these are fewer than the $(2s + 1)$ one would have in the massive case. You may be tempted to relegate this fact to the realm of technical curiosities of QFT, but you shouldn't: as we will see, it is the physical origin of all we know and love (or fear) about GR.

1.2 Local field operators

Now, in relativistic QFT we describe particles via local covariant (scalar, spinor, vector, etc.) field operators. To do so, we need a finite-dimensional representation of the Lorentz group to use as polarization spinor, vector, etc. For instance, we know that to describe a massive spin-1 particle we need polarization four-vectors e^μ. Indeed, in non-relativistic QM the spin-1 representation of

$SO(3)$ is that of three-dimensional vectors, with the three J_z eigenstates being

$$
\begin{aligned}
\vec{e}_{-1} &= \tfrac{1}{\sqrt{2}}(1,-i,0) && (1.1)\\
\vec{e}_{0} &= (0,0,1) && (1.2)\\
\vec{e}_{+1} &= -\tfrac{1}{\sqrt{2}}(1,+i,0) && (1.3)
\end{aligned}
$$

So, in the rest-frame we define the polarization four-vector for a particle of polarization $\sigma = 0, \pm 1$ simply as

$$e^\mu_\sigma \equiv (0; \vec{e}_\sigma) \; , \tag{1.4}$$

and in any other frame as the boosted versions of these. Then, when we construct the field operator in the usual way, very schematically,

$$Z^\mu(x) \sim \sum_{\sigma=0,\pm1} e^\mu_\sigma(\vec{p}) a_\sigma(\vec{p})\, e^{i\,p\cdot x} + \text{h.c.} \; , \tag{1.5}$$

where $a_\sigma(\vec{p})$ is the annihilation operator for a particle of momentum $\vec{p}$ and spin σ, we get a nice covariant 4-vector field, because all ingredients transform as they should under Lorentz transformations. This means that we can construct Lorentz invariant Lagrangians for Z^μ just by contracting the indices in the usual way.

To appreciate how non-trivial it is that everything works out, now let's try to do the same for massless particles of spin 1—i.e., photons. We need polarization vectors to keep track of their spin dof, which now is just the helicity. We cannot define these polarization vectors in the rest frame—there is no rest frame. However, we just have to define two four-vectors that under rotations around the $\hat{p}$-axis transform into themselves times a phase $e^{\pm i\theta}$. This is the definition of having helicity ± 1. This is quite easy. For instance, for a particle moving along the z-axis,

$$p^\mu = (p, 0, 0, p) \; , \tag{1.6}$$

the desired four-vectors are simply

$$e^\mu_{\pm1}(\vec{p}) = \tfrac{1}{\sqrt{2}}(0, 1, \pm i, 0) \; . \tag{1.7}$$

For a more general propagation direction, the generalization is obvious. Notice however a disturbing feature of this definition: all these polarization vectors have vanishing time component,

$$e^0_{\pm1}(\vec{p}) \; . \tag{1.8}$$

This seems—and is—at odds with Lorentz invariance. Notice that in the massive case we had $e^0_\sigma = 0$ in one particular reference frame only—the particle's rest frame. In other reference frames the same statement takes a Lorentz-invariant form: $p_\mu e^\mu_\sigma = 0$. Here instead eq. (1.8) holds in an arbitrary frame, for all particles, of arbitrary momenta. Clearly something went wrong. What went wrong is quite simple: our insisting on calling the $e^\mu_{\pm1}(\vec{p})$ 'four-vectors' is unwarranted. Or equivalently, if we *decide* to transform them as four-vectors, we do not get what we think. It is straightforward to show that a generic Lorentz boost acting on the $e^\mu_{\pm1}$'s yields, up to an overall phase,

$$\Lambda^\mu{}_\nu \, e^\nu_{\pm1}(\vec{p}) = e^\mu_{\pm1}(\vec{\Lambda p}) + \alpha_\Lambda \, (\Lambda p)^\mu \; , \tag{1.9}$$

where α_Λ is a parameter that depends on the boost being performed. The first term is what you would have expected: the polarization vector for a particle with the same helicity as before moving with the new (boosted) momentum. The second piece is unexpected. It is proportional to the new four-momentum. That is, under Lorentz transformations, the naive polarization vectors that we defined above, get 'contaminated' with the four-momentum. If we pretend that nothing happened, and we go ahead and construct the photon field operator,

$$A^\mu(x) \sim \sum_{h=\pm1} e^\mu_h(\vec{p}) a_\sigma(\vec{p}) \, e^{i\,p\cdot x} + \text{h.c.} \; , \tag{1.10}$$

we discover that, despite our insisting on using a μ index, A^μ *does not transform covariantly under Lorentz transformations.* Under a generic boost one gets

$$A^\mu(x) \to (\Lambda^{-1})^\mu{}_\nu \, A^\nu(\Lambda \cdot x) + \partial^\mu \lambda_\Lambda(x) \tag{1.11}$$

where $\lambda_\Lambda(x)$ is some Λ-dependent linear combination (i.e., integral in d^3p with $\vec{p}$-dependent coefficients) of creation and annihilation operators whose explicit form we are not interested in. The first piece is what you would expect for a covariant vector field. The second piece tells you that $A^\mu(x)$ is *not* a covariant vector field.

With minor modifications, the same remarks apply to massless spin 2 particles, i.e. gravitons. Here we need two polarization *tensors* to describe the two helicity states. We can simply take

$$e^{\mu\nu}_{\pm2}(\vec{p}) = e^\mu_{\pm1}(\vec{p}) e^\nu_{\pm1}(\vec{p}) \; , \tag{1.12}$$

where $e^\mu_{\pm1}$ are the photon polarization four-vectors defined above. Since the latter are eigenstates of rotations around the $\hat{p}$-axis with helicity ±1, the $e^{\mu\nu}_{\pm2}$ have helicity ±2,

$$R_z(\theta)^\mu{}_\rho \, R_z(\theta)^\nu{}_\sigma \, e^{\rho\sigma}_{\pm2} = e^{\pm 2i\theta} e^{\mu\nu}_{\pm2} \; , \tag{1.13}$$

as desired. Explicitly, for a graviton propagating along the z-axis, they are

$$e^{\mu\nu}_{\pm2} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & \pm i & 0 \\ 0 & \pm i & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \; . \tag{1.14}$$

Notice that in terms of the standard 'plus' and 'cross' polarizations for gravitational waves, they read $e^{\mu\nu}_{\pm2} = \frac{1}{2}(e^{\mu\nu}_+ \pm i \, e^{\mu\nu}_\times)$. They correspond to circularly polarized gravitational waves.

Next, we naively construct the field operator as we did before

$$h^{\mu\nu}(x) \sim \sum_{h=\pm2} e^{\mu\nu}_h(\vec{p}) \, a_h(\vec{p}) \, e^{ip\cdot x} + \text{h.c.} \; , \tag{1.15}$$

and as before we realize that this is not a good Lorentz-tensor. Indeed from eq. (1.12) and from the properties of the photon polarization vectors (1.8) we see that

$$h^{\mu 0} = 0 \, , \tag{1.16}$$

in all reference frames—clearly a non Lorentz-invariant constraint *if* we assume that $h^{\mu\nu}$ is a tensor.

Since our polarization tensors are just products of the photon's helicity polarization four-vectors, we can use the results we reviewed above. Under a generic Lorentz transformation we have, up to an overall phase (see eq. (1.9)),

$$\begin{aligned} \Lambda^\mu{}_\rho \, \Lambda^\nu{}_\sigma \, e^{\rho\sigma}_{\pm 2}(\vec{p}) &= \big(e^\mu_{\pm 1}(\vec{\Lambda p}) + \alpha_\Lambda \, (\Lambda p)^\mu\big)\big(e^\nu_{\pm 1}(\vec{\Lambda p}) + \alpha_\Lambda \, (\Lambda p)^\nu\big) \\ &= e^{\mu\nu}_{\pm 2}(\vec{\Lambda p}) + (\Lambda p)^\mu v^\nu + (\Lambda p)^\nu v^\mu \, , \end{aligned} \tag{1.17}$$

where v^μ is a four-vector that depends on Λ and on $\vec{p}$,

$$v^\mu \equiv \alpha_\Lambda \, e^\mu_{\pm 1}(\vec{\Lambda p}) + \tfrac{1}{2}\alpha_\Lambda^2 \, (\Lambda p)^\mu \, . \tag{1.18}$$

As a consequence, the field operator transforms as

$$h^{\mu\nu}(x) \to (\Lambda^{-1})^\mu{}_\rho (\Lambda^{-1})^\nu{}_\sigma \, h^{\rho\sigma}(\Lambda x) + \partial^\mu \xi^\nu_\Lambda(x) + \partial^\nu \xi^\mu_\Lambda(x) \, , \tag{1.19}$$

where $\xi^\mu_\Lambda(x)$ is a Λ-dependent, four-vector linear combination of creation and annihilation operators. Despite its carrying $\mu\nu$ indices, the graviton field operator is not a covariant tensor under Lorentz transformations.

1.3 Gauge Invariance from Lorentz Invariance

Since—as we just found out—the photon and graviton field operators do not transform covariantly under Lorentz transformations, we cannot just stick them into a Lorentz invariant-*looking* Lagrangian to get a *truly* Lorentz-invariant theory. For instance, a photon quartic self-interaction like $(A_\mu A^\mu)^2$ looks Lorentz-invariant, but given the transformation properties of A^μ (1.11), it isn't. However, eq. (1.11) itself provides us with the solution to the problem. If, besides looking Lorentz-invariant, the Lagrangian is invariant under the formal replacement

$$A_\mu(x) \to A_\mu(x) + \partial_\mu \lambda(x) \, , \tag{1.20}$$

for a generic function $\lambda(x)$, then the extra piece in the Lorentz-transformation of A_μ is harmless, and the Lagrangian *is* Lorentz invariant. For instance, out of the photon field we can construct the 'electromagnetic field-strength' operator

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu \, . \tag{1.21}$$

This is unaffected by the replacement (1.20), and, as a consequence, it is a true Lorentz-tensor

$$F_{\mu\nu}(x) \to F_{\mu\nu}(x) = (\Lambda^{-1})_\mu{}^\rho (\Lambda^{-1})_\nu{}^\sigma F_{\rho\sigma}(\Lambda \cdot x) \; . \tag{1.22}$$

This means that now we can use $F_{\mu\nu}$ in a Lagrangian in the usual way, just by contracting indices in a Lorentz-invariant fashion. For instance, the quartic term $(F_{\mu\nu}F^{\mu\nu})^2$ now describes a truly Lorentz-invariant photon self-interaction. In addition to using $F_{\mu\nu}$, we can couple directly A_μ to a conserved 'current' $J^\mu(x)$—that is, a (genuine) four-vector made up of other fields (electron, etc.) that obeys a local conservation equation

$$\partial_\mu J^\mu = 0 \; . \tag{1.23}$$

If the Lagrangian term that describes the interaction between A_μ and these other fields is

$$\mathcal{L}_{\rm int} = A_\mu(x) J^\mu(x) \; , \tag{1.24}$$

then upon the replacement (1.20) this Lagrangian term in unchanged:

$$\mathcal{L}_{\rm int} \to \mathcal{L}_{\rm int} + \partial_\mu \lambda \, J^\mu = \mathcal{L}_{\rm int} - \lambda \, \partial_\mu J^\mu = \mathcal{L}_{\rm int} \; , \tag{1.25}$$

where we integrated by parts, and we used the conservation of J^μ. As a result, $\mathcal{L}_{\rm int}$ describes a Lorentz-invariant interaction.

Likewise, for the graviton field $h_{\mu\nu}$ we get a Lorentz-invariant theory if the Lagrangian has all indices contracted in a Lorentz invariant fashion *and* is also invariant under the replacement

$$h_{\mu\nu}(x) \to h_{\mu\nu}(x) + \partial_\mu \xi_\nu(x) + \partial_\nu \xi_\mu(x) \; , \tag{1.26}$$

for a generic four-vector field $\xi_\mu(x)$. This way, the non-covariant part in the transformation of $h_{\mu\nu}$ (1.19) has no effect on the Lagrangian, which then is Lorentz-invariant. Like for the photon, we can construct a true Lorentz-tensor by taking suitable derivatives of $h_{\mu\nu}$. The simplest is the (linearized) 'Riemann tensor':

$$R_{\mu\nu\rho\sigma} \equiv \tfrac{1}{2}(\partial_\mu \partial_\rho h_{\nu\sigma} - \partial_\nu \partial_\rho h_{\mu\sigma} - \partial_\mu \partial_\sigma h_{\nu\rho} + \partial_\nu \partial_\sigma h_{\mu\rho}) \; . \tag{1.27}$$

It is straightforward to check that this is unaffected by eq. (1.26). Then, it is a true tensor, for the non-covariant pieces in the Lorentz transformation of $h_{\mu\nu}$ cancel among different terms. We can use the Riemann tensor to build Lorentz invariant interactions for the graviton. Alternatively, just like we did for the photon, we can couple $h_{\mu\nu}$ directly to a *conserved symmetric* tensor $J^{\mu\nu}$: a symmetric true Lorentz tensor made up of other fields obeying the conservation law

$$\partial_\mu J^{\mu\nu} = 0 \; . \tag{1.28}$$

The interaction Lagrangian

$$\mathcal{L}_{\rm int} = h_{\mu\nu}(x) J^{\mu\nu}(x) \tag{1.29}$$

is invariant under the replacement (1.26),

$$\mathcal{L}_{\rm int} \to \mathcal{L}_{\rm int} + 2 \, \partial_\mu \xi_\nu \; J^{\mu\nu} = \mathcal{L}_{\rm int} - 2 \, \xi_\nu \; \partial_\mu J^{\mu\nu} = \mathcal{L}_{\rm int} \; , \tag{1.30}$$

and is therefore Lorentz-invariant. We will elaborate on this interaction Lagrangian below.

1.4 Remarks on gauge invariance

Of course invariance under the formal replacements (1.20, 1.26) is what usually deserves the name of 'gauge invariance', or 'diff-invariance', in the case of the graviton. From the constructive viewpoint we have taken here, it is clear that gauge invariance is not a symmetry on which we have any choice: it is forced upon us by combining quantum mechanics and Lorentz-invariance. In particular, if we want to describe massless particles through local field operators like $A_\mu(x)$ and $h_{\mu\nu}(x)$, then the theory is truly Lorentz invariant only if it is gauge-invariant—because these field operators do not transform covariantly under Lorentz transformations. However *gauge-invariance is not a symmetry in any physical sense*: unlike ordinary symmetries, gauge-invariance does not yield any non-trivial relations between different observables like e.g. different scattering amplitudes—because it acts trivially on the particles' states! More properly, gauge invariance should be thought of as a *redundancy* in our description of massless particles: field configurations that differ only by a gauge transformation, describe the same physical state.

We could dispose of gauge invariance altogether, if in our Lagrangian we never used A_μ and $h_{\mu\nu}$ alone, but only used truly covariant field operators like $F_{\mu\nu}$ and $R_{\mu\nu\rho\sigma}$. In fact these are true Lorentz tensors precisely because they are gauge invariant—so that all non-covariant pieces in the Lorentz-trasformation law of A_μ and $h_{\mu\nu}$ cancel. By using true tensor we would not need to introduce gauge tranformations at all. Any Lorentz invariant-looking Lagrangian would be truly Lorentz-invariant. Put another away: with gauge-invariant fields, and gauge-invariant states, gauge transformations would act trivially on everthing, so why bother introducing them?

It is useful to go back to the field expansion of A_μ (1.10), and see what it implies for $F_{\mu\nu}$,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \sim \sum_{\sigma=\pm 1} \big(ip_\mu \, e^\sigma_\nu(\vec p) - ip_\nu \, e^\sigma_\mu(\vec p) \big) a_\sigma(\vec p) \, e^{ip\cdot x} + \text{h.c.} \; . \tag{1.31}$$

That is, $F_{\mu\nu}$ describes a photon of helicity ± 1 and momentum $\vec p$ through the polarization tensor

$$ip_\mu \, e^{\pm 1}_\nu(\vec p) - ip_\nu \, e^{\pm 1}_\mu(\vec p) \; . \tag{1.32}$$

This is obviously an eigenstate of rotations around $\vec p$ with helicity ± 1, because the $e^{\pm 1}_\mu(\vec p)$ is, and p_μ is invariant. However now this polarization tensor *does* transform covariantly under Lorentz boosts—the unwanted terms in the transformation (1.9) proportional to $(\Lambda p)^\mu$ cancel between the two terms in (1.32). Hence, all ingredients that make up $F_{\mu\nu}$ transform as they should, and as a result $F_{\mu\nu}$ is a true tensor. Similar remarks hold for the field operator $R_{\mu\nu\rho\sigma}$, which employs the helicity ± 2 polarization tensors

$$\big(ip_\mu \, e^{\pm 1}_\nu(\vec p) - ip_\nu \, e^{\pm 1}_\mu(\vec p) \big) \big(ip_\rho \, e^{\pm 1}_\sigma(\vec p) - ip_\sigma \, e^{\pm 1}_\rho(\vec p) \big) \; . \tag{1.33}$$

Why not use the true tensors $F_{\mu\nu}$ and $R_{\mu\nu\rho\sigma}$ as *the* field operators for photons and gravitons, then? The problem is that the explicit powers of p^μ appearing in eqs. (1.32, 1.33) would make any amplitude for emitting (or absorbing) a photon or a graviton proportional to some powers of the energy of the emitted (or absorbed) particle. This becomes small at low energies, or, equivalently, at large distances. For instance, the Coulomb and Newton force would be proportional to $1/r^4$ and $1/r^6$, respectively. There would be nothing wrong in theory, but apparently Nature has chosen a different route.

1.5 GR

Let's now focus on the graviton, which is what interests us the most. In perturbation theory we can split the full action as

$$S = S_g^0[h_{\mu\nu}] + S_{\rm matter}[\psi] + S_{\rm int}[h_{\mu\nu}, \psi] \ , \tag{1.34}$$

where:

- S_g^0 is the free graviton action. It should be quadratic in $h_{\mu\nu}$, involve two derivatives, and be *truly* Lorentz invariant, i.e., Lorentz invariant-looking and gauge-invariant. The only form of S_g^0 compatible with these requirements is—not surprisingly—the quadratic truncation of the Einstein-Hilbert action.

- ψ denotes collectively the matter degrees of freedom, and $S_{\rm matter}[\psi]$ is their full action including all interactions except those with the gravitational field. We treat this part of the action as given. Think of it as the standard model (SM) Lagrangian, for instance.

- $S_{\rm int}[h_{\mu\nu}, \psi]$ includes gravitational interactions of matter as well as self-interactions of the gravitational field $h_{\mu\nu}$. For the moment, this part of the action is unknown. We want to determine it.

The action must be Lorentz invariant-looking and invariant under gauge transformations (1.20). S_g^0 and $S_{\rm matter}[\psi]$ are gauge-invariant by themselves—the former by construction, the latter because the matter fields ψ are unaffected by gauge transformations. As to the interaction action, we have

$$\begin{aligned} S_{\rm int}[h_{\mu\nu}, \psi] \ &\to \ S_{\rm int}[h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu, \psi] & (1.35) \\ &= \ S_{\rm int}[h_{\mu\nu}, \psi] + \int d^4x \, 2 \partial_\mu \xi_\nu \, \frac{\delta S_{\rm int}[h_{\mu\nu}, \psi]}{\delta h_{\mu\nu}} + \dots \ , & (1.36) \end{aligned}$$

where we only kept the first order in ξ^μ, and the factor of 2 comes from symmetry of the functional derivative w.r.t. $h_{\mu\nu}$. For the variation to vanish for all ξ^μ, we need

$$\partial_\mu J^{\mu\nu} = 0 \ , \qquad J^{\mu\nu}(x) \equiv \frac{\delta S_{\rm int}[h_{\mu\nu}, \psi]}{\delta h_{\mu\nu}(x)} \ . \tag{1.37}$$

So, $J^{\mu\nu}$ thus defined is symmetric and conserved. We know of another local tensor with the same properties—the stress-energy tensor $T^{\mu\nu}$. Are they the same? One can prove that in an interacting theory the only four-vector that is conserved by all non-trivial physical processes—like e.g. scattering events—is the total four-momentum, and that, as a consequence, $J^{\mu\nu}$ must be proportional to the stress-energy tensor, up to total derivative additions that are irrelevant at low energies/large distances.

In conclusion, the full action for gravity coupled to matter is eq. (1.34), where at low energies/large distances

$$\frac{\delta S_{\rm int}}{\delta h_{\mu\nu}} \simeq \alpha T^{\mu\nu}[h_{\mu\nu}, \psi] \ , \tag{1.38}$$

for some constant α. $T^{\mu\nu}$ is the (symmetric) *full* stress-energy tensor, conserved in the *ordinary, flat-space* sense:

$$\partial_\mu T^{\mu\nu} = 0 \ . \tag{1.39}$$

In particular, it must include the gravitational contributions to energy and momentum.

The problem of constructing the correct QFT for gravity, is thus reduced to finding the full stress-energy tensor for interacting matter fields and gravitons. In principle, one could do it perturbatively in the gravitational field:

0. At zeroth order, $T^{\mu\nu}$ is just that of matter, which we can derive straightforwardly from $S_{\rm matter}[\psi]$ via Noether's theorem (we have to symmetrize it, using the Belinfante procedure). This $T^{\mu\nu}$ acts as a linear source for $h^{\mu\nu}$, which develops a non-trivial solution.

1. We have to include this $h^{\mu\nu}$'s energy and momentum into $T^{\mu\nu}$. To be consistent, since the solution we have for $h_{\mu\nu}$ was obtained in the linearized approximation, we have to include in $T^{\mu\nu}$ terms up to second order in $h_{\mu\nu}$, schematically of the form hT and $\partial h \partial h$. Now the solution for $h^{\mu\nu}$ changes, because its source—$T^{\mu\nu}$—has changed.

2. And so on.

At every step, there is no ambiguity, because of the uniqueness of the stress-energy tensor: once all interactions up to n-th order are given, there is only one $T^{(n)}_{\mu\nu}$ that is conserved by them.

But now we can stop pretending that we do not know about GR: GR obeys *all* the properties that we have derived above for the graviton field $h_{\mu\nu}$. Therefore, given the uniqueness of the procedure sketched above, *the end result must be GR coupled to the ψ matter sector in the usual generally covariant way.* [2]

1.6 Moral

I quickly went through this derivation for several reasons: First, by elucidating what the assumptions that *unavoidably* lead to GR are, we now know what physical properties we have to give up if we want to modify gravity in the IR. We should do (at least) one of the following:

- Give mass to the gravitons.

[2] In practice, there is a more direct way of actually seeing GR emerge via a non-linear trick devised by Boulware and Deser. But the argument sketched above suffices to argue that the perturbative procedure is unambiguous at every step, and that the end result must therefore be unique.

- Introduce new degrees of freedom, that are not massless spin-2 particles.
- Break Lorentz invariance.

Second, this derivation shows how far one can get by applying systematically some theoretical rigor. I propose that we do the same for modified gravity. We should not ignore the consistency problems that we encounter—we should try to solve them. This might allow us to narrow down our search to a few (perhaps one? zero?) fully consistent candidate theories of modified gravity. Third—and this has nothing to do with modified gravity—this result is just an impressive triumph of theoretical physics, which puts the various aspects of gravity into perspective, and GR onto a much firmer ground, and I felt that all physicists interested in gravity should be aware of it. And finally, it is always useful to remind ourselves where the need for gauge- and diff-invariance fundamentally comes from, and how devoid of physical meaning they are—they are statements of redundancy, not of symmetry!

2 Massive gravity

The most obvious way we have to modify the long distance behavior of the gravitational force, is to give a mass to the particles that mediate it. This replaces the $1/r$ potential with a Yukawa-type potential, $1/r \cdot e^{-mr}$, which is exponentially weak at distances larger that the gravitons' Compton wavelength. As we will see, this is not the only effect, and most of the additional effects are problematic, as repeatedly pointed out during the subject's long history. Recent progress however suggests that the theory may be substantially healthier than what originally thought.

2.1 The Fierz-Pauli tuning

For a flat background metric, gravitons are described by a (Lorentz-) tensor field $h_{\mu\nu}$, defined by

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \; . \tag{2.1}$$

The Einstein-Hilbert action, expanded at second order in $h_{\mu\nu}$, [3]

$$S_{\rm EH} = \tfrac{1}{2} M_{\rm Pl}^2 \int d^4x \, \sqrt{-g} \, R \tag{2.2}$$

$$\simeq -\tfrac{1}{2} M_{\rm Pl}^2 \int d^4x \left[\partial_\lambda h_{\mu\nu} \partial^\lambda h^{\mu\nu} - 2 \partial_\mu h_{\nu\lambda} \partial^\nu h^{\mu\lambda} + 2 \partial_\mu h^{\mu\nu} \partial_\nu h - \partial_\mu h \partial^\mu h \right] , \tag{2.3}$$

can be thought of as a kinetic term for the gravitons—the analog of the photons' $-\frac{1}{2} F_{\mu\nu} F^{\mu\nu}$. To give the gravitons a mass, we add terms that are quadratic in $h_{\mu\nu}$, with no derivatives. There are two independent structures,

$$S_{\rm mass} = -\tfrac{1}{2} M_{\rm Pl}^2 \int d^4x \left[m_1^2 \, h_{\mu\nu} h^{\mu\nu} + m_2^2 \, h^2 \right] , \tag{2.4}$$

where the $M_{\rm Pl}^2$ upfront takes into account the non-canonical normalization of $h_{\mu\nu}$ —see eq. (2.3). As we will see in a second, the two mass parameters $m_{1,2}^2$ are not independent. We also assume that gravity is coupled to matter in the usual way—we want to preserve the equivalence principle—which at linear order in $h_{\mu\nu}$ reads

$$S_{\rm matter} = +\tfrac{1}{2} \int d^4x \, h_{\mu\nu} T^{\mu\nu} \; . \tag{2.5}$$

At the order we are working, the full action for $h_{\mu\nu}$ is

$$S = S_{\rm EH} + S_{\rm mass} + S_{\rm matter} \; . \tag{2.6}$$

The linearized equations of motion (eom) follow from varying the action w.r.t. to $h_{\mu\nu}$. We get

$$M_{\rm Pl}^2 \big[G^L_{\mu\nu} + m_1^2 h_{\mu\nu} + m_2^2 \, h \, \eta_{\mu\nu} \big] = T_{\mu\nu} \; , \tag{2.7}$$

[3] Indices are raised and lowered via $\eta_{\mu\nu}$, and $h \equiv h^\mu{}_\mu$.

where $G^L_{\mu\nu}$ is the linearized Einstein tensor.

It is instructive to compare this eom to what we would have for a massive spin-1 field Z_μ,

$$\partial_\mu F^{\mu\nu} + m^2 Z^\nu = J^\nu \ , \tag{2.8}$$

where $F_{\mu\nu} \equiv \partial_\mu Z_\nu - \partial_\nu Z_\mu$, and J^ν is a current that sources our Z^μ field. A massive spin-1 particle has three independent degrees of freedom, which should show up in the source-free eom as three wave solutions with independent polarizations. The 4-vector Z^μ has four components—apparently, one too many degrees of freedom. However, if we take the divergence of the source-free eom, because $F^{\mu\nu}$ is anti-symmetric, we get

$$m^2 \partial_\mu Z^\mu = 0 \tag{2.9}$$

For nonzero m^2, this acts as a constraint on Z^μ: the polarization vector ϵ^μ for a wave with momentum k^μ has to be transverse, $\epsilon^\mu k_\mu = 0$. This lowers the number of degrees of freedom from four to three—the correct number.

We can now go back to our massive gravity eom (2.7) and check whether we have the correct number of independent wave solutions. A massive spin-2 particle should have $(2s+1) = 5$ independent degrees of freedom. $h_{\mu\nu}$, being a symmetric 4×4 matrix, has ten independent components—five too many. If we take the divergence of its eom in the absence of sources, we get

$$m_1^2 \partial^\mu h_{\mu\nu} + m_2^2 \, \partial_\nu h = 0 \ . \tag{2.10}$$

The divergence of $G^L_{\mu\nu}$ vanishes because of the Bianchi identities—the full Einstein tensor obeys $\nabla^\mu G_{\mu\nu} = 0$, which at linear order reduces to $\partial^\mu G^L_{\mu\nu} = 0$. We are thus left with four constraints—one for each value of ν—of the form

$$m_1^2 \, \epsilon_{\mu\nu} \, k^\mu + m_2^2 \, \epsilon^\mu{}_\mu \, k_\nu = 0 \ , \tag{2.11}$$

where $\epsilon_{\mu\nu}$ is the graviton's polarization tensor, $h_{\mu\nu} \sim \epsilon_{\mu\nu} e^{ik\cdot x}$. These four constraints reduce the number of independent polarizations a graviton can have, from ten to six. But we still have one too many. How do we get rid of it?

The (linearized) Einstein tensor is a sum of several terms, each involving two derivatives acting on $h_{\mu\nu}$. Its trace has the same structure, but it has to be a scalar, so it can only be a linear combination of $\Box h$ and $\partial_\mu \partial_\nu h^{\mu\nu}$. The actual combination is

$$G^{L\mu}{}_\mu = \partial_\mu \partial_\nu h^{\mu\nu} - \Box h \ . \tag{2.12}$$

So, if we act with another divergence on eq. (2.10), and we *choose* $m_1^2 = -m^2 \equiv m^2$, we get

$$0 = m^2 \big(\partial_\mu \partial_\nu h^{\mu\nu} - \Box h \big) = m^2 G^{L\mu}{}_\mu \tag{2.13}$$

For this particular choice of mass parameters, the trace of $G^L_{\mu\nu}$ has to vanish. Which means that if we go back to the undifferentiated eom, eq. (2.7), and we trace it, in the absence of sources we get

$$m^2 \, h = 0 \tag{2.14}$$

which is yet another constraint on the graviton's possible polarizations: $\epsilon^\mu{}_\mu = 0$. This brings the number of independent polarizations down to five, which is the desired value. The choice $m_1^2 = -m_2^2$ is called the "Fierz-Pauli tuning". It is necessary if we want to describe a massive graviton rather than a massive graviton plus something else. It turns out that this something else—the so-called 'sixth mode'—is always problematic, as we will see shortly in the Stückeleberg formalism, where many of the peculiar properties of massive gravity become transparent.

With the Fierz-Pauli tuning, the mass term in the action reduces to

$$S_{\rm mass} \to -\tfrac{1}{2} M_{\rm Pl}^2 m^2 \int d^4x \left[h_{\mu\nu} h^{\mu\nu} - h^2 \right] , \tag{2.15}$$

and the eom becomes

$$G^L_{\mu\nu} + m^2 \big(h_{\mu\nu} - h\, \eta_{\mu\nu} \big) = \tfrac{1}{M_{\rm Pl}^2} T_{\mu\nu} \ . \tag{2.16}$$

2.2 The vDVZ discontinuity

It is interesting to consider the behavior of the theory at momenta much bigger than the graviton mass, or, equivalently, at distances much shorter than the graviton's Compton wavelength. The reason is that the massless theory—GR—passes all solar system tests of gravitational physics, so our massive theory should reduce to GR at least at solar system length scales. If we want to modify gravity at cosmological distances, we need a mass of order of the present Hubble constant,

$$m \sim H_0 \sim \big(10^{28} \text{ cm} \big)^{-1} \ . \tag{2.17}$$

For the moment this is the only length scale in our system, which for the gravitational potential demarcates the $1/r$ behavior from the exponential Yukawa-type one. It is then natural to demand that the theory reduce to GR for larger momenta, $p \gg m \sim H_0$, or shorter distances $r \ll m^{-1} \sim H_0^{-1}$.

In the literature you will find that this large-momentum, short-distance limit is often referred to as the "$m \to 0$ limit". This is somewhat misleading: we just went through some moderate pain to introduce a mass, and we decided that it should be of order of the Hubble constant; why should we set to zero now? However, like all limits involving *dimensionful* quantities, "$m \to 0$" should be interpreted as "m much smaller than some other relevant energy scale." In our case, this other energy scale is the momentum, or the inverse distance, so that in this sense the $m \to 0$ limit is precisely the limit we have been talking about above.

Notice that for a theory of massive particles with spin $s \geq 1$, the $m^2 \to 0$ limit is in a sense always discontinuous. This is because the number of dof for any finite mass is $(2s+1) \geq 3$, but for vanishing m^2 it should drop to two—the two helicity states. This is not particularly worrisome, or interesting, unless it leads to measurable effects.

In the spin-1 case, the extra dof's coupling to external (conserved) currents drops to zero continuously when m^2 approaches zero, so that its presence becomes undetectable: its contribution to the long-range force between two external sources becomes negligible, and so does its production rate by external sources. This is encoded by the particle's (Feynman) propagator,

$$\langle T(Z_\mu Z_\nu)\rangle = \frac{i}{-p^2 - m^2 + i\epsilon}\left[\eta_{\mu\nu} + \frac{p_\mu p_\nu}{m^2}\right] . \tag{2.18}$$

The propagator is formally divergent for $m \to 0$. Yet, if our Z_μ field is coupled to a conserved current J^μ, whenever either end of the propagator is attached to a J^μ vertex, the $\frac{p_\mu p_\nu}{m^2}$ does not contribute, because current conservation in momentum space reads $p_\mu J^\mu(p) = 0$. Taking into account this, the $m \to 0$ limit is smooth, and one simply gets

$$\langle T(Z_\mu Z_\nu)\rangle \to \frac{i}{-p^2 + i\epsilon}\eta_{\mu\nu} , \tag{2.19}$$

which is the correct propagator for a *massless*, spin-1 particle in Feynman gauge. We have been focusing on the propagator because that is all one needs to compute physical observables. For instance, the potential between two non-relativistic external sources mediated by our spin-1 particle can be computed as the zero-frequency, three-dimensional Fourier transform of the propagator above, attached to the two sources. Or the rate at which an external source produces our particles, can be computed—via the optical theorem—by taking the imaginary part of a Feynman diagram involving just our propagator with the same source on the two ends. The bottom line is that if the propagator coupled to conserved currents has a smooth $m \to 0$ limit, all physical observables do.

In the spin-2 case, things do not work out as straightforwardly. As usual, the massive graviton propagator can be computed by 'inverting' its quadratic action, or, equivalently, by solving its linearized equation of motion with a delta-like source. The result is

$$\langle T(h_{\mu\nu} h_{\alpha\beta})\rangle = \frac{1}{M_{\rm Pl}^2}\cdot\frac{i}{-p^2 - m^2 + i\epsilon}\left[\tfrac{1}{2}\big(P_{\mu\alpha}P_{\nu\beta} + P_{\mu\beta}P_{\nu\alpha}\big) - \tfrac{1}{3}\,P_{\mu\nu}P_{\alpha\beta}\right] , \tag{2.20}$$

where $P_{\mu\nu}$ is the same structure appearing in the massive spin-1 case,

$$P_{\mu\nu} = \eta_{\mu\nu} + \frac{p_\mu p_\nu}{m^2} . \tag{2.21}$$

Now, since our graviton couples to a conserved stress-energy tensor, all terms involving p^μ in the propagator can be ignored, since in momentum space $p_\mu T^{\mu\nu}(p)$. We can then replace all P's with η's, and use the 'reduced' propagator

$$\langle T(h_{\mu\nu} h_{\alpha\beta})\rangle = \frac{1}{M_{\rm Pl}^2}\cdot\frac{i}{-p^2 - m^2 + i\epsilon}\left[\tfrac{1}{2}\big(\eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha}\big) - \tfrac{1}{3}\,\eta_{\mu\nu}\eta_{\alpha\beta}\right] . \tag{2.22}$$

The $m \to 0$ limit looks smooth, and we get

$$\langle T(h_{\mu\nu} h_{\alpha\beta})\rangle \to \frac{1}{M_{\rm Pl}^2}\cdot\frac{i}{-p^2 + i\epsilon}\left[\tfrac{1}{2}\big(\eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha}\big) - \tfrac{1}{3}\,\eta_{\mu\nu}\eta_{\alpha\beta}\right] . \tag{2.23}$$

So far, this is analogous to the spin-1 case. However, if we now compare this propagator with that of an *exactly* massless spin-2 particle (in DeDonder gauge),

$$\langle T(h_{\mu\nu}h_{\alpha\beta})\rangle_{m=0} = \frac{1}{M_{\rm Pl}^2}\cdot\frac{i}{-p^2+i\epsilon}\big[\tfrac{1}{2}\big(\eta_{\mu\alpha}\eta_{\nu\beta}+\eta_{\mu\beta}\eta_{\nu\alpha}\big)-\tfrac{1}{2}\,\eta_{\mu\nu}\eta_{\alpha\beta}\big] \; , \tag{2.24}$$

we discover an $O(1)$ difference—1/3 versus 1/2 in the last term. This difference is physical, i.e., measurable: Once we attach the propagator to external $T_{\mu\nu}$'s, the last term selects their traces. For non-relativistic sources, the trace is (minus) the energy density, and so the last term contributes to whatever observable we are computing. For ultra-relativistic sources, like e.g. a light ray, the trace vanishes, and so only the other terms contribute to, say, light bending. This is usually phrased as the statement that the value of G we infer from Cavendish-type experiments does not match the value we infer from the bending of light rays by the sun. It looks like we can experimentally tell the difference between an exactly vanishing graviton mass and an infinitesimal one. This funny discontinuity goes under the name of van Dam-Veltman-Zakharov (vDVZ) discontinuity. Its origin is manifest in the Stückelberg formalism, which will also show us how it is in fact resolved.

2.3 The Stückelberg formalism

The behavior of the theory in the $m \to 0$ limit, or, more properly, at momenta much higher than the mass, is better understood via the following trick. Consider an ultra-relativistic massive graviton, with $p \gg m$. Formally, all its five polarization states belong to the same irreducible representation of the Lorentz group, i.e., they can be brought into one another by suitable Lorentz transformations. In practice however, for bigger and bigger p's, one needs faster and faster boosts to do so. This is because in the infinite momentum limit, when our particle travels at the speed of light, the helicity $h = \vec{J}\cdot\hat{p}$ is Lorentz-invariant—different helicity states cannot be brought into one another via Lorentz transformations. As a result, for $p \gg m$, we can conveniently think of the five polarization states as being—in first approximation, i.e., up to corrections suppressed by m/p—five separate helicity states, with $h = 0, \pm 1, \pm 2$.

To describe the $h = \pm 2$ states we need a tensor field $h_{\mu\nu}$ with standard massless spin-2 gauge invariance. To describe the $h = \pm 1$ states we need a vector field A_μ with standard massless spin-1 gauge invariance. To describe the $h = 0$ state we need a scalar field ϕ. Let's therefore take our action, eq. (2.6), and replace everywhere $h_{\mu\nu}$ with

$$H_{\mu\nu} = h_{\mu\nu} + \partial_\mu A_\nu + \partial_\nu A_\mu + 2\partial_\mu\partial_\nu\phi \; . \tag{2.25}$$

The action is now invariant under the spin-2 and spin-1 gauge transformations

$$h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu \; , \qquad A_\mu \to A_\mu - \xi_\mu \tag{2.26}$$

and

$$A_\mu \to A_\mu + \partial_\mu\Lambda \; , \qquad \phi \to \phi - \Lambda \; , \tag{2.27}$$

because $H_{\mu\nu}$ is. We can always choose a gauge in which A_μ and ϕ vanish ($\Lambda = \phi$, $\xi_\mu = A_\mu$), in which case $H_{\mu\nu}$ reduces to $h_{\mu\nu}$, and we recover the original action (this gauge choice is called "unitary gauge"). This means that our introducing new A_μ and ϕ fields with associated gauge invariances is harmless—the new theory is physically identical to the original one. As usual, gauge invariance, rather than a physical symmetry, is the signal that we chose a redundant description for the physical degrees of freedom. However, such a redundancy is definitely useful in the massless case, and as result the new action will have an obvious physical interpretation in the high momentum ($p \gg m$) limit.

Since the Einstein-Hilbert action is gauge invariant, and A_μ and ϕ appear in $H_{\mu\nu}$ as gauge parameters would, we get no new term from this part of the action. Nor do we get anything new from the coupling to matter, (2.5), because of conservation of $T_{\mu\nu}$. So, only the mass term is sensitive to the replacement (2.25). Let's first keep m_1 and m_2 generic, to see the need for the Fierz-Pauli tuning in this new language. If we focus on the terms that involve the scalar ϕ only, we get

$$S_{\rm mass} \to -2M_{\rm Pl}^2 \int d^4x \left[m_1^2 \left(\partial_\mu \partial_\nu \phi\right)^2 + m_2^2 \left(\Box\phi\right)^2 \right] \tag{2.28}$$

$$= -2M_{\rm Pl}^2 \int d^4x \, (m_1^2 + m_2^2) \left(\Box\phi\right)^2 \, , \tag{2.29}$$

where in the second step we integrated by parts the first term twice. This is a *four-derivative* kinetic term for ϕ. It yields a four-derivative linearized eom, which needs four initial conditions to be solved. It thus corresponds to *two* scalar degrees of freedom. It turns out that one of them is always a *ghost*—it has negative energy. I briefly review this, and the problems associated with having ghosts, in the Appendix. The bottom line is that having a four-derivative kinetic term is bad, and we should try to avoid it. The Fierz-Pauli tuning, $m_2^2 = -m_1^2$ arises in this formalism as the requirement that there are no ghosts. If we then choose $m_1^2 = -m_2^2 \equiv m^2$, upon the replacement (2.25) we get

$$S_{\rm mass} \to S_{\rm mass} - \tfrac{1}{2} M_{\rm Pl}^2 m^2 \int d^4x \left[\left(\partial_\mu A_\nu + \partial_\nu A_\mu\right)^2 - 2(\partial\cdot A)^2 \right. \tag{2.30}$$

$$+ \, 2\partial_\mu\partial_\nu\phi \left(h^{\mu\nu} - h\,\eta^{\mu\nu}\right) \tag{2.31}$$

$$\left. + \, h\,\partial A \right] \, . \tag{2.32}$$

As we will see, the schematic terms collected in the third line will be negligible—so there is no point in keeping track of their precise structure. As to the first line, it can be written in a much more familiar form by integrating by parts the second term twice, and combining it with the terms in parentheses (schematically: $(a+b)^2 - 4ab = (a-b)^2$):

$$\left(\partial_\mu A_\nu + \partial_\nu A_\mu\right)^2 - 4(\partial\cdot A)^2 \to \left(\partial_\mu A_\nu - \partial_\nu A_\mu\right)^2 = F_{\mu\nu}F^{\mu\nu} \, , \tag{2.33}$$

which is the standard kinetic term for a massless spin-1 field. As to the second line, it is a kinetic mixing between ϕ and $h_{\mu\nu}$. Since ϕ does not have a kinetic term on its own—the only available one had four derivatives, which we decided was bad, and we canceled it—this is the leading source of kinetic energy for ϕ. We can diagonalize the $h_{\mu\nu}$-ϕ system via the field redefinition (not a gauge transformation)

$$h_{\mu\nu} = \hat{h}_{\mu\nu} + m^2 \phi \, \eta_{\mu\nu} \; . \tag{2.34}$$

Plugging this into the (quadratic) Einstein-Hilbert action (2.3), we get, schematically,

$$S_{\rm EH}[h_{\mu\nu}] = S_{\rm EH}[\hat{h}_{\mu\nu}] + \int d^4x \big[\partial \hat{h} \, \partial\phi + \partial\phi \, \partial\phi \big] \; . \tag{2.35}$$

The $\partial\hat{h}\,\partial\phi$ terms, once integrated by parts, precisely cancel against the mixing terms (2.31). The $\partial\phi\,\partial\phi$ terms combine with similar ones coming from implementing the field redefinition in (2.31), to yield an healthy kinetic term for ϕ,

$$-\, 3 M_{\rm Pl}^2 m^4 \int d^4x \, (\partial\phi)^2 \; . \tag{2.36}$$

Now, the crucial fact is that the field redefinition (2.34) introduces a direct coupling of ϕ to matter—more specifically, to the trace of $T_{\mu\nu}$—via $h_{\mu\nu}T^{\mu\nu}$. Putting everything together, the (quadratic) action is

$$\mathcal{L} = \mathcal{L}_{EH}[\hat{h}_{\mu\nu}] + \tfrac{1}{2} \hat{h}_{\mu\nu} T^{\mu\nu} \tag{2.37}$$
$$-\, 3 M_{\rm Pl}^2 m^4 (\partial\phi)^2 + \tfrac{1}{2} m^2 \, \phi \, T \tag{2.38}$$
$$-\, \tfrac{1}{2} M_{\rm Pl}^2 m^2 F_{\mu\nu} F^{\mu\nu} \tag{2.39}$$
$$+\, M_{\rm Pl}^2 m^2 \, h \, \partial A + M_{\rm Pl}^2 m^2 (\hat{h} + m^2\phi)^2 \; . \tag{2.40}$$

Once again, we are schematically collecting in the last line terms that will turn out to be negligible in the limit we are interested in. In particular the very last term comes from the graviton's mass term upon the field redefinition (2.34).

It is particularly convenient to choose the same normalization for all fields, as determined by the their kinetic terms—say canonical normalization:

$$\hat{h}^c_{\mu\nu} = M_{\rm Pl} \, \hat{h}_{\mu\nu} \; , \qquad A^c_\mu = m M_{\rm Pl} \, A_\mu \; , \qquad \phi^c = m^2 M_{\rm Pl} \, \phi \; . \tag{2.41}$$

We thus get

$$\mathcal{L} = \mathcal{L}_{EH}[\hat{h}^c_{\mu\nu}] + \tfrac{1}{2} \tfrac{1}{M_{\rm Pl}} \hat{h}^c_{\mu\nu} T^{\mu\nu} \tag{2.42}$$
$$-\, 3 (\partial\phi^c)^2 + \tfrac{1}{2} \tfrac{1}{M_{\rm Pl}} \phi^c \, T \tag{2.43}$$
$$-\, \tfrac{1}{2} F^c_{\mu\nu} F^{c\mu\nu} \tag{2.44}$$
$$+\, m \, h^c \, \partial A^c + m^2 (\hat{h}^c + \phi^c)^2 \; . \tag{2.45}$$

(with an abuse of notation I am using $\mathcal{L}_{EH}[\hat{h}^c_{\mu\nu}]$ to denote the Einstein-Hilbert action without any $M_{\rm Pl}$ upfront—I hope this will not lead to confusion.) Then, the physical picture is transparent: The first line is just linearized GR. The second line describes a scalar field, coupled with gravitational strength to the trace of matter's stress-energy tensor. This is the origin of the vDVZ discontinuity. In the $m \to 0$, this scalar is still there, with the same coupling. The third line describes a free massless spin-1 field, which for most purposes can be ignored—since it does not couple to anything else. The terms in the fourth line are the only ones suppressed by powers of m. As anticipated, they can thus be ignored in the $m \to 0$ ($p \gg m$) limit.

2.4 Interactions and strong coupling

We saw that for distances much smaller than the graviton's Compton wavelength, massive gravity should be thought of as GR plus a massless scalar coupled with gravitational strength to matter's $T_{\mu\nu}$. The Stückelberg formulation turns out to be invaluable to study the interactions among these degrees of freedom. We already know that GR, once expanded in powers of $h_{\mu\nu}$, contains infinitely many interaction terms. These become important for classical solutions whenever $h_{\mu\nu}$ becomes large, of order one, like for instance close to a black hole's horizon. Quantum mechanically, they become important for processes that involve gravitons with energies of order of $M_{\rm Pl}$. In massive gravity, however, there are interactions that are much more relevant, in the sense that they become important for much weaker gravitational fields, and for much less energetic quanta.

This discussion is quite topical. Let's forget for a second about gravity, and let's consider massive gauge bosons, for some non-abelian symmetry group. The action that describes them is

$$S = \int d^4x \big[-\tfrac{1}{2}{\rm tr} F_{\mu\nu}F^{\mu\nu} + \tfrac{1}{2}m^2 {\rm tr} A_\mu A^\mu \big] \; , \tag{2.46}$$

where the trace is taken over the symmetry group's algebra, and as usual for non-abelian theories

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + g\,[A_\mu, A_\nu] \; , \tag{2.47}$$

where g is a small coupling constant. If one expands the F^2 term in the action in powers of A, one gets interaction terms with three and four A's. If the theory were massless, these interaction terms would lead to perfectly well-behaved (tree-level) scattering amplitudes at all energies. For instance, a $2 \to 2$ scattering amplitude would be of order g^2—with no power of energy, since g is already dimensionless—which is small as long as g is. That is to say, at least at tree level, the massless theory is perturbative at all energies. For the massive theory instead there is a twist. If we start scattering longitudinally polarized gauge bosons, we get enhancement factors in the amplitude coming from the longitudinal polarization vector $\epsilon^\mu = p^\mu/m$. These enhancement factors grow with energy, and make the theory violate unitarity—that is, the $2 \to 2$ scattering amplitude becomes of order one—at a scattering energy roughly of order m/g. It is this phenomenon that, once applied to the W bosons of electroweak interactions, tells us that some new physics should show up before we violate unitarity. This new physics could in principle take many forms, a weakly coupled standard model Higgs boson being just the simplest possibility. Apparently, it is also the correct one.

Back to massive gravity now. Addressing the importance of the new interactions is quite non-trivial in the unitary gauge description—for instance, already in the massive gauge boson example, there are cancellations that make the strong coupling scale substantially higher than naively expected (m/g instead of $m/\sqrt{g}$). In the Stückelberg formalism however, everything is much more transparent: these interactions correspond to self-interactions of the scalar ϕ. The reason is quite clear from the massive gauge boson example: first, these interactions become important at energies parametrically bigger than the mass. Then we can use the Stückelberg decomposition in terms of states of definite helicity. Second, the strong-coupling phenomenon has to do with the longitudinal polarizations, because their polarization vectors carries one power of p/m. In the massive gravity case, the helicity-one polarization tensor will carry a p/m, while the helicity-zero one will carry $(p/m)^2$.

To find the new interactions, we have to extend the Stückelberg trick to non-linear order. It is actually quite straightforward. Let's extend the replacement (2.25) to non-linear oder by just demanding that $H_{\mu\nu}$ be obtained from $h_{\mu\nu}$ via a non-linear gauge transformation,

$$H_{\mu\nu} = h_{\mu\nu} + \partial_\mu \pi_\nu + \partial_\nu \pi_\mu - \partial_\mu \pi^\alpha \partial_\nu \pi_\alpha \tag{2.48}$$

(this expression follows from the standard transformation law $g \to \frac{\partial x}{\partial x'} \frac{\partial x}{\partial x'} g$ for the metric—see e.g [23]), with gauge parameter

$$\pi^\mu = A^\mu + \partial^\mu \phi \; . \tag{2.49}$$

Consider then the non-linear theory, made up of the full Einstein-Hilbert term plus the mass term (2.15), plus the standard non-linear diff-invariant coupling of the metric to matter. If we replace $h_{\mu\nu}$ with $H_{\mu\nu}$ as given in (2.48), we get no new terms from the Einstein-Hilbert or the matter parts of the action, since these are diff-invariant, and $H_{\mu\nu}$ is obtained from $h_{\mu\nu}$ by performing a diff (this is the reason why (2.48) is a clever non-linear generalization of (2.25)). The mass term however is not diff-invariant, and, even though it is quadratic in $h_{\mu\nu}$, it gives us non-trivial interactions for the Stückelberg fields, since these appear non-linearly in $H_{\mu\nu}$. Focusing on the interaction terms involving ϕ only, we get, schematically,

$$\mathcal{L}_{\rm mass} \to \mathcal{L}_{\rm int} \sim m^2 M_{\rm Pl}^2 \big[(\partial^2 \phi)^3 + (\partial^2 \phi)^4 \big] \; , \tag{2.50}$$

with some specific Lorentz contractions. Going to canonical normalization—eq. (2.41)—these become

$$\mathcal{L}_{\rm int} \sim \frac{(\partial^2 \phi^c)^3}{m^4 M_{\rm Pl}} + \frac{(\partial^2 \phi^c)^4}{m^6 M_{\rm Pl}^2} \; . \tag{2.51}$$

It is useful to recast each denominator as a power of a single energy scale,

$$\mathcal{L}_{\rm int} \sim \frac{(\partial^2 \phi^c)^3}{\Lambda_5^5} + \frac{(\partial^2 \phi^c)^4}{\Lambda_4^8} \; , \tag{2.52}$$

with

$$\Lambda_5 \equiv (m^4 M_{\rm Pl})^{1/5} \quad \ll \quad \Lambda_4 \equiv (m^3 M_{\rm Pl})^{1/4} \tag{2.53}$$

(the hierarchy follows straightforwardly from $m \ll M_{\rm Pl}$.)

The fact that the cubic interaction is weighed at the denominator by an energy scale that is much lower than that weighing the quartic one, tells us that the cubic interaction is much more important. For instance, in a scattering experiment with some typical momentum p, the contribution to the $2 \to 2$ scattering amplitude coming from two cubic vertices connected by a propagator, just by dimensional analysis, will be of order p^{10}/Λ_5^{10}, whereas that due to the contact quartic interaction, will be of order p^8/Λ_4^8. Their ratio is $(p/m)^2$ which is much bigger than one at ultra-relativistic momenta—the limit we are considering. Related to this, the first contribution becomes strongly coupled (i.e., of order one) at energies of order Λ_5, much before the other contribution does. So, at least for quantum mechanical processes, Λ_5 is the strong coupling scale of the theory—the analog of the massive gauge bosons' m/g.

To put things into perspective, the quantum mechanical strong coupling scale for GR is the Planck scale. The ratio between the two strong coupling scales is

$$\frac{\Lambda_5}{M_{\rm Pl}} = \Big(\frac{m}{M_{\rm Pl}}\Big)^{4/5} \sim 10^{-48} \tag{2.54}$$

(I have used $m \sim H_0$), which is a measure of how much narrower the regime of validity of the massive gravity effective field theory is w.r.t. that of GR [4]. It is clear that the apparently harmless action of adding a mass for the graviton has far reaching implications for the consistency of the theory.

In getting to eq. (2.52) we have kept only the interaction terms involving ϕ. It is straightforward to check that the other interactions we get from the graviton's mass term upon the replacement (2.48)—for instance those of the form $(\partial A)^3$—are all subleading w.r.t those we explicitly kept, in the sense they are weighed by energy scales that are much higher than Λ_5. They can thus be neglected.

2.5 The Vainshtein effect

Besides quantum mechanical strong-coupling, a more relevant question for astrophysical or cosmological observables, is whether the interactions above show up as sizable non-linearities for classical solutions. For GR, classical non-linearities are important when the (dimensionless) $h_{\mu\nu}$ becomes of order one. In our canonical normalization, this would correspond to $h^c_{\mu\nu}$ of order $M_{\rm Pl}$. For a simple spherical source, we know that this happens at distances of order of the Schwarzschild radius $R_S \sim M_*/M_{\rm Pl}^2$, where M_* is the source's mass. We can estimate very easily the analogous distance scale at which ϕ's self-interactions become important in massive gravity. Consider the

[4] At energies of order of the strong coupling scale and above, an effective field theory cannot be trusted anymore, and has to be replaced by a more complete theory—the so-called 'UV completion'. For instance, for the W and Z gauge bosons of electroweak interactions, a possible UV completion is the Standard Model with a Higgs boson. For GR, a possible UV completion is string theory. For massive gravity, at present we have no idea about what a possible UV completion could be.

quadratic Lagrangian (2.42). In the presence of a point-like source, with $T_{00} = M_* \delta^3(\vec{x})$, there is a simple $1/r$ linearized solution for ϕ,

$$\phi^c \sim \frac{M_*}{M_{\rm Pl}} \frac{1}{r} \, . \tag{2.55}$$

To assess at which distance from the source non-linearities are important, we can plug this solution into the quadratic Lagrangian and into the self-interaction terms (2.52). When the two pieces become of the same order, the linear approximation breaks down. Using $\partial \sim 1/r$, we get

$$\mathcal{L}_2 \equiv (\partial \phi^c)^2 \sim \Big(\frac{M_*}{M_{\rm Pl}} \Big)^2 \frac{1}{r^4} \tag{2.56}$$

$$\mathcal{L}_3 \equiv \frac{(\partial^2 \phi^c)^3}{\Lambda_5^5} \sim \Big(\frac{M_*}{M_{\rm Pl}} \Big)^3 \frac{1}{m^4 M_{\rm Pl}} \cdot \frac{1}{r^9} \tag{2.57}$$

$$\mathcal{L}_4 \equiv \frac{(\partial^2 \phi^c)^4}{\Lambda_4^8} \sim \Big(\frac{M_*}{M_{\rm Pl}} \Big)^4 \frac{1}{m^6 M_{\rm Pl}^2} \cdot \frac{1}{r^{12}} \tag{2.58}$$

Notice that the cubic and quartic terms at large distances are suppressed by large powers of $1/r$. However their prefactors are enhanced by large powers of $1/m$, and of $(M_*/M_{\rm Pl})$ (we are considering massive astrophysical objects, for which $M_* \gg M_{\rm Pl}$.) So, in moving in from $r = \infty$ towards to the origin, there will be a critical distance at which one of them becomes of the same order as the quadratic term. The ratios we have to focus on are

$$\frac{\mathcal{L}_3}{\mathcal{L}_2} \sim \frac{M_*}{M_{\rm Pl}^2} \frac{1}{m^4} \frac{1}{r^5} \, , \qquad \frac{\mathcal{L}_4}{\mathcal{L}_2} \sim \Big(\frac{M_*}{M_{\rm Pl}^2} \frac{1}{m^3} \frac{1}{r^4} \Big)^2 \, . \tag{2.59}$$

We see that, in our linearized approximation, the cubic and the quartic become important, respectively, at the critical distances $r \sim (R_S m^{-4})^{1/5}$ and $r \sim (R_S m^{-3})^{1/4}$. Since the former is the larger, it is the one that signals the breakdown of the linearized approximation in approaching the source from infinity. We see that even for classical non-linearities, the cubic interaction is more important than the quartic one.

To summarize: when computing the gravitational field produced by a macroscopic source, we have to take into account ϕ's self-interactions. In approaching the source, they become relevant at a distance of order of the so-called Vainshtein radius

$$R_V \equiv \big(R_S \, m^{-4} \big)^{1/5} \, . \tag{2.60}$$

That is a *huge* distance. For instance for a solar-mass source ($R_S \sim$ km), and for $m \sim H_0$, we get $R_V \sim 10^{18}$ km. This is much much bigger than the size of the solar system (this is yet another signal that interactions in massive gravity are much much stronger than in GR.) This means that the gravitational field of the sun cannot be computed reliably within the linearized approximation at solar system scales. This is actually *good*: recall that linearized massive gravity is order-one *wrong* in its predictions for solar system tests, because of the vDVZ discontinuity. The vDVZ discontinuity is due precisely to the linearized exchange of the scalar ϕ. Is it possible that the non-linearities we are studying actually "screen" the scalar on scales smaller than R_V, thus recovering agreement with solar system tests? This actually happens quite naturally, at least in some examples that we will mention below. But before seeing this effect—the so-called 'Vainshtein effect'—at work, we have to solve another problem.

2.6 Is the sixth mode back?

Let us focus on the scalar ϕ, and let's consider its dynamics in the presence of a macroscopic source, like the Sun. The source induces a classical background $\Phi^c(x)$. In the linearized approximation it is given in (2.55). (Now we are using Φ^c to denote the background, and ϕ^c to denote the full field—background plus fluctuations.) To study the stability of this solution we expand the Lagrangian to quadratic order in the fluctuation $\varphi \equiv \phi^c - \Phi^c$. The result is schematically of the form

$$\mathcal{L}_\varphi = -(\partial\varphi)^2 + \frac{(\partial^2\Phi^c)}{{\Lambda_5}^5}(\partial^2\varphi)^2 \; . \tag{2.61}$$

We see that the background gives a four-derivative contribution to the φ kinetic term. As mentioned earlier and as discussed in the Appendix, this results in the appearance of a massive *ghost*, now with an x-dependent mass

$$m^2_{\rm ghost}(x) \sim \frac{\Lambda_5^5}{\partial^2\Phi^c(x)} \; . \tag{2.62}$$

Notice that in the absence of sources, when the background solution vanishes, the ghost formally becomes infinitely massive. So, we can interpret this result in the following way: the Fierz-Pauli tuning that we introduced to eliminate the sixth mode, actually made it infinitely massive. A non-trivial background gravitational field is enough to bring it back, down to finite values of the mass.

The question is whether this ghost is a mathematical artifact, or whether it is actually relevant for low energy physics, in the following sense. Remember that we are dealing with an effective theory with a tiny UV cutoff Λ_5. Therefore, we should not worry until the mass of the ghost drops below Λ_5, since no prediction of our theory can be trusted at energies above it. In approaching the source from far away, this happens at a distance $R_{\rm ghost}$ such that $\partial^2\Phi^c \sim \Lambda_5^3$. Unfortunately this is a huge distance, parametrically larger than the (already huge) Vainshtein radius R_V. Indeed, for a source of mass M_*, at distances $r \gg R_V$ the background field goes as $\Phi^c(r) \sim (M_*/M_{\rm Pl}) \cdot 1/r$, so that

$$R_{\rm ghost} \sim \frac{1}{\Lambda_5}\left(\frac{M_*}{M_{\rm Pl}}\right)^{1/3} \gg R_V \sim \frac{1}{\Lambda_5}\left(\frac{M_*}{M_{\rm Pl}}\right)^{1/5} \; . \tag{2.63}$$

(You can easily convince yourself that this way of writing the Vainshtein radius is equivalent to our definition, eq. (2.60)). The ghost is thus going to show up in an extremely weak background field, still well within the linear regime.

Inside $R_{\rm ghost}$, in the spirit of Appendix A, one is forced to postulate that additional physics lighter than the local ghost mass cures the instability, that is the cutoff must be lowered from Λ_5

to $m_{\rm ghost}(x)$. A byproduct of this in general would be that interactions strengthen, being weighted by the new cutoff scale rather than by Λ_5. But let us optimistically assume that, instead, the only effect of this new physics is to cure the ghost instability. However, when the local ghost mass is of order of the inverse distance from the source there is no way of proceeding further without specifying the UV completion of the theory, since the background itself has a typical length scale of order of the UV cutoff. One can easily check that this happens at the Vainshtein radius R_V. *There is no sense in which one can trust the classical solution below* R_V. Since one can hope to recover GR only in the region inside R_V, where non-linear effects can hide the scalar (Vainshtein effect), this also means that *there is no range of distances where GR is a good approximation to massive gravity.*

2.7 Improving the theory

There is one last hope. In going from the linearized theory of sect. 2.1 to the interacting one, we just promoted the quadratic Einstein-Hilbert term to the fully non-linear one, and the linearized coupling to matter, $h_{\mu\nu}T^{\mu\nu}$, also to its fully non-linear version. This process does not introduce interactions for the Stückelberg fields ϕ and A_μ, because we were careful enough to define them at non-linear level simply as gauge-parameters, which therefore cannot show up in fully diff-invariant pieces of the action. We did nothing to the quadratic mass term though, eq. (2.15). We know that precisely that term encodes the dynamics of the Stückelberg fields, upon replacing $h_{\mu\nu} \to H_{\mu\nu}$. If we add higher order $h^n_{\mu\nu}$ terms to it, we introduce *new* interactions for ϕ and A_μ. in particular, since $H_{\mu\nu}$ starts linear in the fields, with a cubic $h^3_{\mu\nu}$ addition we generate—among other things—new cubic interactions for ϕ.

We are thus led to consider the possibility of eliminating the problematic trilinear interaction of ϕ that we discussed above, by adding appropriate cubic terms in $H_{\mu\nu}$ to the Fierz-Pauli mass term. The three independent contractions are H^3, $H(H_{\mu\nu})^2$, and $(H_{\mu\nu})^3$, where the last stands for the cyclic contraction of the indices. These contain interaction terms for the Stückelberg field ϕ of the form $(\Box\phi)^3$, $\Box\phi(\partial_\mu\partial_\nu\phi)^2$, and $(\partial_\mu\partial_\nu\phi)^3$ which, for the proper choice of coefficients, cancel the trilinear interaction of eq. (2.52). However, in this way one introduces further quartic interactions $(\partial^2\phi)^4$ on top of those already present in the Fierz-Pauli mass term, because of the non-linear relation between $H_{\mu\nu}$ and ϕ of eq. (2.48). These are problematic for exactly the same reason as before, and the same problem shows up at any order: an interaction term of the form $(\partial^2\phi)^n$ evaluated around a background gives a contribution to the equation of motion for the fluctuations with too many derivatives. This signals the presence of a ghost instability. Again one can check that the cutoff—*i.e.* the ghost mass—becomes of order of the inverse radius at the new Vainshtein scale, always defined as the distance from the source at which non-linearities become relevant. One never recovers GR.

The only possibility is therefore to concentrate on theories in which *all* the interactions of the form $(\partial^2\phi)^n$ are set to zero by properly choosing infinitely many coefficients—those of the $H^n_{\mu\nu}$ terms in the non-linear generalization of the Fierz-Pauli mass term. We can look at this procedure as an extension at non-linear order of the Fierz-Pauli tuning which, as discussed in sect. 2.3, leads to the cancellation of the $(\Box\phi)^2$ terms. The cancellation of all the $(\partial^2\phi)^n$ interactions is also a way of raising the strong interaction scale to

$$\Lambda_3 = (m^2 M_{\rm Pl})^{1/3} \; , \tag{2.64}$$

which is much bigger than Λ_5, but still tiny compared to $M_{\rm Pl}$. In fact, after the cancellation, the leading interactions are of the form

$$m^2 M_{\rm Pl}^2 \, (\partial A)^2 (\partial^2\phi)^n \quad \text{and} \quad m^2 M_{\rm Pl}^2 \, (\hat{h}_{\mu\nu} + m^2 \, \eta_{\mu\nu}\phi)(\partial^2\phi)^n \; . \tag{2.65}$$

When the fields are canonically normalized all these terms are suppressed by the scale Λ_3, while additional interactions are weighed by higher scales. Related to this, the Vainshtein radius now shrinks to

$$R_V = \frac{1}{\Lambda_3}\Big(\frac{M_*}{M_{\rm Pl}}\Big)^{1/3} \sim \big(R_S \, m^{-2}\big)^{1/3} \; , \tag{2.66}$$

which is much smaller than before (see eq. (2.60)). At this radius, all non-linear terms of the form $(\hat{h}_{\mu\nu} + m_g^2\eta_{\mu\nu}\phi)(\partial^2\phi)^n$ become relevant; on the other hand there are no terms linear in A in the Lagrangian, so that A is sourced neither by matter nor by the other fields. Interactions involving this field are therefore irrelevant for our purposes, and can be consistently neglected.

Have we eliminated the ghost? It does not look like we have. If we neglect all interactions that are suppressed by scales higher than Λ_3, the Lagrangian takes the form, schematically,

$$\mathcal{L} \sim (\partial\hat{h}^c)^2 + (\partial\phi^c)^2 + \frac{(\hat{h}^c + \phi^c)(\partial^2\phi^c)^n}{\Lambda_3^{3n-3}} + \frac{1}{M_{\rm Pl}}\hat{h}^c T + \frac{1}{M_{\rm Pl}}\phi^c T \; . \tag{2.67}$$

In the presence of a massive source, $\hat{h}^c$ and ϕ^c develop a $1/r$ background solution that corrects the kinetic energy for φ fluctuations, which now includes four-derivative terms of the form

$$\frac{(\hat{H}^c + \Phi^c)(\partial^2\Phi^c)^{(n-2)}}{\Lambda_3^{3n-3}} \, (\partial^2\varphi)^2 \; , \tag{2.68}$$

where $\hat{H}^c$ and Φ^c are the background fields. Once again, it looks like we have a ghost with an x-dependent mass, which becomes lighter and lighter as we approach the source.

However there is a subtlety, which turns out to be relevant here. A four-derivative linearized eom is a problem only if it involves more than two *time*-derivatives, because it is the number of time derivatives that tells us how many initial conditions we are supposed to give, i.e., how many degrees of freedom we have. If we have no more that two time derivatives, than there is no room for the infamous sixth mode. Remarkably, it turns out that after all $(\partial^2\phi)^n$ have been cancelled via the non-linear extension of the Fierz-Pauli tuning, the Lagrangian one is left with, schematically given in (2.67), has the following properties:

1. It stops at quintic order, i.e., there are no interaction terms higher than $(\hat{h}^c + \phi^c)(\partial^2\phi^c)^4$.

2. Each term involves at most *two* time derivatives.

The second property seems at odds with Lorentz invariance—why is the time direction special? It's not. The point is that each term involves at most two derivatives along any given direction, be it x, y, z, or t. Then, given the structure of interactions in (2.67), property 2 has a chance to hold only if property 1 does.

There is a very convenient way of rewriting the resulting Lagrangian—once again keeping the leading interactions only—which makes use of a non-linear generalization of the field redefinition (2.34), schematically of the form $h_{\mu\nu} = \hat{h} + m^2 \eta_{\mu\nu} + m^2 \partial_\mu \phi \partial_\nu \phi$. This eliminates most of the mixed $\hat{h}$-ϕ interactions one has in (2.67), and one is left with

$$\mathcal{L} = \mathcal{L}_{\rm EH}[\hat{h}^c_{\mu\nu}] + \tfrac{1}{2} \tfrac{1}{M_{\rm Pl}} \, \hat{h}^c_{\mu\nu} T^{\mu\nu} \tag{2.69}$$

$$+ \mathcal{L}_{\rm galileon}[\phi^c] + \tfrac{1}{2} \tfrac{1}{M_{\rm Pl}} \, \phi^c T \tag{2.70}$$

$$+ \frac{1}{\Lambda_3^6} \hat{h}^c_{\mu\nu} \, X^{\mu\nu}_{(3)}[\phi^c] \; . \tag{2.71}$$

The first line describes linearized GR, for the (canonically normalized) graviton field $\hat{h}^c_{\mu\nu}$. The second line is the so-called galileon Lagrangian, which we will describe at length in the following. For the moment, suffice it to say that $\mathcal{L}_{\rm galileon}$ is a sum of terms each of the form

$$\partial\phi \, \partial\phi \, (\partial^2\phi)^n \; , \tag{2.72}$$

from $n = 0$ (the kinetic term), to $n = 3$ (a quintic interaction), and which all yield exactly *second order* equations of motion—i.e., they do not reintroduce the sixth mode about non-trivial backgrounds. The third line is a quartic interaction between the graviton and ϕ, via the tensor $X^{\mu\nu}_{(3)}$, which is schematically of the form $(\partial^2\phi^c)^3$, with some non-trivial Lorentz structure. This last interaction cannot be removed via a field redefinition, as can be checked by computing the tree-level amplitude for the scattering process $\hat{h}\phi \to \phi\phi$: the amplitude is nonzero, and there is no other Lagrangian term that can contribute to it.

3 DGP and the galileon

3.1 DGP

Another model that has received a lot of attention in the past decade or so is the Dvali-Gabadadze-Porrati (DGP) model. From a theoretical perspective, it appeared to be considerably healthier than massive gravity—at least before it was realized that the improved version of massive gravity that we briefly described is actually not that sick. From a phenomenological perspective, it attracted interest because there is a cosmological solution—the so-called self-accelerating solution—in which the universe accelerates in the absence of any stress-energy tensor, and, in particular, in the absence of a cc.

The model is defined via an extra-dimensional construction, but most of its properties—its interesting features as well its pathologies—are visible already, and, in fact, manifest, at the level of a 4D effective theory, which is a good description of the system at short distances (in the same spirit as the Stückelberg formalism for massive gravity). Let's start with the extra-dimensional construction. It involves a 5D Minkowski space, with a 4D hypersurface—a 'brane'—in it. Gravity can propagate in the 'bulk'—that is, in the whole 5D spacetime—whereas all other particles, which *we* are made of, are confined to the 4D brane. Notice that, unlike other popular extra-dimensional models, here the fifth dimension is neither compact nor warped: the bulk is a flat, infinite Minkoswki space. Usually this would yield 5D gravity, that is, a $1/r^2$ Newtonian potential rather than $1/r$, which would be of course ruled out. What saves the day here is that gravity features a large kinetic term localized on the 4D brane, which—at least in some distance range—confines the propagation of gravity to the brane. In formulae, the action for gravity is a sum of a 5D Einstein-Hilbert term and a 4D one, with different Newton's constants:

$$S_{\rm DGP} = 2M_5^3 \int_{5D} d^5X \, \sqrt{-G} \, R[G] + 2M_4^2 \int_{4D} d^4x \, \sqrt{-g} \, R[g] \; , \tag{3.1}$$

where G_{MN} is the 5D metric while $g_{\mu\nu}$ is its projection (the so-called induced metric) onto the brane [5]. To this action, we should add a 4D action for matter, localized on the brane, possibly including a cosmological constant, that is, a brane tension.

A special role is played by the length scale

$$L_{\rm DGP} \equiv 1/m \equiv \frac{M_4^2}{M_5^3} \; . \tag{3.2}$$

Below $L_{\rm DGP}$ gravity looks four-dimensional, while at larger length scales it enters the five-dimensional regime. To see this, consider the brane-to-brane propagator for the graviton. This is what we need in order to compute the gravitational potential between two sources that are localized on the brane, and, as usual, it can be computed by inverting the quadratic action, or by solving

[5] We are using $M, N, \dots = 0, \dots 4$ as 5D Lorentz indices, and $\mu, \nu, \dots = 0, \dots, 3$ as 4D ones. Likewise, X^M are the 5D bulk coordinates, and x^μ are the 4D ones along the brane.

the linearized eom with delta-like sources. Neglecting terms that vanish because of stress-energy conservation, one finds

$$\langle T(h_{\mu\nu}h_{\alpha\beta})\rangle_{\text{brane}} = \frac{1}{M_{\text{Pl}}^2}\cdot\frac{i}{-p^2-m|p|}\left[\tfrac{1}{2}\left(\eta_{\mu\alpha}\eta_{\nu\beta}+\eta_{\mu\beta}\eta_{\nu\alpha}\right)-\tfrac{1}{3}\,\eta_{\mu\nu}\eta_{\alpha\beta}\right] , \tag{3.3}$$

where $|p| \equiv \sqrt{p^2}$ is the magnitude of the 4D momentum, and I am avoiding being explicit about $i\epsilon$'s at the denominator because now the analytic structure is more complicated [6]. Notice that to talk about the brane-to-brane correlator in momentum space, we are implicitly using a mixed (y, p^μ) representation, where y is the fifth coordinate, and p^μ is the 4D momentum. The denominator encodes the behavior of the gravitational potential: At large distances or small momenta

$$p \ll m , \qquad r \gg L_{\text{DGP}} , \tag{3.4}$$

the denominator is dominated by the second term, and the propagator scales like $1/p$, whose 3D Fourier transform is $1/r^2$. This is the expected behavior for 5D gravity. And in fact the overall normalization is also the expected one: $\frac{1}{M_4^3}\frac{1}{m} = \frac{1}{M_5^3}$, which is the 5D Newton constant. At small distances or large momenta

$$p \gg m , \qquad r \ll L_{\text{DGP}} , \tag{3.5}$$

the denominator is dominated by the first term, and the propagator scales like $1/p^2$, whose 3D Fourier transform is $1/r$. This is the correct behavior for 4D gravity, and we can identify the $\frac{1}{M_4^2}$ prefactor with Newton's constant, $\frac{1}{M_{\text{Pl}}^2}$.

So, from a purely 4D perspective, DGP is an infrared modification of the gravitational force characterized by a critical distance L_{DGP}, beyond which the gravitational potential weakens from $1/r$ to $1/r^2$. L_{DGP} is know as the 'crossover scale'. If we want the 5D regime to kick in at cosmological distances, we need

$$m \sim H_0 , \qquad L_{\text{DGP}} \sim H_0^{-1} , \tag{3.6}$$

which, given the definition (3.2), corresponds to a huge hierarchy between the 4D Planck mass and the 5D one

$$M_5 \ll M_4 . \tag{3.7}$$

With an abuse of notation, sometimes we will refer to the parameter m as the 'graviton mass', simply to remind ourselves that it divides the two drastically different regimes $p \ll m$ and $p \gg m$.

Now let's turn our attention to the tensor structure of the propagator. It is identical to that of a massive graviton. In particular, there is a vDVZ discontinuity, which means that effectively there is a scalar degree of freedom that, at least at linear order, does not decouple from conserved sources when $m \to 0$, i.e., when $p \gg m$. For massive gravity we learned a lot by using a formulation of the theory that makes this scalar dof manifest. Fortunately here we can do the same, although

[6] For an explicit study of the analytic structure of the DGP propagator, see [5]

finding the analog of the Stückelberg parameterization is not straighforward. I will quote here the final result only, without deriving it.

Since the stress-energy tensor is localized on the boundary, it is convenient to 'integrate out' the bulk modes and rewrite the theory as a 4D theory. Namely, one can solve the linearized 5D Einstein equations for given sources and given 4D boundary conditions, plug the solution back into the action, and thus get an effective four-dimensional action that only involves the sources and the 4D field configuration on the boundary. The downside of doing so is that such an effective action is non-local, because massless degrees of freedom have been integrated out. Remarkably, at short-distances the 4D effective action is approximately local [7]. During this process, one discovers that there is an analog of the Stückelberg ϕ field, the so-called brane-bending mode π: in one gauge it measures fluctuations in the position of the brane along the fifth dimension; in another gauge the brane sits at $y = 0$, and π appears as a component of the five-dimensional metric. The gauge-invariant, physical statement is that, like for massive gravity, in the short distance limit

1. there is a scalar dof that couples directly to $T_{\mu\nu}$, and thus mediates a physically measurable force;

2. there are large self-interactions for this scalar, which are much more relevant than the standard non-linearities of GR.

In particular, at distances much shorter than $1/m = L_{\rm DGP}$, the 4D effective action reduces simply to

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\rm EH}[\hat{h}^c_{\mu\nu}] + \tfrac{1}{2}\tfrac{1}{M_{\rm Pl}}\,\hat{h}^c_{\mu\nu}T^{\mu\nu} \\ - 3(\partial\pi^c)^2 - \tfrac{1}{\Lambda_3^3}(\partial\pi^c)^2\Box\pi^c + \tfrac{1}{2}\tfrac{1}{M_{\rm Pl}}\,\pi^c T\,.\end{aligned} \tag{3.8}$$

Like in the case of massive gravity, the first line describes linearized GR, for the canonically normalized graviton $\hat{h}^c_{\mu\nu}$ field. The second line describes a scalar field, universally coupled to matter's stress-energy tensor, with a peculiar cubic self-interaction. The scale suppressing this interaction is the same Λ_3 defined in (2.64)—where of course m should now be interpreted as our new m parameter.

To arrive to this simple form of the action, one neglects terms that are suppressed for $p \gg m$. In particular, there is another field playing the role of the Stückelberg A_μ, which does not couple to external sources, and whose interactions with π and $\hat{h}$ are negligible at large momenta. More importantly, as we mentioned above, there are actually *non-local* terms that are negligible at large momenta, but become relevant at momenta of order of m or lower. For instance, at the quadratic action level, they take the form $m\,\pi^c\sqrt{-\Box}\,\pi^c$, $m\,\hat{h}^c_{\mu\nu}\sqrt{-\Box}\,\hat{h}^{c\mu\nu}$, etc., where $\sqrt{-\Box}$ is a placeholder for $\sqrt{p^2}$ in Fourier transform (these corrections are of course related to the funny $|p|$ in the propagator). These are subleading for $p \gg m$, w.r.t. the kinetic terms we explicitly kept in (3.8), but they become important at larger length scales. And this is unavoidable: we

[7] At least for the quadratic action, this is related to the statement that linearized gravity looks four-dimensional at short distances.

cannot hope to rewrite the original 5D model as a local 4D one *exactly.* The fact that we can do it approximately at least in some interesting distance range is quite remarkable already, and deserves being exploited.

A final comment: the direct coupling between π and T comes—like in massive a gravity—from a redefinition of the graviton field needed to demix the graviton from π:

$$h_{\mu\nu} = \hat{h}_{\mu\nu} + m\pi\, \eta_{\mu\nu} \; , \tag{3.9}$$

where $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$ is the perturbation in the 4D induced metric that appears in the original action, eq. (3.1), and π is related to the canonically normalized scalar by $\pi^c = mM_{\rm Pl}\, \pi$. [8]

3.2 Self-accelerating solution

As we mentioned, one of the attractive features of DGP is a cosmological solution that accelerates in the absence of a 4D cc, i.e., of a brane tension, or of any other form of energy. Geometrically, in the 5D picture, the solution is very simple: the bulk is flat—5D Minkowski space—and the brane is curled up into a maximally-symmetric hyperboloid, obeying

$$\eta_{MN} X^M X^N = L_{\rm DGP}^2 \; . \tag{3.10}$$

This is the *definition* of four-dimensional deSitter space. So, the statement is that the field equations deriving from the action (3.1) in the absence of sources, admit a solution that looks like 4D deSitter space, with Hubble rate

$$H = m = \frac{1}{L_{\rm DGP}} \; . \tag{3.11}$$

It is natural to wonder whether this can be the origin of our universe's acceleration.

To study this solution, rather than using the full 5D picture, it is much easier and illuminating to restrict to short distances $r \ll L_{\rm DGP}$, and use the 4D effective action (3.8). From this simplified (but correct) picture we will *(a)* recover the self-accelerating solution, and *(b)* show that it is unstable against π-perturbations, and thus, untenable. In the 4D description (3.8), the existence of a non-trivial solution such as the self-accelerating one relies crucially on π's self-interactions. Indeed, by varying the action w.r.t. π we get the non-linear eom

$$3\,\Box\pi^c + \tfrac{1}{\Lambda_3^3}\big((\Box\pi)^2 - (\partial_\mu\partial_\nu\pi^c)^2\big) = -\tfrac{1}{4}\tfrac{1}{M_{\rm Pl}} T \; . \tag{3.12}$$

We see that in principle, even in the absence of sources ($T \to 0$ in the r.h.s.), we can have non-trivial solutions sustained by the non-linear terms, as long as π's second derivatives are large enough, $\partial^2\pi^c \sim \Lambda_3^3$. In particular, we are after a maximally symmetric solution, with non-trivial second-derivatives, so that a natural ansatz is

$$\pi^c = \tfrac{1}{2} A\, x_\alpha x^\alpha \; , \tag{3.13}$$

[8] It is this π, with dimensions of length, that measures in some gauge the brane displacement in the fifth dimension.

with constant A. Plugging this into the eom and using $\partial_\mu\partial_\nu(x_\alpha x^\alpha) = 2\,\eta_{\mu\nu}$, we get a quadratic equation for A:

$$12A + \tfrac{1}{\Lambda_3^3} 12A^2 = 0 \ , \tag{3.14}$$

with a trivial solution, $A = 0$, and a non-trivial one, $A = -\Lambda_3^3$. The guess—which we are now going to check—is that this non-trivial solution corresponds precisely to the short-distance behavior of the self-accelerating solution described above.

To see this, consider first of all π's contribution to the physical metric perceived by matter. Recall that to write the 4D action in the nice demixed form (3.8) one had to redefine the metric perturbation according to (3.9). Moreover, in the absence of sources the $\hat h_{\mu\nu}$ field has no non-trivial maximally symmetric solution: $\hat h_{\mu\nu} = 0$. So, the 4D metric perturbation associated with our π solution is simply

$$h_{\mu\nu} = m\pi\,\eta_{\mu\nu} = \tfrac{1}{M_{\rm Pl}}\,\pi^c\,\eta_{\mu\nu} = m^2\,(x_\alpha x^\alpha)\,\eta_{\mu\nu} \ . \tag{3.15}$$

Is this what deSitter space looks like on short distances, $r \ll H^{-1}$? To address this question, consider the hyperboloid (3.10) in the vicinity of the point $X^M = (0,0,0,0,L_{\rm DGP})$. (The hyperboloid is maximally symmetric, so any point is as good as any other.) In a small neighborhood of this point, we can identify the x^μ coordinates along the hyperboloid with the first four of the 5D Minkwoski coordinates:

$$X^\mu = x^\mu \ , \qquad \mu = 0, \ldots 3 \ . \tag{3.16}$$

Then, the hyperboloid equation (3.10) gives the fifth coordinate X^4 as a function of the first four. We get

$$X^4(x) = \sqrt{L_{\rm DGP}^2 - x_\mu x^\mu} \simeq L_{\rm DGP} - \tfrac{1}{2} m\, x_\mu x^\mu \ , \tag{3.17}$$

where in the second step we focused on small distances from the origin, $x \ll L_{\rm DGP}$. We can then compute very easily the induced metric in this 4D coordinate system:

$$g_{\mu\nu} = \frac{\partial X^M}{\partial x^\mu}\frac{\partial X^N}{\partial x^\nu}\eta_{MN} \simeq \eta_{\mu\nu} + m^2\, x_\mu x_\nu \ , \tag{3.18}$$

which corresponds to the metric perturbation

$$h_{\mu\nu} = m^2\, x_\mu x_\nu \ . \tag{3.19}$$

This does not look the same as (3.15), but in fact it is, being related to it by an infinitesimal diff with $\xi_\mu = \frac{1}{4} m^2\, x_\mu\, x_\alpha x^\alpha$.

Now that we convinced ourselves that the 4D effective description captures correctly the short-distance behavior of the self-accelerating solution, we can forget about the complicated 5D picture (and equations, which we did not even write down), and just use the 4D description for short distance questions. For instance, we can ask whether the solution we just found is stable against short wavelength perturbations, with $k \gg m$. This is of course a necessary condition for overall stability. To answer this question, we just take the action (3.8) and expand it to second order in perturbations about the deSitter background

$$\pi^c = \pi^c_{\rm dS} + \varphi \,, \qquad \pi^c_{\rm dS} = -\tfrac{1}{2}\Lambda_3^3\, x_\alpha x^\alpha \,. \tag{3.20}$$

We can focus on the scalar sector—the $\hat{h}_{\mu\nu}$ sector is just linearized GR in Minkowski space, whose dynamics we know very well. We get

$$S_\varphi = \int d^4x \left[-3(\partial\varphi)^2 - \frac{1}{\Lambda_3^3}\left(\Box\pi^c_{\rm dS}\,(\partial\varphi)^2 + 2\,\partial_\mu\pi^c_{\rm dS}\,\partial^\mu\varphi\,\Box\varphi\right)\right] . \tag{3.21}$$

By using the identity $\partial^\mu\varphi\Box\varphi = \partial_\nu[\partial^\nu\varphi\partial^\mu\varphi - \frac{1}{2}\eta^{\mu\nu}(\partial\varphi)^2]$, and integrating by parts, we find

$$S_\varphi = \int d^4x \left[-3(\partial\varphi)^2 + \frac{2}{\Lambda_3^3}\left(\partial_\mu\partial_\nu\pi^c_{\rm dS} - \eta_{\mu\nu}\Box\pi^c_{\rm dS}\right)\partial^\mu\varphi\,\partial^\nu\varphi\right] . \tag{3.22}$$

So far we have not used anything about the specific form of our background solution. This quadratic action—which we will re-use later—describes how any π background affects the propagation of small perturbations. If now use $\pi^c_{\rm dS} = -\frac{1}{2}\Lambda_3^3\, x_\alpha x^\alpha$, we simply get

$$S_\varphi = \int d^4x\, 3(\partial\varphi)^2 \,, \tag{3.23}$$

which looks like the original kinetic term, but, crucially, with the *opposite* sign. We thus see that π's perturbations about the self-accelerating solution have negative kinetic energy. They thus yield ghost-like instabilities, and make the solution untenable. This result has been confirmed via (much more laborious) fully five-dimensional analyses.

3.3 Vainshtein effect

With the simple action we have for π, we can check whether DGP features the Vainshtein effect: is the scalar screened at short distances from massive sources, thus eliminating the vDVZ discontinuity and recovering agreement with solar system tests?

Let's first run some order-of-magnitude estimates. Consider π's eom (3.12) in the presence of a point-like source, $T = -M_*\delta^3(\vec{x})$. At large distances, where the linearized approximation holds, the solution is roughly

$$\pi^c \sim \frac{M_*}{M_{\rm Pl}}\frac{1}{r} \,, \qquad r\to\infty \,. \tag{3.24}$$

Plugging this solution into the non-linear terms, we discover that they become of the same order as the linear one at distances of order of the Vainshtein radius

$$R_V = \frac{1}{\Lambda_3^3}\left(\frac{M_*}{M_{\rm Pl}}\right)^{1/3} \sim \left(R_S\, m^{-2}\right)^{1/3} . \tag{3.25}$$

This is a huge distance. For solar-mass sources, it corresponds to roughly 10^{15} km. At much shorter distances, we expect the non-linear terms to dominate, and assuming π behaves in a power-law fashion

$$\pi \sim A\, r^\alpha \ , \qquad r \to 0 \ , \tag{3.26}$$

we get [9]

$$\frac{1}{\Lambda_3^3}(\partial^2\pi)^2 \sim \frac{1}{\Lambda_3^3} A^2\, r^{2\alpha-4} \sim \frac{M_*}{M_{\rm Pl}}\frac{1}{r^3} \ , \tag{3.27}$$

that is, $\alpha = 1/2$ and $A^2 \sim \Lambda_3^3 M_*/M_{\rm Pl}$. It appears that the short-distance behavior of π is very suppressed compared to the linearized approximation—$\sqrt{r}$ vs. $1/r$.

This has a direct implication for the force mediated by π: a non-relativistic test particle moving in this π field will feel an acceleration due to π of order $\frac{1}{M_{\rm Pl}}\vec\nabla\pi^c$. You can convince yourselves that this is correct by recalling that the physical metric to which matter is minimally coupled is affected by π as in eq. (3.9). So, if our estimates are correct, it looks like the vDVZ discontinuity is gone, and we recover agreement with solar system tests: at distances much shorter than the Vainshtein radius—which is huge, much bigger than the solar system—the force mediated by the scalar increases only like $1/\sqrt{r}$, which quickly becomes negligible w.r.t. the gravitational force, which scales like $1/r^2$.

It is actually straightforward to find an exact spherically symmetric solution to π's non-linear equation of motion. The reason is that the eom is a non-linear generalization of Gauss's law, which can then be integrated straightforwardly and cast as an *algebraic* equation for $\pi'(r)$, which is directly the physical quantity we are interested in. There is a deep reason why this happens, which is spelled out in [30]. Let's first rewrite (3.12) as

$$3\,\Box\pi^c - \tfrac{1}{2}\tfrac{1}{\Lambda_3^3}\Box(\partial\pi^c)^2 + \tfrac{1}{\Lambda_3^3}\partial_\mu(\partial^\mu\pi^c\Box\pi^c) = -\tfrac{1}{4}\tfrac{1}{M_{\rm Pl}}T \ . \tag{3.28}$$

You can just expand the derivatives and convince yourselves that this is equivalent to (3.12). But now we see that the l.h.s. is a total divergence

$$\partial_\mu\Big[3\,\partial^\mu\pi^c - \tfrac{1}{2}\tfrac{1}{\Lambda_3^3}\partial^\mu(\partial\pi^c)^2 + \tfrac{1}{\Lambda_3^3}\partial^\mu\pi^c\Box\pi^c\Big] = -\tfrac{1}{4}\tfrac{1}{M_{\rm Pl}}T \ . \tag{3.29}$$

In particular, for a static configuration with a point-like source we have

$$\vec\nabla\cdot\Big[3\,\vec\nabla\pi^c - \tfrac{1}{2}\tfrac{1}{\Lambda_3^3}\vec\nabla(\vec\nabla\pi^c)^2 + \tfrac{1}{\Lambda_3^3}\vec\nabla\pi^c\nabla^2\pi^c\Big] \equiv \vec\nabla\cdot\vec E = \tfrac{1}{4}\tfrac{M_*}{M_{\rm Pl}}\,\delta^3(\vec x) \ . \tag{3.30}$$

This is Gauss's law for the vector field $\vec E$, which is, in general, a non-linear combination of π's first and second derivatives. For spherically symmetric solutions, Gauss's law can be integrated as usual—by integrating both sides in a spherical volume, and then using Gauss's theorem for the l.h.s. One gets

$$\vec E(r) = \frac{M_*}{16\pi M_{\rm Pl}}\frac{\hat r}{r^2} \ . \tag{3.31}$$

[9]For our order-of-magnitude estimates, we can replace the $\delta^3(\vec x)$ on the r.h.s. with $1/r^3$, since it has the scaling property $\delta^3(\omega\vec x) = 1/\omega^3\cdot\delta^3(\vec x)$, and it contains no dimensionful quantities but r.

Now, it turns out that when π depends on r only, there is a simplification in the expression of $\vec{E}$ in terms of π's derivatives,

$$-\tfrac{1}{2}\vec{\nabla}(\vec{\nabla}\pi)^2 + \vec{\nabla}\pi\nabla^2\pi \to \hat{r}\big[-\tfrac{1}{2}\tfrac{d}{dr}(\pi')^2 + \pi'\big(\pi'' + \tfrac{2}{r}\pi'\big)\big] = \tfrac{2}{r}\,\hat{r}(\pi')^2 \,, \tag{3.32}$$

and $\vec{E}$ ends up depending (non-linearly) on π' only:

$$\vec{E} = \hat{r}\Big[3\,\pi_c' + \frac{1}{\Lambda_3^3}\frac{2(\pi_c')^2}{r}\Big] \tag{3.33}$$

Our solution for $\vec{E}$—eq. (3.31)—can thus be interpreted as a quadratic equation for $\pi'(r)$, with solutions

$$\pi_c'(r) = \frac{\Lambda_3^3}{4r}\Big[\pm\sqrt{9r^4 + \tfrac{1}{2\pi}\,R_V^3 r} - 3\,r^2\Big] \,, \tag{3.34}$$

where R_V is the Vainshtein radius defined above. The + solution is the one we should pick—it decays like $1/r^2$ at large r, while the other blows up like r^2. At very small r, it behaves as

$$\pi'(r \ll R_V) \sim \frac{1}{\sqrt{r}} \,, \tag{3.35}$$

as predicted.

3.4 The galileon

The scalar self-interaction in DGP, $(\partial\pi)^2\Box\pi$, has two remarkable mathematical properties, which we implicitly used in our analysis above:

1. Despite having a higher-derivative structure—there is a π acted upon by two derivatives—it yields equations of motion that involve exactly *two* derivatives on each field. This requires non-trivial cancellations: in deriving the eom from the action, one gets terms with three derivatives on a single field, like in eq. (3.28). They all cancel.

2. Once expanded to second order in fluctuations about *any* background, even though there are formally three-derivative terms, one can integrate them by parts and get a purely two-derivative action for the perturbations, eq. (3.22). The (tensor) coefficient that multiplies the perturbations' derivatives is made up of the background field's second derivatives.

From a technical viewpoint, these two properties make the model particularly tractable. The eom can be thought of as an *algebraic* (non-linear) equation for a single tensor, $\partial_\mu\partial_\nu\pi$. With an high enough degree of symmetry, it can be solved straightforwardly. Moreover, the quadratic action for small perturbations of any given solution depends algebraically on the same tensor $\partial_\mu\partial_\nu\pi$, evaluated on the background solution. That is, the background eom can alternatively be thought of directly as an algebraic equation for the kinetic term of small perturbations. This makes addressing questions of stability of a given solution particularly easy. Indeed, sometimes these questions can be answered without even finding the background solution. For instance, it can be proved in absolute generality that in the presence of non-relativistic, positive-energy sources, a solution that decays at infinity is stable everywhere [29].

More importantly, from a physical standpoint, the properties above make the model particularly well behaved. The fact that the eom are second order, and, related to this, that the action for small perturbations about any given background has two derivatives, means that we do not have the ghost problems associated with higher-derivative equations of motion that caused us so much headache in the case of massive gravity. Of course, for any given background, we still have to make sure that the perturbations' kinetic energy has the right sign—for instance for the self-accelerating solution, it does not—but this is now demoted from being an unavoidable pathology of the theory to a background-dependent question.

There is another interesting property, not directly related to π's internal dynamics, but rather to how it contributes to the geometry perceived by matter:

3. π's contributes to the geometry only via its second derivatives, $\partial_\mu \partial_\nu \pi$.

This seems to contradict the relation between the metric and π that we have been using so far, eq. (3.9). However recall that the metric is not physical—i.e., observable—only its gauge invariant combinations are. These are the curvature tensors, and their covariant derivatives. In particular, since in the limit we are considering we are keeping non-linearities only in the dynamics of π, but never in the dynamics of the metric nor in π's contribution to the metric, then the curvature tensors are simply gauge-invariant combinations of $h_{\mu\nu}$'s second derivatives. Via eq. (3.9), these get 'contaminated' by the second derivatives of π. For instance, for the Ricci tensor eq. (3.9) implies

$$R_{\mu\nu} = \hat{R}_{\mu\nu} - m\big(\partial_\mu \partial_\nu \pi + \tfrac{1}{2}\eta_{\mu\nu}\Box\pi\big) \ . \tag{3.36}$$

Then, the non-linear equation for π, which as we emphasized is algebraic in the tensor $\partial_\mu \partial_\nu \pi$, can also be thought of as an algebraic non-linear equation directly for π's corrections to the geometry. This makes the Vainshtein effect potentially very robust: we checked that it works for an idealized situation involving a spherical source in otherwise empty space. But what about a more realistic situation, in which we have many sources, with overlapping Vainsthein regions (recall that the Vainshtein radius is huge), and no symmetry whatsoever? At any given point x, given the local sources at x, the eom decides immediately whether π's contribution to the local curvature is unscreened (linear regime, small $\partial\partial\pi$), or screened (non-linear regime, large $\partial\partial\pi$). The model has a chance of implementing the Vainshtein effect automatically in a wide range of physical situations.

Properties 1, 2, and 3 above invite interpreting $\partial_\mu \partial_\nu \pi$, rather that π or $\partial_\mu \pi$, as the physically relevant quantity. This can be rephrased as a symmetry statement: the physics of the model is invariant under transformations of π and $\partial_\mu \pi$ that do not change $\partial_\mu \partial_\nu \pi$. The most general such a transformation is

$$\pi(x) \to \pi(x) + c + b_\mu x^\mu \ , \tag{3.37}$$

with constant c and b^μ. It is natural to ask whether the DGP scalar self-interactions is unique in this respect, or whether it admits interesting generalizations.

First of all, notice that (3.37) is a field-theoretical, four-dimensional generalization of the *galilean transformations* of mechanics:

$$\vec{x}(t) \to \vec{x}(t) + \vec{x}_0 + \vec{v}_0\, t \; . \tag{3.38}$$

In our field-theory case, x^μ plays the role of t, and $\pi(x)$ plays the role of the field $\vec{x}(t)$. So, in the following, we will refer to (3.37) as 'galilean shift'. Second, notice that naively all invariants we can write down have to involve at least two derivatives acting on each π—precisely because the symmetry was designed to select $\partial_\mu \partial_\nu \pi$ and discard π and $\partial_\mu \pi$. And, in fact, in the DGP eom for the scalar, there are only second derivatives. However the action has a non-trivial structure, involving first derivatives as well. In what sense is it invariant under our galilean shift (3.37)? It is, up to total derivatives. Indeed, performing (3.37) we get

$$\delta\big((\partial\pi)^2 \Box\pi\big) = 2b_\mu\, \partial^\mu \pi \Box \pi = 2b_\mu \partial_\nu \left[\partial^\nu \pi \partial^\mu \pi - \tfrac{1}{2}\eta^{\mu\nu}(\partial\pi)^2\right] \; , \tag{3.39}$$

which is a total divergence.

Notice in passing that the kinetic term, $(\partial\pi)^2$, has precisely the same property,

$$\delta\big((\partial\pi)^2\big) = 2b_\mu\, \partial^\mu \pi \; , \tag{3.40}$$

which has an obvious analog in Newtonian mechanics: the kinetic energy $\frac{1}{2} m\, \dot{\vec{x}}^{\,2}$, is *not* invariant under a Galileo boost—it changes by a total derivative $m\, \vec{v}_0 \cdot \dot{\vec{x}}$, which is enough to guarantee the invariance of the eom.

To summarize, if we want to impose the symmetry (3.37) at the level of the action, we have two options: we have 'obvious' invariants, which involve at least two derivatives on each field, or we can look for less obvious ones, which involve fewer derivatives per field, and are invariant only up to total derivatives. The first class of invariants is problematic: we are going to get higher derivative equations of motion, with associated ghost instabilities, like in the case of (non-improved) massive gravity. The second possibility, more in line with what we learned from the DGP example, is more promising: with the right overall number of derivatives, we can hope to get an eom that involves no more than two derivatives per field, and thus no ghost associated with higher derivative terms. On the other hand, for the eom to be invariant under the galilean shift, it has to involve *at least* two derivatives per field. We are thus let to consider actions that give us eom with *exactly* two derivatives per field, like in the DGP case.

We expect these Lagrangian terms to have the schematic form

$$\mathcal{L}_n \sim \partial\pi\partial\pi\, (\partial^2\pi)^{n-2} \; , \tag{3.41}$$

so that once we vary w.r.t. π to derive the eom, we end up with the right number of derivatives—two per field. Notice that the DGP interaction has precisely this form, with $n = 3$. So does the kinetic term, with $n = 2$. Of course not all possible Lorentz contractions of a structure like this are going to work for our purposes: the invariance under galilean shifts, or equivalently, the cancellation of all three-derivative and four-derivative terms in the eom, is going to require some non-trivial tunings.

It turns out that for a single scalar π, there is exactly one galilean-invariant structure of this form at each order in π. Moreover, in four spacetime dimensions, we stop at fifth order (in D dimensions we stop at order $D+1$). The actual galilean invariant combinations are

$$\mathcal{L}_1 = \pi \tag{3.42}$$
$$\mathcal{L}_2 = -\tfrac{1}{2}\,\partial\pi\cdot\partial\pi \tag{3.43}$$
$$\mathcal{L}_3 = -\tfrac{1}{2}\,[\Pi]\,\partial\pi\cdot\partial\pi \tag{3.44}$$
$$\mathcal{L}_4 = -\tfrac{1}{4}\big([\Pi]^2\,\partial\pi\cdot\partial\pi - 2\,[\Pi]\,\partial\pi\cdot\Pi\cdot\partial\pi - [\Pi^2]\,\partial\pi\cdot\partial\pi + 2\,\partial\pi\cdot\Pi^2\cdot\partial\pi\big) \tag{3.45}$$
$$\begin{aligned}\mathcal{L}_5 = -\tfrac{1}{5}\big(&[\Pi]^3\,\partial\pi\cdot\partial\pi - 3[\Pi]^2\,\partial\pi\cdot\Pi\cdot\partial\pi - 3[\Pi][\Pi^2]\,\partial\pi\cdot\partial\pi + 6[\Pi]\,\partial\pi\cdot\Pi^2\cdot\partial\pi\\ &+2[\Pi^3]\,\partial\pi\cdot\partial\pi + 3[\Pi^2]\,\partial\pi\cdot\Pi\cdot\partial\pi - 6\,\partial\pi\cdot\Pi^3\cdot\partial\pi\big)\end{aligned} \tag{3.46}$$

where we are using the following notation: Π stands for the matrix of second derivatives of π, $\Pi^\mu{}_\nu \equiv \partial^\mu\partial_\nu\pi$; the brackets $[\ldots]$ stand for the trace operator; the '$\cdot$' stands for the standard Lorentz-invariant contraction of indices. So, for instance, in $\mathcal{L}_3$

$$[\Pi]\,\partial\pi\cdot\partial\pi \equiv \Box\pi\,\partial_\mu\pi\partial^\mu\pi\ , \tag{3.47}$$

which is precisely the DGP cubic interaction. The overall normalizations have been chosen to have simple normalizations in the equations of motion. Defining $\mathcal{E}_i \equiv \frac{\delta\mathcal{L}_i}{\delta\pi}$, one gets

$$\mathcal{E}_1 = 1 \tag{3.48}$$
$$\mathcal{E}_2 = \Box\pi \tag{3.49}$$
$$\mathcal{E}_3 = (\Box\pi)^2 - (\partial_\mu\partial_\nu\pi)^2 \tag{3.50}$$
$$\mathcal{E}_4 = (\Box\pi)^3 - 3\,\Box\pi(\partial_\mu\partial_\nu\pi)^2 + 2(\partial_\mu\partial_\nu\pi)^3 \tag{3.51}$$
$$\mathcal{E}_5 = (\Box\pi)^4 - 6(\Box\pi)^2(\partial_\mu\partial_\nu\pi)^2 + 8\,\Box\pi(\partial_\mu\partial_\nu\pi)^3 + 3\big[(\partial_\mu\partial_\nu\pi)^2\big]^2 - 6(\partial_\mu\partial_\nu\pi)^4 \tag{3.52}$$

where by $(\partial_\mu\partial_\nu\pi)^n$ we denote the cyclic contraction, $(\partial_\mu\partial_\nu\pi)^n \equiv [\Pi^n]$.

We can thus define the following generalization of the DGP four-dimensional effective theory:

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\rm EH}[\hat h_{\mu\nu}] + \tfrac{1}{2}\,\hat h_{\mu\nu}T^{\mu\nu}\\ + \sum_{i=1}^{5} c_i\,\mathcal{L}_i[\pi] + \tfrac{1}{2}\pi\,T\ ,\end{aligned} \tag{3.53}$$

which goes under the name of 'galileon theory'. It has five free coefficients, the c_i's. All $\mathcal{L}_i$'s are as remarkable as the DGP scalar interaction, that is, they all obey the properties 1, 2, and 3 that we discussed above. As a result, the system is physically very well behaved—there are no ghosts from higher derivative terms—and mathematically quite tractable—the eom, stability, and Vainshtein screening are all related *algebraically*.

Given its nice physical properties, the galileon provides us with a template for what a well-behaved IR modification of gravity should look like at sub-cosmological distances: if there is a short-distance range where the modification of the gravitational force can be attributed to the exchange of a single scalar field π—both massive gravity and DGP have this property—then this scalar should behave as a galileon. The fact that the improved version of massive gravity—which does not have ghost-instabilities from higher derivative terms—conforms to this expectation, is quite remarkable.

I would like to stress that, by construction, the galileon should be thought of as a description of some IR modification of gravity only at short distances. In all interesting examples, the gravity-plus-scalar splitting stops making sense at large distances: in massive gravity, at low momenta the scalar becomes one of the five physical components of a massive spin-2 state; in DGP, the four-dimensional scalar 'melts' into five-dimensional degrees of freedom (there are no stable gravitational degrees of freedom confined to the brane). We do not know whether the most general galileon theory can follow from some more complete theory defined also at large length scales. However, in the regimes where the galileon effective theory holds, the extra freedom it has w.r.t. DGP—those five free coefficients—make it perform better than DGP. In particular, one can adjust those coefficients to have *stable, deSitter-like solutions with Vainshtein-like non-linear perturbations, also stable against small fluctuations.* That is, a fairly realistic picture of our universe.

3.5 The problem: Superluminality

We conclude with an outstanding problem for these theories. Consider the DGP case for simplicity. Let's take the Vainshtein solution described in sect. 3.3, and let's consider small fluctuations about it. Their quadratic action is simply (3.22), with our Vainshtein background solution replacing the deSitter one. In spherical coordinates one gets

$$\mathcal{L}_\varphi = \Big[3 + \frac{2}{\Lambda_3^3}\Big(\pi_c'' + \frac{2\,\pi_c'}{r}\Big)\Big]\dot\varphi^2 - \Big[3 + \frac{4}{\Lambda_3^3}\,\frac{\pi_c'}{r}\Big](\partial_r\varphi)^2 - \Big[3 + \frac{2}{\Lambda_3^3}\Big(\pi_c'' + \frac{\pi_c'}{r}\Big)\Big](\partial_\Omega\varphi)^2 \, , \tag{3.54}$$

where $(\partial_\Omega\varphi)^2$ is the angular part of $(\vec\nabla\varphi)^2$,

$$(\partial_\Omega\varphi)^2 = \frac{1}{r^2}(\partial_\theta\varphi)^2 + \frac{1}{r^2\sin^2\theta}(\partial_\phi\varphi)^2 \, . \tag{3.55}$$

With the explicit form of the Vainshtein solution, eq. (3.34), one can check straightforwardly that all terms in brackets are positive, i.e., that the Lagrangian has the correct $(+,-,-,-)$ signature corresponding to stability (this is the case for the $+$ solution, which is the one we are supposed to choose.) So, the Vainshtein solution is stable against small perturbations, which is good.

There is however a subtler problem. Consider the *speed* at which fluctuations propagate. This is easy to compute from the Lagrangian above, for fluctuations that have wavelengths much shorter than the typical variation scale of the background field, $\lambda \ll r$. For those, about any point r, we can zoom in on a neighborhood of that point, much bigger than the fluctuations wavelength but

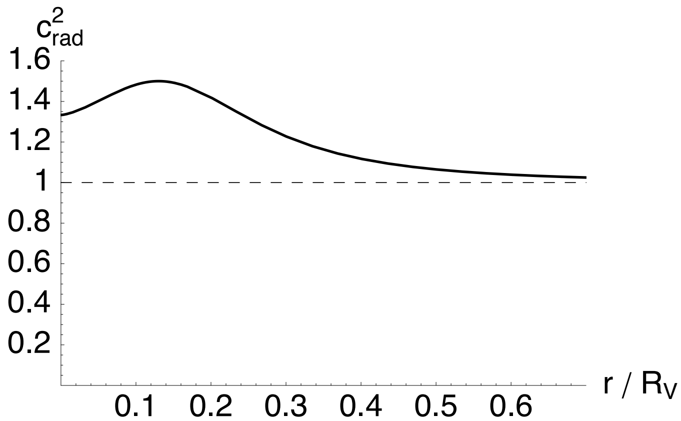

Figure 1: *The speed of radially moving fluctuations for a Vainshtein solution in DGP.*

much smaller than r. In that neighborhood we can treat the Lagrangian coefficients as constant—i.e., r-independent—and we can neglect the curvature of the angular coordinates θ and ϕ, and treat them as flat coordinates. If we factor out the (constant, in this approximation) coefficient of the $(\dot\varphi)^2$ term, the Lagrangian reduces to

$$\mathcal{L}_\varphi \to \dot\varphi^2 - c^2_{\rm rad}(\partial_r\varphi)^2 - c^2_{\rm ang}(\vec\nabla_\perp\varphi)^2 \; , \tag{3.56}$$

where $\vec\nabla_\perp$ denotes the gradient w.r.t. the transverse (i.e., angular) variables, and the quantities

$$c^2_{\rm rad} \equiv \frac{\left[3 + \frac{4}{\Lambda_3^3}\,\frac{\pi_c'}{r}\right]}{\left[3 + \frac{2}{\Lambda_3^3}\left(\pi_c'' + \frac{2\,\pi_c'}{r}\right)\right]} \; , \qquad c^2_{\rm ang} \equiv \frac{\left[3 + \frac{2}{\Lambda_3^3}\left(\pi_c'' + \frac{\pi_c'}{r}\right)\right]}{\left[3 + \frac{2}{\Lambda_3^3}\left(\pi_c'' + \frac{2\,\pi_c'}{r}\right)\right]} \tag{3.57}$$

can be interpreted as the radial and angular (squared) propagation speeds, respectively, because in this approximation the fluctuation φ obeys the wave equation

$$\ddot\varphi - c^2_{\rm rad}\,\partial_r^2\varphi - c^2_{\rm ang}\,\nabla_\perp^2\varphi = 0 \; . \tag{3.58}$$

The bottom line is that short-wavelength perturbations traveling in the radial or angular directions, have local speeds given by eq. (3.57).

It is just a fact that for the background solution eq. (3.34), $c^2_{\rm rad}$ is *larger than one* for any r. A plot of $c^2_{\rm rad}$ versus r is given in fig. 1: it starts from $4/3$ at $r=0$, reaches a maximum of $3/2$ at $r \sim R_V$ and asymptotes to 1 (from above!) for $r \to \infty$. This is an $\mathcal{O}(1)$ deviation from the speed of light in an enormous region of space; for instance for the Sun, R_V is $\sim 10^{15}$ km.

Unless one is able to construct closed time-like curves, superluminal signal propagation in a relativistic theory is not necessarily an inconsistency of the theory. However, recall that the theories we are talking about should only be thought of as low-energy effective field theories: they have a finite UV cutoff—quite low, in fact—beyond which the theory does not make quantum mechanical sense, and has to be replaced by a more complete theory—the UV completion. It is then natural to ask whether there can exist sensible Lorentz-invariant UV completions that admit superluminal signal propagation at low energies. As far as we know, the answer is 'no'. For instance, for the UV-complete theories that we normally use for particle physics, that is, renormalizable Lorentz-invariant QFTs, the answer is quite simple: canonical quantization combined with Lorentz invariance demand that local observables commute at space-like separations. This is an absolute, exact, non-perturbative statement that has to hold at the operator-level, i.e., for *any* state. Turning on a background field, like our Vainshtein solution above, corresponds in the quantum theory to choosing a particular state. But this cannot impair the property that local observables commute at spacelike distances—that is, there cannot be superluminal signal propagation in a relativistic renormalizable QFT.

One can then consider moving from DGP to a more generic galileon model, and trying to choose the c_i coefficients in such a way as to avoid this problem. However there is a very generic tension, which makes the superluminality problem particularly robust: it turns out that any choice of the c_i's compatible with the existence and stability of Vainshtein-like solutions—a desirable feature if we want the theory to describe the solar system—necessarily implies superluminal signal propagation.

4 Further Reading

The cosmological constant problem is reviewed extensively in the classic review by Weinberg [6], who also proposed an anthropic solution [7].

The QFT approach to GR was initiated by Weinberg, in two seminal papers [8, 9]. It was eventually completed by Deser [10], and by Boulware and Deser [11]. Feynman's gravity lectures [12] also take this approach. In modern QFT terms, it is spelled out in Weinberg's QFT texbook [13]. For a pedagogical review of the method, consistent with these lectures, see my lecture notes [14].

For massive gravity, the original Fierz-Pauli theory was constructed in [15]. The vDVZ discontinuity was pointed out by van Dam and Veltman [16], and by Zakharov [17]. The Vainshtein effect was proposed in [18]. The general problem with the sixth mode in the non-linear theory was pointed out by Boulware and Deser in [19]. The modern approach based on the Stückelberg fields was initiated in [20]. It was used in [21] to understand the Boulware-Deser sixth mode problem in effective field theory terms. This problem was recently solved by de Rham and Gabadadze in [22], which also pointed out the relation between the improved theory of massive gravity and the galileon. This gave rise to a lot of activity on the subject, part of which is reviewed in detail by Hinterbichler [23]. The methods that worked for massive gravity have also been extended to 'bi-metric' and 'multi-metric' theories—see e.g. [24].

The DGP model was introduced in [25]. Its self-accelerating cosmological solution was studied in [26, 27]. The 4D effective theory for the brane-bending mode was derived in [28], and further analyzed in [29], where many of its remarkable properties were pointed out. It was then generalized to the galileon effective field theory in [30]. The superluminality problem for these theories was pointed out—along with some associated S-matrix analiticity problems—in [31], and discussed further for the galileon in [30, 32, 33].

Appendix

A Ghosts from higher derivative kinetic terms

Let us be very specific about why higher derivative kinetic terms give rise to ghost-like instabilities—i.e., to excitations with negative kinetic energy. Consider for instance a massless scalar field ϕ with Lagrangian density

$$\mathcal{L} = -\tfrac{1}{2}(\partial\phi)^2 + \frac{a}{2\Lambda^2}(\Box\phi)^2 - V_{\rm int}(\phi) \tag{A.1}$$

(let me remind you that I am using the $(-,+,+,+)$ signature!), where Λ is some energy scale, a is a sign, $a = \pm 1$, and $V_{\rm int}$ is a self-interaction potential. Regardless of the sign a of the second term, the system is plagued by ghosts. To show this, it is particularly convenient to reduce to a purely two-derivative kinetic Lagrangian, from which we know how to extract the stability properties of the system. Let's introduce an auxiliary scalar field χ and a new Lagrangian

$$\mathcal{L}' = -\tfrac{1}{2}(\partial\phi)^2 - a\,\partial_\mu\chi\partial^\mu\phi - \tfrac{1}{2}a\,\Lambda^2\,\chi^2 - V_{\rm int}(\phi)\ , \tag{A.2}$$

which reduces to $\mathcal{L}$ *exactly* once χ is "integrated out" [10]. Indeed χ's eom reads

$$a\Box\phi - a\Lambda^2\chi = 0\ . \tag{A.3}$$

Plugging the solution

$$\chi = \Box\phi/\Lambda^2 \tag{A.4}$$

back into $\mathcal{L}'$, and integrating by parts we get

$$\int d^4x\,\mathcal{L}' = \int d^4x\,\mathcal{L}\ . \tag{A.5}$$

So the two Lagrangians, $\mathcal{L}$ and $\mathcal{L}'$, are physically equivalent. Yet $\mathcal{L}'$ is easier to interpret, since it involves at most first-derivatives of the fields. One could for instance derive the Hamiltonian in the usual way, and study its positivity properties. Alternatively, one can diagonalize $\mathcal{L}'$ via the substitution $\phi = \phi' - a\chi$. One gets

$$\mathcal{L}' = -\tfrac{1}{2}(\partial\phi')^2 + \tfrac{1}{2}(\partial\chi)^2 - \tfrac{1}{2}a\,\Lambda^2\,\chi^2 - V_{\rm int}(\phi',\chi)\ . \tag{A.6}$$

The (in)stability properties of the system are pretty manifest now: the system has two scalar degrees of freedom, one of which (χ) has negative kinetic energy.

[10] For classical systems, integrating out a degree of freedom simply means solving its equations of motion, and plugging back the solution into the action.

B The problem with ghosts

Notice in passing that in the example above χ can also be a tachyon, for $a = +1$: in this case χ has exponentially growing modes, for low enough k's,

$$\chi \sim e^{i\vec{k}\cdot\vec{x}} e^{-i\omega t} + \text{c.c.} \, , \qquad \omega = \sqrt{k^2 - a\Lambda^2} \, . \tag{B.1}$$

But let us neglect this possibility and concentrate on the ghost instability, i.e., on the negativity of χ's kinetic energy, which is unavoidable. A ghost, unlike a tachyon, is not unstable by itself: its equation of motion is perfectly healthy at the linear level, and does not admit any exponentially growing solution: its quadratic Lagrangian has an overall negative sign, but the equations motion are insensitive to the overall normalization of the Lagrangian, including its sign. The problem is that its Hamiltonian is negative-definite, so that when couplings to ordinary 'healthy' matter are taken into account (the potential term in our example above), the system is unstable: with zero net energy one can excite both sectors—the positive energy one and the negative energy one—without bound, and this exchange of energy happens spontaneously already at classical level. In a quantum system with ghosts in the physical spectrum this translates into an instability of the vacuum. The decay rate is UV divergent due to an infinite degeneracy of the final state phase space. It is impossible to cutoff this divergence in a Lorentz invariant way: the reason is that the initial state—the vacuum—is boost-invariant, so that if it can decay into some non-trivial final state, say made up of two ghosts and two gravitons, then it can decay with exactly the same amplitude into any boosted version of such a final state. Since Lorentz-boosts form a non-compact group, the phase space generated by applying these boosts has infinite volume. So, it looks like we cannot make much sense of a Lorentz-invariant theory with ghosts.

However the situation is not as bad as it seems: our ghost χ in eq. (A.6) has a (normal or tachyonic) mass Λ, so that it shows up only at energies of order of or greater than Λ, *i.e.* when the four derivative kinetic term in eq. (A.1) starts being comparable to the usual two derivative one. We can consistently use our scalar field theory eq. (A.1) at energies below Λ, and postulate that some new degree of freedom or some new physics enters at energies of order Λ and takes care of the ghost instability. For example, one could add a term $-(\partial\chi)^2$ to eq. (A.2) (for simplicity we stick to the non-tachyonic case $a = +1$, and we set $V_{\rm int} = 0$),

$$\mathcal{L}_{\rm UV} = -\tfrac{1}{2}(\partial\phi)^2 - \partial_\mu\chi\partial^\mu\phi - (\partial\chi)^2 - \tfrac{1}{2}\,\Lambda^2\,\chi^2 \, . \tag{B.2}$$

This drastically changes the high-energy picture, since the resulting Lagrangian obtained by de-mixing ($\phi = \phi' - \chi$), now describes two perfectly healthy scalars, one massless and the other with mass Λ,

$$\mathcal{L}_{\rm UV} = -\tfrac{1}{2}(\partial\phi')^2 - \tfrac{1}{2}(\partial\chi)^2 - \tfrac{1}{2}\,\Lambda^2\,\chi^2 \, . \tag{B.3}$$

At the same time, at energies below Λ the heavy field χ can be integrated out from $\mathcal{L}_{\rm UV}$,

$$\Box\phi + 2\Box\chi - \Lambda^2\chi = 0 \qquad \Rightarrow \qquad \chi = \frac{1}{\Lambda^2 - 2\Box}\Box\phi = \frac{1}{\Lambda^2}\Big[1 + 2\frac{\Box}{\Lambda^2} + 4\frac{\Box^2}{\Lambda^4} + \dots\Big]\Box\phi \tag{B.4}$$

thus giving the starting Lagrangian eq. (A.1) up to terms suppressed by additional powers of $(\Box/\Lambda^2)$. This simple example shows that in principle a ghost instability can be cured by suitable new physics at the scale Λ without modifying the low-energy behavior of theory. Because of this, a higher-derivative theory like eq. (A.1) makes perfect sense as an *effective field theory* with UV cutoff Λ.

References

[1] Z. Komargodski and A. Schwimmer, "On Renormalization Group Flows in Four Dimensions," JHEP **1112**, 099 (2011) [arXiv:1107.3987 [hep-th]].

[2] R. M. Wald, "General Relativity," *Chicago, Usa: Univ. Pr. (1984) 491p*

[3] S. Weinberg, "Gravitation and Cosmology," *John Wiley & Sons 1972*

[4] C. W. Misner, K. S. Thorne and J. A. Wheeler, "Gravitation," *San Francisco 1973, 12*

[5] K. Hinterbichler, A. Nicolis and M. Porrati, "Superluminality in DGP," JHEP **0909**, 089 (2009) [arXiv:0905.2359 [hep-th]].

[6] S. Weinberg, "The Cosmological Constant Problem," Rev. Mod. Phys. **61**, 1 (1989).

[7] S. Weinberg, "Anthropic Bound on the Cosmological Constant," Phys. Rev. Lett. **59**, 2607 (1987).

[8] S. Weinberg, "Photons and gravitons in S-matrix theory: derivation of charge conservation and equality of gravitational and inertial mass," Phys. Rev. **135**, B1049 (1964).

[9] S. Weinberg, "Photons and gravitons in perturbation theory: Derivation of Maxwell's and Einstein's equations," Phys. Rev. **138**, B988 (1965).

[10] S. Deser, "Selfinteraction and gauge invariance," Gen. Rel. Grav. **1**, 9 (1970) [gr-qc/0411023].

[11] D. G. Boulware and S. Deser, "Classical General Relativity Derived from Quantum Gravity," Annals Phys. **89**, 193 (1975).

[12] R. P. Feynman, F. B. Morinigo, W. G. Wagner and B. Hatfield, (ed.), "Feynman lectures on gravitation," Reading, USA: Addison-Wesley (1995) 232 p. (The advanced book program)

[13] S. Weinberg, "The Quantum theory of fields. Vol. 1: Foundations," *Cambridge, UK: Univ. Pr. (1995) 609 p.*

[14] A. Nicolis, lecture notes available for download at http://phys.columbia.edu/˜nicolis/G8099.html

[15] M. Fierz and W. Pauli, "On relativistic wave equations for particles of arbitrary spin in an electromagnetic field," Proc. Roy. Soc. Lond. A **173**, 211 (1939).

[16] H. van Dam and M. J. G. Veltman, "Massive and massless Yang-Mills and gravitational fields," Nucl. Phys. B **22**, 397 (1970).

[17] V. I. Zakharov, "Linearized gravitation theory and the graviton mass," JETP Lett. **12**, 312 (1970) [Pisma Zh. Eksp. Teor. Fiz. **12**, 447 (1970)].

[18] A. I. Vainshtein, "To the problem of nonvanishing gravitation mass," Phys. Lett. B **39**, 393 (1972).

[19] D. G. Boulware and S. Deser, "Can gravitation have a finite range?," Phys. Rev. D **6**, 3368 (1972).

[20] N. Arkani-Hamed, H. Georgi and M. D. Schwartz, "Effective field theory for massive gravitons and gravity in theory space," Annals Phys. **305**, 96 (2003) [hep-th/0210184].

[21] P. Creminelli, A. Nicolis, M. Papucci and E. Trincherini, "Ghosts in massive gravity," JHEP **0509**, 003 (2005) [hep-th/0505147].

[22] C. de Rham and G. Gabadadze, "Generalization of the Fierz-Pauli Action," Phys. Rev. D **82**, 044020 (2010) [arXiv:1007.0443 [hep-th]].

[23] K. Hinterbichler, "Theoretical Aspects of Massive Gravity," Rev. Mod. Phys. **84**, 671 (2012) [arXiv:1105.3735 [hep-th]].

[24] K. Hinterbichler and R. A. Rosen, "Interacting Spin-2 Fields," JHEP **1207**, 047 (2012) [arXiv:1203.5783 [hep-th]].

[25] G. R. Dvali, G. Gabadadze and M. Porrati, "4-D gravity on a brane in 5-D Minkowski space," Phys. Lett. B **485**, 208 (2000) [hep-th/0005016].

[26] C. Deffayet, "Cosmology on a brane in Minkowski bulk," Phys. Lett. B **502**, 199 (2001) [hep-th/0010186].

[27] C. Deffayet, G. R. Dvali and G. Gabadadze, "Accelerated universe from gravity leaking to extra dimensions," Phys. Rev. D **65**, 044023 (2002) [astro-ph/0105068].

[28] M. A. Luty, M. Porrati and R. Rattazzi, "Strong interactions and stability in the DGP model," JHEP **0309**, 029 (2003) [hep-th/0303116].

[29] A. Nicolis and R. Rattazzi, "Classical and quantum consistency of the DGP model," JHEP **0406**, 059 (2004) [hep-th/0404159].

[30] A. Nicolis, R. Rattazzi and E. Trincherini, "The Galileon as a local modification of gravity," Phys. Rev. D **79**, 064036 (2009) [arXiv:0811.2197 [hep-th]].

[31] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis and R. Rattazzi, "Causality, analyticity and an IR obstruction to UV completion," JHEP **0610**, 014 (2006) [hep-th/0602178].

[32] A. Nicolis, R. Rattazzi and E. Trincherini, "Energy's and amplitudes' positivity," JHEP **1005**, 095 (2010) [Erratum-ibid. **1111**, 128 (2011)] [arXiv:0912.4258 [hep-th]].

[33] P. Creminelli, A. Nicolis and E. Trincherini, "Galilean Genesis: An Alternative to inflation," JCAP **1011**, 021 (2010) [arXiv:1007.0027 [hep-th]].