# The cosmological constant

Austin Joyce

May 19, 2016

**Abstract**

Notes on the cosmological constant problem.

## Contents



## 1 The cosmological constant in Einstein gravity

Einstein's equations read

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{1}{M_{\rm Pl}^2}T_{\mu\nu}, \tag{1.1}$$

where $R_{\mu\nu}$ is the Ricci curvature tensor and $T_{\mu\nu}$ is the stress-energy tensor for matter components. The principle underlying these equations is diffemorphism covariance—under a diffeomorphism, these objects transform like tensors. Einstein gravity may be though of as the lowest order term in a derivative expansion of the metric: $R \sim \partial^2 g$.

Aside from higher-derivative corrections $R^2 \sim \partial^4 g$, there is a *lower* derivative term which can be added to Einstein's equations, which is also diffeomorphism covariant $\sim \Lambda g_{\mu\nu}$ so that we have

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{1}{M_{\rm Pl}^2} T_{\mu\nu}. \tag{1.2}$$

This term is known as the *cosmological constant* (CC). The possibility of adding this term was actually realized by Einstein himself [1], though his motivation was to construct a static cosmological model.

For a long time (for reasons that we will elaborate on), the cosmological constant was thought to be exactly zero in our universe. Consequently, it was very surprising in 1998 when observations of type Ia supernovae led to the discovery that expansion of the universe is accelerating [2, 3], implying that $\Lambda$ is nonzero and positive. The observed value of the CC is rather small, in particle physics units $\Lambda_{\rm obs.} \sim (10^{-3}{\rm eV})^4$, while in units more natural to gravity,

$$\Lambda_{\rm obs.} \sim (10^{-30} M_{\rm Pl})^4. \tag{1.3}$$

This is a very small number. In what follows, we will try to understand why this should concern us—the smallness of the observed $\Lambda$ is the essence of the *cosmological constant problem*. Our naïve expectation is that the natural value for the CC to take is $\sim M_{\rm Pl}^4$, so we would like to understand why this is not the case.

In what follows we will elucidate the precise nature of the cosmological constant problem, and point out how it differs from other "smallness" problems in physics. After understanding this, we will discuss some possible ways to address the CC problem, and difficulties they face. Off the bat it should be said that there are no known solutions to the cosmological constant problem—if you find one, please let me know!

An inspiring review of the cosmological constant problem (from which much of the following is drawn) is Weinberg's article [4]. Other nice reviews include [5], which examines a wide variety of possible solutions (and shows why they don't work) and [6], which shows in great detail many of the calculations underlying our understanding of the nature of the CC problem.

## 2 The cosmological constant problem

In this Section, we will describe what precisely is meant by the cosmological constant "problem." It should be stressed that this problem is not a problem in the usual sense—it is not an inconsistency between experiment and our description of a system, and does not arise in a regime where we expect our approximations to break down. Instead the CC problem is one of *naturalness*. This is a very loaded word, it has a precise technical meaning in the context of quantum field theory, which we will review, which goes beyond the aesthetic discomfort that the number appearing in (1.3) is small.

Oftentimes people separate the CC problem into the "old" CC problem and the "new" CC problem; the distinction is essentially the following:

- **Old:** Why is the CC not large?
- **New:** Why is the CC small (but nonzero)? And why is it comparable to the matter energy density today?

To me this is not really a useful distinction to make—the real problem is the "old" problem, as we will see.

What is our expectation for the size of the CC anyway? We expect at least a contribution[1] to the cosmological constant of the form [4]

$$\langle T_{\mu\nu}\rangle \sim -\langle\rho\rangle g_{\mu\nu} \ , \tag{2.1}$$

from quantum-mechanical processes involving Standard Model fields. Let's begin by heuristically estimating its size by modeling SM fields as a collection of independent harmonic oscillators at each point in space and then summing over their zero-point energies

$$\langle\rho\rangle \sim \int_0^{\Lambda_{\rm UV}} \frac{{\rm d}^3k}{(2\pi)^3}\frac{1}{2}\hbar E_k \sim \int_0^{\Lambda_{\rm UV}} {\rm d}k\ k^2\sqrt{k^2+m^2} \sim \Lambda_{\rm UV}^4 + \Lambda_{\rm UV}^2 m^2 - \frac{m^4}{2}\log\frac{\Lambda_{\rm UV}}{m} + \cdots \ . \tag{2.2}$$

Here $\Lambda_{\rm UV}$ is the cutoff of our theory—the energy scale up to which we can trust predictions. Most conservatively, the Standard Model has been extremely well tested up to energies around electron mass, $\Lambda_{\rm UV} \sim m \sim$ MeV. Plugging in this value, we find a theoretical expectation for the cosmological constant to be around

$$\Lambda_{\rm theory} \sim ({\rm MeV})^4 \sim\ 10^{-88}\ M_{\rm Pl}^4, \tag{2.3}$$

which is extremely far from the observed value (1.3) (off by 32 orders of magnitude). Let's do this calculation a little bit more carefully in order to understand better what is going on.

## 2.1 Contribution from matter fields

Let's calculate carefully how matter fields contribute to the effective cosmological constant. This is an inherently quantum-mechanical question, so we have to employ some quantum field theory formalism. Don't be afraid, the only stuff we will need is basically fancy gaussian integration. Recall the formula for a gaussian integral:

$$\int {\rm d}x\, e^{-\frac{ax^2}{2}} = \left(\frac{2\pi}{a}\right)^{1/2} . \tag{2.4}$$

[1] The form of this contribution may be deduced by noting that on flat space, Lorentz invariance forces $\langle T_{\mu\nu}\rangle \propto \eta_{\mu\nu}$. Then, we invoke the equivalence principle to promote $\eta_{\mu\nu} \mapsto g_{\mu\nu}$.

This formula has an immediate generalization to the matrix gaussian integral ($A$ is an $N \times N$ matrix):

$$\int \mathrm{d}\vec{x}\, e^{-\frac{1}{2}(\vec{x},A\vec{x})} = \frac{(2\pi)^{N/2}}{(\det A)^{1/2}}, \tag{2.5}$$

which is proven by diagonalizing the matrix and then applying (2.4).

The only piece of information we need in addition to this from QFT is that the object

$$Z = \int \mathcal{D}\phi \exp iS[\phi], \tag{2.6}$$

contains the full quantum-mechanical information about a system. This is the Feynman path integral, where $\mathcal{D}\phi$ means that we should integrate over all field configurations, and $S[\phi]$ is the action for some collection of fields, $\phi$.

We will proceed to do part of this integral. For simplicity, we only consider a non-self-interacting heavy scalar $\Phi$—coupled to the fields of interest—so that the action is quadratic in $\Phi$. In this case, the path integral over $\Phi$ can be done exactly, and the effective action for $\phi$ is given by a functional determinant (in essentially the same way as (2.5))

$$\exp iS_{\text{eff}}[\phi] = \int \mathcal{D}\Phi\, \exp iS_{\phi,\Phi} = \det\left(\frac{\delta^2 S_{\phi,\Phi}}{\delta\Phi\delta\Phi}\right)^{-1/2} \exp iS_{\phi,\Phi=0}\,. \tag{2.7}$$

We want to think of this determinant as a contribution to the effective action for the remaining degrees of freedom:

$$\exp i\Delta_\Phi S_{\text{eff}} = \det\left(\frac{\delta^2 S_{\phi,\Phi}}{\delta\Phi\delta\Phi}\right)^{-1/2} = \exp\left(-\frac{1}{2}\mathrm{Tr}\log\frac{\delta^2 S_{\phi,\Phi}}{\delta\Phi\delta\Phi}\right). \tag{2.8}$$

**Integrating out a heavy field coupled to gravity:** We begin by considering the action

$$S = \int \mathrm{d}^4x\sqrt{-g}\left(\frac{M_{\rm Pl}^2 R}{2} - \Lambda - \delta\Lambda + \frac{1}{2}\Phi(\Box - M^2)\Phi\right). \tag{2.9}$$

Here $\delta\Lambda$ is a counter-term which we will see that we need later on. As before, the integral over $\Phi$ is Gaussian, so the contribution to the effective action for $g_{\mu\nu}$ from integrating it out is of the form

$$\Delta_\Phi S_{\text{eff}} = \frac{i}{2}\mathrm{Tr}\log\frac{\delta^2 S}{\delta\Phi\delta\Phi} = \frac{i}{2}\mathrm{Tr}\log\sqrt{-g}(\Box - M^2). \tag{2.10}$$

We can discard the metric determinant piece, as it vanishes in dimensional regularization.[2] What we are trying to compute is the correction to the $\sqrt{-g}$ term which appears in the action (2.9). We

[2]To see this explicitly, note that for an arbitrary differential operator $\hat{\mathcal{O}}$ we have

$$\mathrm{Tr}\log\sqrt{-g}\,\hat{\mathcal{O}} = \mathrm{Tr}\left[\log\sqrt{-g} + \log\hat{\mathcal{O}}\right] = \int \mathrm{d}^4x\,\langle x|\log\sqrt{-g}|x\rangle + \mathrm{Tr}\log\hat{\mathcal{O}} = \int \mathrm{d}^4x\log\sqrt{-g(x)}\,\delta^4(0) + \mathrm{Tr}\log\hat{\mathcal{O}}, \tag{2.11}$$

and the divergence $\delta^4(0)$ is zero in dimensional regularization, so we can consider just the $\mathrm{Tr}\log\hat{\mathcal{O}}$ piece.

will do this perturbatively in $h_{\mu\nu}$ by expanding $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. Note that when we do this, the metric determinant has an infinite number of terms

$$\sqrt{-g} \approx 1 + \frac{1}{2}h + \frac{1}{8}h^2 - \frac{1}{4}h_{\mu\nu}h^{\mu\nu} + \cdots \tag{2.12}$$

so it would seem like we are making more work for ourselves. However, we will exploit diffeomorphism invariance. We will only compute the correction to the $h$ term, diff invariance will fix all the other coefficients in the series in terms of this coefficient.

We are therefore interested in computing

$$\Delta_\Phi S_{\text{eff}} = \frac{i}{2}\text{Tr}\log(\Box - M^2) = \frac{i}{2}\text{Tr}\log(\partial^2 - M^2 - h^{\mu\nu}\partial_\mu\partial_\nu - \partial^\nu h_\nu{}^\mu\partial_\mu + \frac{1}{2}\partial^\alpha h\,\partial_\alpha + \ldots)\,. \tag{2.13}$$

We can pull out a factor of the scalar propagator (this is basically the vacuum bubble contribution, so we will neglect it). Beginning at linear order in $h_{\mu\nu}$, we need to calculate

$$\Delta_\Phi S_{\text{eff}} = \frac{i}{2}\text{Tr}\log\left[\mathbb{1} + \frac{1}{\partial^2 - M^2}\left(-h^{\mu\nu}\partial_\mu\partial_\nu - \partial^\nu h_\nu{}^\mu\partial_\mu + \frac{1}{2}\partial^\alpha h\,\partial_\alpha\right) + \ldots\right] \tag{2.14}$$

$$= \frac{i}{2}\text{Tr}\left[\frac{1}{\partial^2 - M^2}\left(-h^{\mu\nu}\partial_\mu\partial_\nu - \partial^\nu h_\nu{}^\mu\partial_\mu + \frac{1}{2}\partial^\alpha h\,\partial_\alpha\right)\right] + \cdots\,, \tag{2.15}$$

where we have used $\log(1+\alpha) = \alpha - \frac{\alpha^2}{2} + \frac{\alpha^3}{3} + \cdots$. The only term that contributes to the $h$ term in the action is

$$\Delta_\Phi S_{\text{eff}} \supset -\frac{i}{2}\text{Tr}\left[\frac{1}{\partial^2 - M^2}h^{\mu\nu}\partial_\mu\partial_\nu\right]\,, \tag{2.16}$$

so we just need to compute this trace. To do this, we make a quantum-mechanics analogy. We promote the coordinates and derivatives to operators acting on a Hilbert space as

$$x^\mu \mapsto \hat{x}^\mu\,, \qquad \partial_\mu \mapsto i\hat{p}_\mu. \tag{2.17}$$

We then introduce sets of position eigenstates, $|x\rangle$ and momentum eigenstates $|p\rangle$, which satisfy the orthogonality and completeness relations

$$\langle x|y\rangle = \delta(x-y) \qquad \langle p|k\rangle = \delta(p-k) \tag{2.18}$$

$$\int \mathrm{d}^d x\,|x\rangle\langle x| = \mathbb{1} \qquad \int \mathrm{d}^d p\,|p\rangle\langle p| = \mathbb{1}. \tag{2.19}$$

The inner product between these two bases is given by[3]

$$\langle x|p\rangle = \frac{1}{(2\pi)^{d/2}}e^{ip\cdot x}. \tag{2.21}$$

[3]We employ the following convention for Fourier transformation

$$f(x) = \int \frac{\mathrm{d}^d k}{(2\pi)^d}\,e^{ik\cdot x}\tilde{f}(k) \qquad \tilde{f}(k) = \int \mathrm{d}^d x\,e^{-ik\cdot x}f(x). \tag{2.20}$$

The action of the $\hat{x}^\mu$ and $\hat{p}_\mu$ operators on these eigenstates is the obvious one

$$\hat{x}^\mu|x\rangle = x^\mu|x\rangle \tag{2.22}$$
$$\hat{p}_\mu|p\rangle = p_\mu|p\rangle. \tag{2.23}$$

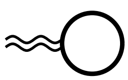
The operator $h^{\mu\nu}\partial_\mu\partial_\nu$ then gets mapped to

$$h^{\mu\nu}\partial_\mu\partial_\nu \mapsto h^{\mu\nu}(\hat{x})(i\hat{p})_\mu(i\hat{p})_\nu. \tag{2.24}$$

We can then evaluate the matrix element of this operator between momentum eigenstates by inserting a complete set of position eigenstates

$$\begin{aligned}\langle k|h^{\mu\nu}(\hat{x})(i\hat{p})_\mu(i\hat{p})_\nu|p\rangle &= \int \mathrm{d}^d x\, \langle k|h^{\mu\nu}(\hat{x})|x\rangle\langle x|(i\hat{p})_\mu(i\hat{p})_\nu|p\rangle = \int \mathrm{d}^d x\, h^{\mu\nu}(x) i p_\mu i p_\nu \langle k|x\rangle\langle x|p\rangle \\ &= \int \frac{\mathrm{d}^d x}{(2\pi)^d}\, e^{-i(k-p)\cdot x} h^{\mu\nu}(x) i p_\mu i p_\nu = (2\pi)^{-d}\tilde{h}(k-p) i p_\mu i p_\nu,\end{aligned} \tag{2.25}$$

Using this, we find (note that in the trace, the object $\tilde{h}(0)$ appears, we interpret this as the $\int \mathrm{d}^d x h^{\mu\nu}(x)$)



$$= -\frac{i}{2}\mu^{4-d}\int \mathrm{d}^d p\, \langle p|\frac{1}{\partial^2 - M^2} h^{\mu\nu}\partial_\mu\partial_\nu|p\rangle = -\frac{i}{2}\mu^{4-d}\int \mathrm{d}^d x\, h^{\mu\nu}(x)\int \mathrm{d}^d p\, \frac{p_\mu p_\nu}{p^2 + M^2}. \tag{2.26}$$

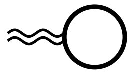
Note that this is exactly the result we would have gotten from a traditional Feynman diagram computation, but here we haven't had to worry about Feynman rules or anything, the trace (2.10) is the only thing we have to worry about. Now we just have to to the integral in (2.26). First, note that we can replace $p_\mu p_\nu \mapsto \frac{1}{d}p^2\eta_{\mu\nu}$ because the non-trace piece will vanish by symmetry (integral of an odd function over a symmetric domain) so we are left with computing (we first send $p_0 \mapsto i p_d$, to yield a Euclidean integral)

$$\frac{1}{d}\int \mathrm{d}^d p\, \frac{p^2}{p^2 + M^2} = \frac{i}{d}\int \frac{\mathrm{d}\Omega_{d-1}\mathrm{d}p}{(2\pi)^d}\frac{p^{d+1}}{p^2 + M^2} = \frac{i}{d(4\pi)^{d/2}} M^d \frac{\Gamma\left(1 + \frac{d}{2}\right)\Gamma\left(-\frac{d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}, \tag{2.27}$$

where $\Gamma$ is the Euler gamma function. The reason that we had to do this stuff in $d$-dimensions is that the integral we wanted to do is divergent in $d = 4$, as can be seen by writing $d = 4 - \epsilon$ and taking $\epsilon \to 0$ in (2.27)

$$\frac{i}{d(4\pi)^{d/2}} M^d \frac{\Gamma\left(1 + \frac{d}{2}\right)\Gamma\left(-\frac{d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \xrightarrow{d\to 4} \frac{iM^4}{32\pi\epsilon} - \frac{iM^4}{64\pi^2}\left(\log\frac{M^2}{\mu^2} - \log 4\pi + \gamma - \frac{3}{2}\right) + \cdots \tag{2.28}$$

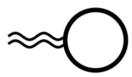
This expression has a pole at $\epsilon = 0$. What we have just done is an example of dimensional regularization. Note also that the scale $\mu$ that we introduced is completely unphysical, we merely had to introduce it in order to make all the dimensions correct, so we are free to redefine it to absorb the finite pieces $-\log 4\pi + \gamma - \frac{3}{2}$. Note that we *cannot* change the coefficient of the logarithmic term (though we can change the argument). This is a recurring theme in QFT—the coefficients of logs are physical things. Putting this all together, we find a contribution to the effective action of the form



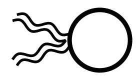
$$= \int \mathrm{d}^4 x \, \frac{1}{2} h \, \frac{M^4}{64\pi^2} \left( \frac{2}{\epsilon} - \log \frac{M^2}{\mu^2} \right). \tag{2.29}$$

As noted above, we should think of this as the first term in an infinite series which will re-sum to give a contribution to the cosmological constant

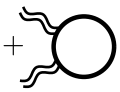
$$\sqrt{-g} \ \sim$$



$$+$$



$$+$$



$$+ \cdots \tag{2.30}$$

which will be of the form

$$S_{\text{eff}} \supset \int \mathrm{d}^4 x \, \sqrt{-g} \left[ -\Lambda - \delta\Lambda + \frac{M^4}{64\pi^2} \left( \frac{2}{\epsilon} - \log \frac{M^2}{\mu^2} \right) \right]. \tag{2.31}$$

Now for the magic of renormalization: none of these terms individually are physical, it is only really their sum that can be measured. Therefore, we are free to choose $\delta\Lambda$ to cancel the (infinite) contribution $\frac{M^4}{64\pi^2} \frac{2}{\epsilon}$, which leaves a "renormalized" value for the CC of

$$\Lambda_R = \Lambda + \frac{M^4}{64\pi^2} \log \frac{M^2}{\mu^2}. \tag{2.32}$$

This value is still not really physical, as it still depends on the unphysical scale $\mu$, and can also be freely adjusted by a finite amount by changing $\delta\Lambda$, but this is fine; it is no different from anything else in QFT, we go out and *measure* $\Lambda_R$ (we can think of doing the measurement at the scale $\mu$ and this fixes $\Lambda$ at that scale, then they both change in a compensatory way as we change $\mu$).

The thing I want to focus on is that the contribution from $\Phi$ to the CC scales as $\sim M^4$. This means that if we imagine that we compute $\Lambda_R$, go out and measure it, and then imagine that we can shift $M \mapsto M + \Delta M$, then $\Lambda_R$ will be power-law sensitive to this change. Alternatively, we could imagine holding everything fixed and introducing another heavy particle into the theory, the value of $\Lambda$ would also be power-law sensitive to this introduction. This is the essence of the cosmological constant problem—the renormalized value of the CC is extremely sensitive to the masses of *other* particles in the theory.[4] The Higgs hierarchy problem in the Standard Model is essentially the same.

## 2.2 A foil: fermion masses

Here I want to illustrate that the answer that we got in the CC case is *not* how things always go. To see this, consider a toy example

$$\mathcal{L} = \frac{1}{2}\Phi(\Box - M^2)\Phi + \bar{\psi}(i\partial\!\!\!/ - m)\psi + \lambda\Phi\bar{\psi}\psi, \tag{2.33}$$

[4] Often, people will phrase the CC problem as having to tune $\Lambda_R$ at 1-loop and then having to change the tuning at 2-loops and so on. I think that this characterization misses the point. Above we computed an example which is 1-loop exact, and there is still a CC problem. The real essence of the problem is that the value of the CC is extremely sensitive to *deformations* of the theory by changing the masses of the heavy states holding regularization and renormalization schemes fixed.

which consists of a scalar field coupled to a Dirac fermion. Similar to what we did with the CC we can integrate out the scalar and see how it corrects the mass term for the fermion. In this case the functional method we used above is a little clumsy, so you will just have to believe me that the mass correction goes like

$$\delta m \sim$$

$$\sim \lambda^2 \int \frac{\mathrm{d}^d p}{(2\pi)^d} \frac{\mathrm{Tr}(\not{p} + \not{k} - m)}{((p-k)^2 + m^2)(k^2 + M^2)} \sim \frac{\lambda^2}{16\pi^2} m \log \frac{m}{\mu}. \tag{2.34}$$

The precise value of the correction is not important, what is important is that $\delta m \propto m$—the mass of the fermion is renormalized proportional to itself. This is known as *multiplicative* renormalization (as opposed to the *additive* renormalization the CC received above). It is for this reason that we do not worry that fermion masses are small in the Standard Model. Indeed, in weak scale units, the electron mass is very small ($\sim 10^{-7}$TeV).

There is a symmetry reason for this result, in the limit $m \to 0$, the lagrangian (2.33) has a chiral symmetry

$$\psi \mapsto \gamma_5 \psi, \qquad\qquad \Phi \mapsto -\Phi, \tag{2.35}$$

which is broken by the presence of the mass term. (Under the symmetry $\bar{\psi}\psi \mapsto -\bar{\psi}\psi$). In this limit, any quantum corrections to the mass term must vanish, because they violate the symmetry, so the corrections that we do get away from this point must be proportional to $m$ so that they vanish in the limit. This symmetry expectation is borne out by the explicit computation (2.34).

If a parameter in a theory leads to an enhanced symmetry in the limit it is taken to be zero, it is said to be *technically natural* for the parameter to be small.[5] The fermion mass considered here is one example of a technically natural parameter (owing to the chiral symmetry). No symmetry is known that would make the CC technically natural.

# 3 Some solutions that don't work

Now that we understand what the problem is, let's try to solve it.

## 3.1 Supersymmetry

Supersymmetry is an extension of the symmetry algebra of spacetime to include fermionic generators (supercharges). In the simplest scenario, the anticommutator of these supercharges is related to the momentum generator of the Poincaré algebra:

$$\{Q_\alpha, \bar{Q}_{\dot{\beta}}\} = \sigma^\mu_{\alpha\dot{\beta}} P_\mu. \tag{3.1}$$

Imagine now that the vacuum which describes our world is supersymmetric, *i.e.*, $Q_\alpha|0\rangle = \bar{Q}_{\dot{\alpha}}|0\rangle = 0$. Then, we can take the inner product of the state $Q_\alpha|0\rangle$ with itself (and do the same for $Q_{\dot{\alpha}}|0\rangle$ and add them) ,

$$\langle 0|Q_\alpha \bar{Q}_{\dot{\beta}} + \bar{Q}_{\dot{\beta}} Q_\alpha|0\rangle = \sigma^\mu_{\alpha\dot{\beta}} \langle 0|P_\mu|0\rangle = 0 \implies \langle 0|P_\mu|0\rangle = 0, \tag{3.2}$$

[5]This nomenclature is due to 't Hooft.

the last equality follows from $\sigma^\mu$ being nonzero. In particular, this implies that $\langle 0|P_0|0\rangle = \langle 0|H|0\rangle = 0$, which means that the CC has to vanish in this vacuum. So, if the world were supersymmetric, the CC would be zero. We could then imagine that SUSY is broken at a scale $\sim$ meV, leading to a residual CC around this magnitude. However, we are pretty sure that the world is not supersymmetric at meV scales, so this proposed solution doesn't work.

The real situation is slightly more complicated than this because we should also make gravity dynamical and ask this question in the context of supergravity, but this additional complication does not change the story very much, see [4].

### 3.2 Weyl invariance

In the same vein as SUSY, we could recognize that the Weyl transformation

$$g_{\mu\nu} \mapsto \Omega^2 g_{\mu\nu}, \tag{3.3}$$

causes the determinant to transform as

$$\sqrt{-g} \mapsto \Omega^4 \sqrt{-g}. \tag{3.4}$$

So if gravity were Weyl invariant, a cosmological constant would not be allowed. However, this idea doesn't really get off the ground, the Einstein–Hilbert term itself isn't Weyl invariant either, so we would need to consider some completely different theory of gravity. One proposal is Weyl gravity

$$S = M_{\rm Pl}^2 \int \mathrm{d}^4x \sqrt{-g}\, C_{\mu\nu\rho\sigma} C^{\mu\nu\rho\sigma}, \tag{3.5}$$

where $C_{\mu\nu\rho\sigma}$ is the traceless part of the Riemann tensor (the Weyl tensor), which *is* Weyl invariant. However, this theory has a ghost. One might also imagine introducing a scalar field as in

$$S = \int \mathrm{d}^4x \sqrt{-g} \left( -(\partial\phi)^2 - \frac{1}{6}\phi^2 R \right), \tag{3.6}$$

which is Weyl invariant where the metric transforms as (3.3) and where the scalar transforms as $\phi \mapsto \Omega^{-1}\phi$. However, in this case, the term $\sqrt{-g}\,\phi^4$ is also Weyl invariant, which is just the CC upon fixing a gauge for $\phi$. The fact that this doesn't work shouldn't really surprise us, $\phi$ is just a Stückelberg field for the Weyl invariance, and introducing gauge symmetries can't change the physics.

### 3.3 Unimodular gravity

Another attempt to solve the CC problem is *unimodular gravity*. In this theory, one considers the traceless Einstein equations

$$R_{\mu\nu} - \frac{1}{4} R g_{\mu\nu} = \frac{1}{M_{\rm Pl}^2} \left( T_{\mu\nu} - \frac{1}{4} T g_{\mu\nu} \right) . \tag{3.7}$$

Taking a covariant divergence and utilizing the Bianchi identity, one finds the following

$$\nabla_\mu R = -\nabla_\mu T \ , \tag{3.8}$$

which upon integrating yields

$$T = -R + \Lambda \ , \tag{3.9}$$

where $\Lambda$ is an integration constant. Plugging this expression back into (3.7) yields precisely the full Einstein equations, but where $\Lambda$ is now an integration constant, unrelated to any bare CC in the action.

The idea is then that one chooses the integration constant to reproduce the observed value of the CC. However, this falls short of being a true solution to the cosmological constant problem. The reason is somewhat subtle, it is basically because we are restricting the solutions to Einstein's equations by imposing some boundary conditions on the fields—essentially fixing the Ricci scalar at infinity. This exact same thing could be done in Einstein gravity, it's not clear that we have gained anything. Indeed, we still have some somewhat arbitrarily fix $\Lambda$ to match the observed value.

This theory is called unimodular because the equations (3.7) can be derived from the Einstein–Hilbert action from a variational principle that restricts the determinant to be -1. Another question surrounding unimodular gravity is that it is not clear to what degree it and Einstein gravity are equivalent at the quantum-mechanical level.

### 3.4 Letting $\Lambda$ be a dynamical field (Weinberg's no-go theorem)

Here we consider an alluring idea—that the cosmological constant may be able to dynamically relax to a small value—and its obstructions. Most prominent among these is a celebrated *no-go* result of Weinberg [4]. In essence, this result says that we cannot achieve anything by tying the value of the CC to the potential of some scalar field—we must fine-tune the potential just as much as we would have had to fine-tune the bare CC.

Here we follow Weinberg's original argument [4], and consider a set-up in which we have a scalar field coupled to gravity in any way we like[6]

$$S = \frac{M_{\rm Pl}^2}{2}\int {\rm d}^4x\sqrt{-g}R + S[\phi, g_{\mu\nu}] \ , \tag{3.10}$$

where $S[\phi, g_{\mu\nu}]$ can depend arbitrarily on $\phi, g_{\mu\nu}$ and their derivatives. We look for a solution where

$$\phi = \bar{\phi} = \text{constant} \tag{3.11}$$

$$g_{\mu\nu} = \eta_{\mu\nu} \ . \tag{3.12}$$

[6] Actually, the theorem proved in [4] is even more general, allowing the for $N$ fields, which can be either scalars or tensors, but here we will restrict to a single scalar for simplicity.

With such an ansatz, the Euler–Lagrange equations become very simple

$$\left.\frac{\delta \mathcal{L}}{\delta \phi}\right|_{g_{\mu\nu};\phi=\text{const.}} = \frac{\partial \mathcal{L}}{\partial \phi} = 0 \tag{3.13}$$

$$\left.\frac{\delta \mathcal{L}}{\delta g^{\mu\nu}}\right|_{g_{\mu\nu};\phi=\text{const.}} = \frac{\partial \mathcal{L}}{\partial g^{\mu\nu}} = 0 \ . \tag{3.14}$$

In order for a solution to (3.13) and (3.14) to be natural, we want the trace of the gravitational equation of motion to be satisfied automatically as a consequence of the scalar equations (this is basically because $\phi$ is the source for $T$). Another way of saying this is that the trace of the metric equation of motion must be of the form:

$$g^{\mu\nu}\frac{\partial \mathcal{L}}{\partial g^{\mu\nu}} = f(\phi)\frac{\partial \mathcal{L}}{\partial \phi}. \tag{3.15}$$

Demanding this equation be satisfied is equivalent to demanding a particular symmetry of the Lagrangian [4]. To see this, note that the variation of the action (3.10) is

$$\delta S = \int \mathrm{d}^4x \left(\frac{\delta \mathcal{L}}{\delta g^{\mu\nu}}\delta g^{\mu\nu} + \frac{\delta \mathcal{L}}{\delta \phi}\delta\phi\right) \ . \tag{3.16}$$

If we consider the variations

$$\delta g^{\mu\nu} = \epsilon g^{\mu\nu} \ ; \qquad \delta\phi = -\epsilon f(\phi) \ , \tag{3.17}$$

then (3.15) implies that the action is invariant under this symmetry when the fields are taken to be constant. If we start with a Lagrangian invariant under (3.17), if it admits a solution $\bar{\phi} = \text{const.}$ with $\frac{\partial \mathcal{L}}{\partial \phi}\big|_{\phi=\bar{\phi}} = 0$, then the gravitational equation will be satisfied. However, this turns out to be impossible to arrange without some degree of fine tuning.

To see this, consider doing a field redefinition of $\phi$ into $\psi$ so that the symmetry transformation (3.17) is [4]

$$\delta g^{\mu\nu} = 2\epsilon g^{\mu\nu}, \qquad \delta\psi = -\epsilon \ . \tag{3.18}$$

This transformation is now nothing but a conformal transformation, with $\psi$ playing the role of a dilaton. This means that when the fields are constant, the Lagrangian can be written as a function of the conformal metric

$$\hat{g}_{\mu\nu} = e^{2\psi} g_{\mu\nu} \ . \tag{3.19}$$

When all the fields are set to constants, all the curvature invariants of this metric vanish, so the on-shell Lagrangian must be of the form

$$\mathcal{L} = \sqrt{-\hat{g}}\mathcal{L}(\sigma) = \sqrt{-g}e^{4\psi}\mathcal{L}(\sigma), \tag{3.20}$$

where $\mathcal{L}(\sigma)$ is meant to stand for other possible fields in the theory. However, the equation $\frac{\partial \mathcal{L}}{\partial g^{\mu\nu}}\big|_{g_{\mu\nu}=\text{const.}} = 0$ implies that we must have

$$\left.\sqrt{-g}e^{4\psi}\mathcal{L}(\sigma)\right|_{g_{\mu\nu};\psi;\sigma=\text{const.}} = e^{4\psi}V(\bar{\sigma}) = 0 \ , \tag{3.21}$$

which is clearly a fine-tuning (we are tuning the potential for $\sigma$ to have a minimum at $V(\bar{\sigma}) = 0$).

It is worthwhile to examine the assumptions which went into this no-go theorem. First, we assumed that there were a finite number of scalar fields—it is possible that the conclusions could be avoided with an infinite number of fields, but to date this loophole has not been exploited. More promising would be to give up the assumption of constant fields; indeed, some people have tried to exploit this loophole in scenarios where the scalar sector has non-trivial coordinate dependence.

# 4 Some ideas that might work

Let's move on to some more promising avenues.

## 4.1 Anthropics and the landscape

A popular approach to the cosmological constant problem is that of *anthropics*. This approach has its origins in an old argument of Weinberg [7] that the cosmological constant in our universe cannot be very large or else cosmological large scale structures would not have been able to form, and consequently we would not be around to ask questions about the cosmological constant. Interestingly, Weinberg predicted that we should observe a cosmological constant somewhere near this upper bound *prior* to the observational discovery of the accelerating universe. It is then a genuine prediction.

Before trying to understand the anthropic approach to the CC problem, it is helpful to make a few anthropic reasonings about the structure of planets in the solar system as an orientation. Consider the following two facts

1. The Earth is at a distance from the Sun that the temperature varies between $\sim 0°\text{C}$ and $\sim 30°\text{C}$.
2. The Moon is at a distance from the Sun that it subtends the same angle as the Sun, allowing for solar eclipses.

Both of these appear to be fine-tunings; however, only one of them is anthropic. As you well know, if the Earth were at any other distance from the Sun, things would be either too hot or too cold for human life to develop. Therefore, it is not surprising that we find ourselves this distance from the Sun. In order for this to be predictive there must be an ensemble of planets at varying distances from stars and some mechanism for populating this ensemble. For the Earth-Sun case, this mechanism is precisely standard planetary formation. Conversely, the fact that solar eclipses can happen is a pure accident, it is just a random fine tuning.

Proponents of the anthropic principle argue that the value of the cosmological constant is like the Earth-Sun distance. In order for large scales structures to form, we must have that $\Omega_\Lambda(z_\text{g}) \lesssim \Omega_m(z_\text{g})$, where $z_\text{g}$ is the redshift where the first galaxies form. This then implies that

$$\frac{\Omega_\Lambda(0)}{\Omega_m(0)} \leq (1 + z_\text{g})^3 \sim 10^3. \tag{4.1}$$

This puts an anthropic upper bound on the magnitude of the CC to be roughly 3 orders of magnitude larger than the observed value. If the CC were negative, it would still be possible to form galaxies, but spacetime would collapse into a singularity on a timescale of order $\Lambda^{-1}$, this puts a lower bound that the CC must be around the matter density today

$$\frac{\Omega_\Lambda(0)}{\Omega_m(0)} \gtrsim -10. \tag{4.2}$$

Weinberg argued that the CC should be observed to be near one of the endpoints of this range, as there was no anthropic reason for it to be smaller. The observed value differs from this by a couple order of magnitude, but this amount of tuning is already much better than the 120 orders of magnitude fine tuning needed above.

At this level, the anthropic argument is tautological; we observe the parameter value we do because we wouldn't be around to observe it otherwise. In order for this argument to be explanatory or predictive, it needs the second aspect that our Earth-Sun example had: there needs to be an ensemble of vacua with different values of the cosmological constant, and some mechanism to populate them. We would then like for the value we observe to be statistically likely in this ensemble. The idea is that the ensemble is provided by the string theory landscape and the mechanism for populating the vacua is eternal inflation.

The current understanding of string theory is that is possesses an extremely large number of solutions (anywhere from $\sim 10^{500}$ to exponentially more) with a wide variety of values of the cosmological constant. The idea is that if eternal inflation took place, Coleman–de Luccia tunneling would populate all of these vacua, leading to different values of the cosmological constant in different regions of the multiverse. The hope is then that we can argue that our observed value of the CC is statistically likely (or at the very least not *un*likely). Even better would be if there were some other property which was correlated with having a small value of $\Lambda$, which we could then go out at measure or observe.

To me, this seems like a promising thing to pursue, but it there are several open problems which must be resolved in order for this to a compelling solution:

- It must be demonstrated that there exist de Sitter solutions in string theory with a variety of values of the cosmological constant (further, these must be much more closely spaced than $10^{-120}M_{\rm Pl}^4$). To date there are *no* fully compelling de Sitter solutions in string theory.
- It must be demonstrated that eternal inflation occurs; this is suspected, but it has not been rigorously established that it is a necessary consequence of inflation (which itself is not guaranteed to have happened).
- It must be established that the combination of eternal inflation and vacuum tunneling populates all of the putative de Sitter vacua.
- Finally, it must be shown that vacua like ours are statistically likely amongst the landscape, or at least are more likely to be populated. This step has proven to be difficult as well—typically many quantities are infinite in eternal inflation, and this requires regulating these infinities in some way. This is known as the measure problem.

### 4.2 Modifying gravity in the infrared

Another approach to the cosmological constant involves changing the rules of gravity—maybe Einstein gravity is not the correct theory of gravity at the longest wavelengths. Before considering changing gravity, let's first explore the assumptions needed to arrive at Einstein gravity in the first place.

#### 4.2.1 Einstein gravity is massless spin-2

It turns out that under the following assumptions:

- Lorentz invariance, locality
- massless spin-2 field, interacting
- low energy (meaning long distances or lowest order in derivatives)

we are led uniquely to Einstein gravity. Let's sketch the argument. Our starting point is the action for a free massless spin-2 field at lowest order in derivatives[7]

$$S = -\frac{1}{2}\int \mathrm{d}^4x\; h^{\mu\nu}\left(\Box h_{\mu\nu} - \partial_\mu\partial^\alpha h_{\alpha\nu} - \partial_\nu\partial^\alpha h_{\alpha\mu} + \partial_\mu\partial_\nu h + \eta_{\mu\nu}\left(\partial^\alpha\partial^\beta h_{\alpha\beta} - \Box h\right)\right) , \tag{4.4}$$

where the field $h_{\mu\nu}$ has dimensions of mass. The action is invariant under the gauge transformation

$$\delta_\xi h_{\mu\nu} = \partial_\mu \xi_\nu + \partial_\nu \xi_\mu \ , \tag{4.5}$$

with gauge parameter $\xi_\mu = \xi_\mu(x)$. These are nothing more than linearized diffeomorphisms. The equation of motion following from this action is

$$\Box h_{\mu\nu} - \partial_\mu\partial^\alpha h_{\alpha\nu} - \partial_\nu\partial^\alpha h_{\alpha\mu} + \partial_\mu\partial_\nu h + \eta_{\mu\nu}\left(\partial^\alpha\partial^\beta h_{\alpha\beta} - \Box h\right) = 0 \ . \tag{4.6}$$

As a theory on its own, (4.4) is perfectly fine, but the field $h^{\mu\nu}$ is free; let's see what happens when we try to introduce interactions by coupling it to its own energy momentum tensor (note that $h_{\mu\nu}$ should couple to the full stress tensor, including its own). The action is schematically

$$S \sim \int \mathrm{d}^4x \left(h^{\mu\nu}\mathcal{E}^{\alpha\beta}_{\mu\nu}h_{\alpha\beta} + h^{\mu\nu}S^{(2)}_{\mu\nu}\right) \ , \tag{4.7}$$

where $\mathcal{E}^{\alpha\beta}_{\mu\nu}$ is the Lichnerowicz operator and $S^{(2)}_{\mu\nu}$ is some tensor quadratic in the field $h$, chosen such that

$$\frac{\delta}{\delta h^{\mu\nu}}\left(h^{\alpha\beta}S^{(2)}_{\alpha\beta}\right) = \Theta^{(2)}_{\mu\nu} \ , \tag{4.8}$$

[7] This structure is imposed upon us by demanding that our Lagrangian be manifestly Lorentz invariant, local and that it describes the two polarizations of a massless spin-2 particle. The field operator, $h_{\mu\nu}$, is *not* a tensor under Lorentz transformations, rather it transforms inhomogeneously [8]

$$U(\Lambda)h_{\mu\nu}U^{-1}(\Lambda) = \Lambda^\alpha{}_\mu\Lambda^\beta{}_\nu h_{\alpha\beta}(\Lambda^{-1}x) + \partial_\mu\xi_\nu(x,\Lambda) + \partial_\nu\xi_\mu(x,\Lambda) \ , \tag{4.3}$$

where the explicit form of $\xi_\mu$ is not important. Therefore, in order to have a Lagrangian which propagates the desired degrees of freedom, we must construct it so that it both *looks* Lorentz invariant and is invariant under the additional transformation $\delta h_{\mu\nu} = \partial_\mu\xi_\nu + \partial_\nu\xi_\mu$, which leads us uniquely to (4.4).

where $\Theta^{(2)}_{\mu\nu}$ is the energy momentum tensor of the quadratic action. The stress tensor can be constructed from the standard Noether procedure. The equations of motion following from this action are

$$\mathcal{E}^{\alpha\beta}_{\mu\nu} h_{\alpha\beta} \sim \Theta^{(2)}_{\mu\nu} \,. \tag{4.9}$$

Now, the left hand side is identically divergence-less, but the right hand side is not conserved, because $\Theta^{(2)}_{\mu\nu}$ is *not* the full energy-momentum tensor for the field $h$—the cubic piece we added to the action also contributes! However, we may correct for this by adding a quartic piece to the action

$$S \sim \int \mathrm{d}^4x \left( h^{\mu\nu} \mathcal{E}^{\alpha\beta}_{\mu\nu} h_{\alpha\beta} + h^{\mu\nu} S^{(2)}_{\mu\nu} + h^{\mu\nu} S^{(3)}_{\mu\nu} \right) \,, \tag{4.10}$$

so that the equation of motion is

$$\mathcal{E}^{\alpha\beta}_{\mu\nu} h_{\alpha\beta} \sim \Theta^{(2)}_{\mu\nu} + \Theta^{(3)}_{\mu\nu} \,, \tag{4.11}$$

where $\Theta^{(3)}_{\mu\nu}$ is stress tensor of the cubic part of the action. However, this still doesn't fully fix the problem, because now the quartic piece we added contributes to the stress tensor. If we continue to iterate the procedure, we will end up with an infinite number of terms in the action, and the claim is that these re-sum to give Einstein gravity

$$S_{\mathrm{EH}} \sim \int \mathrm{d}^4x \left( h\mathcal{E}h + h \sum_{n=2}^{\infty} S^{(n)} \right) \sim \int \mathrm{d}^4x \sqrt{-g} R \,. \tag{4.12}$$

This iteration procedure was performed by a shortcut in [9]. There is also an equivalent, but algebraically simpler, derivation of Deser [10].

### 4.2.2 Generalities

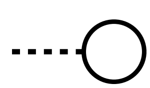
We saw above that relatively few assumptions lead us more or less directly to GR. In order to move beyond Einstein gravity, we must therefore consider breaking one of these assumptions. We like Lorentz invariance and locality, so we would like to keep these. The basic assumption we will consider breaking is the (kind of implicit) assumption that the massless spin-2 field is the only gravitational degree of freedom. Basically we will consider adding more degrees of freedom. For a review of this approach, see [11].

First, to neutralize $\Lambda$ to an accuracy of order $\sim H_0^2 M_{\mathrm{Pl}}^2 \sim (\mathrm{meV})^4$, the scalars must have a mass at most comparable to the present-day Hubble parameter,

$$m_\phi \lesssim H_0 \,. \tag{4.13}$$

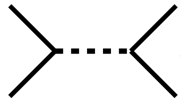
If they were much more massive, they could be integrated out and would be irrelevant to the low energy dynamics.

Secondly, these scalars must couple to Standard Model fields since, as we saw in Section 2.1, SM fields contribute a large amount to the vacuum energy. Another way of saying this is that the tadpole diagram



(4.14)

must be present in the theory. However, by unitarity, so must the exchange diagram

 . (4.15)

This implies that $\phi$ mediates a force between Standard Model fields, with a range around $\sim m_\phi^{-1}$. Given (4.13), this is comparable to the present Hubble radius. Thus, the scalar mediates a fifth force, both at cosmological distances and within the solar system. However, gravity is exquisitely well tested within the solar system so there must necessarily be some mechanism which hides these new fields from local observations. This can be achieved through *screening mechanisms*, which suppress deviations from GR.

**Screening:** Let's try to understand how screening can work in the context of a single scalar field conformally coupled to matter

$$\mathcal{L} = -\frac{1}{2} Z^{\mu\nu}(\phi, \partial\phi, \ldots)\partial_\mu\phi\partial_\nu\phi - V(\phi) + g(\phi)T^\mu_{\ \mu} \,, \tag{4.16}$$

where $Z^{\mu\nu}$ schematically encodes derivative self-interactions of the field, and $T^\mu_{\ \mu}$ is the trace of the matter stress-energy tensor. For non-relativistic sources, we can make the replacement $T^\mu_{\ \mu} \to -\rho$. In the presence of a point source, $\rho = \mathcal{M}\delta^3(\vec{x})$, we can then expand the field about its background solution $\bar{\phi}$ as $\phi = \bar{\phi} + \varphi$ to obtain the equation of motion for the perturbation:

$$Z(\bar{\phi})\Big(\ddot{\varphi} - c_s^2(\bar{\phi})\nabla^2\varphi\Big) + m^2(\bar{\phi})\varphi = g(\bar{\phi})\mathcal{M}\delta^3(\vec{x}) \,, \tag{4.17}$$

where $c_s$ is an effective sound speed. In general, we have in mind that the background value $\bar{\phi}$ is set by other background quantities, such the local density $\bar{\rho}$ or the Newtonian potential $\Phi$. Neglecting the spatial variation of $\bar{\phi}$ over the scales of interest, the resulting static potential is

$$V(r) = -\frac{g^2(\bar{\phi})}{Z(\bar{\phi})c_s^2(\bar{\phi})} \frac{e^{-\frac{m(\bar{\phi})}{\sqrt{Z(\bar{\phi})}c_s(\bar{\phi})}r}}{4\pi r}\mathcal{M} \,. \tag{4.18}$$

The corresponding force is therefore *attractive*, as it should be for scalar mediation.

For a light scalar, and with the other parameters $\mathcal{O}(1)$, we see that $\varphi$ mediates a gravitational-strength long range force $F_\varphi \sim 1/r^2$. Local tests of GR forbid any such force to high precision. However, the fact that the various parameters $g$, $Z$, $c_s$ and $m$ appearing in (4.18) depend on the background value of the field gives us some ways to suppress this force.

- *Weak coupling:* One possibility is to let the coupling to matter, $g$, depend on the environment. In regions of high density—where local tests of gravity are performed—the coupling is very small, and the fifth force sufficiently weak to satisfy all of the constraints.
- *Large mass:* Another option is to let the mass of fluctuations, $m(\bar{\phi})$, depend on the ambient matter density. In regions of high density, such as on Earth, the field acquires a large mass, making its effects short range and hence unobservable. Deep in space, where the mass density is low, the scalar is light and mediates a fifth force of gravitational strength. This idea leads quite naturally to screening of the *chameleon* type.

- *Large inertia:* We may also imagine making the kinetic function, $Z(\bar{\phi})$, large environmentally. This leads us to screening of the *kinetic* type.

**Dark Energy vs. Modified Gravity:** Often with modifications of this type, people assume that some other unknown physics (or symmetry principle) resolves the "old" cosmological constant problem (the contribution from SM fields), and then seek to explain the observed cosmological acceleration through some new stress energy component (dark energy) or additional degrees of freedom (modified gravity). These two approaches are notoriously intertwined—consider adding some new term, $\mathcal{G}_{\mu\nu}$, to Einstein's equations which is built out of the metric and derivatives in some nontrivial way

$$G_{\mu\nu} + \mathcal{G}_{\mu\nu} = \frac{1}{M_{\rm Pl}^2} T_{\mu\nu}. \tag{4.19}$$

We would then call this modified gravity, however, nothing stops us from putting $\mathcal{G}_{\mu\nu}$ on the right hand side and considering it as a new stress energy component and thinking of it as dark energy.

Therefore we see that the boundary between these two approaches is somewhat fuzzy. In [12], we proposed that a boundary can be drawn such that anything which obeys the strong equivalence principle is dark energy while things which obey the weak equivalence principle but not the strong should be thought of as modified gravity.

### 4.2.3 $f(R)$

A modification to gravity, which can exhibit chameleon screening, that has received a lot of attention is so-called $f(R)$ gravity; in these models, the Ricci scalar in the Einstein–Hilbert action is replaced by an arbitrary function of the Ricci scalar. In [13–15], this idea was adapted to explain the late-time acceleration of the universe, without invoking a cosmological constant.

The action for this modification to Einstein gravity is of the form

$$S = \frac{M_{\rm Pl}^2}{2} \int {\rm d}^4x \sqrt{-g} \Big( R + f(R) \Big) + S_{\rm matter}[g_{\mu\nu}, \psi] \ , \tag{4.20}$$

where we have assumed that the matter fields, $\psi$, couple minimally to the metric $g_{\mu\nu}$, which has Ricci scalar $R$. In fact this theory is classically equivalent to a scalar-tensor theory. To see this, consider the alternate action

$$S = \frac{M_{\rm Pl}^2}{2} \int {\rm d}^4x \sqrt{-g} \left( R + f(\Phi) + \frac{{\rm d}f}{{\rm d}\Phi}(R - \Phi) \right) + S_{\rm matter}[g_{\mu\nu}, \psi] \ , \tag{4.21}$$

with equations of motion[8]

$$(1 + f_R) R_{\mu\nu} - \frac{1}{2} \left( R + f - 2\Box f_R \right) g_{\mu\nu} - \nabla_\mu \nabla_\nu f_R = \frac{1}{M_{\rm Pl}^2} T_{\mu\nu}^{\rm matter} \tag{4.22}$$

$$\Phi = R \ , \tag{4.23}$$

[8] Note that we have assumed that $f_{,\Phi\Phi} \neq 0$ in the $\Phi$ equation of motion.

where $f_R \equiv \mathrm{d}f/\mathrm{d}R = \mathrm{d}f/\mathrm{d}\Phi$. From equation (4.23), we see that $\Phi$ is an auxiliary field—its equation of motion is non-dynamical (it does not involve time derivatives of $\Phi$). At the classical level, we may therefore use this equation to eliminate $\Phi$ from the action and reproduce the $f(R)$ action (4.20).[9]

In fact the action (4.21) is nothing more than Einstein gravity plus a canonical scalar non-minimally coupled, albeit in disguise. To make this explicit, we simultaneously make a conformal transformation and a field redefinition

$$\tilde{g}_{\mu\nu} = \left(1 + \frac{\mathrm{d}f}{\mathrm{d}\Phi}\right) g_{\mu\nu} \ , \qquad \phi = -\sqrt{\frac{3}{2}} M_{\mathrm{Pl}} \log\left(1 + \frac{\mathrm{d}f}{\mathrm{d}\Phi}\right) \ . \tag{4.24}$$

This leads to the action

$$S = \int \mathrm{d}^4x \sqrt{-\tilde{g}} \left(\frac{M_{\mathrm{Pl}}^2}{2}\tilde{R} - \frac{1}{2}\tilde{g}^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right) + S_{\mathrm{matter}}[e^{\sqrt{2/3}\phi/M_{\mathrm{Pl}}}\tilde{g}_{\mu\nu}, \psi] \ , \tag{4.25}$$

where we have defined

$$V(\phi) = \frac{M_{\mathrm{Pl}}^2}{2} \frac{\left(\phi\frac{\mathrm{d}f}{\mathrm{d}\phi} - f(\phi)\right)}{\left(1 + \frac{\mathrm{d}f}{\mathrm{d}\phi}\right)^2} \ . \tag{4.26}$$

For a suitable choice of $V(\phi)$ will exhibit chameleon screening. The potential for the scalar field is set by our choice of the function $f$, and an important thing to note is that theories that look extremely complicated in one description may be simple from the other perspective. For example, simple functions of $R$ often correspond to non-analytic potentials for the scalar $\phi$, and vice versa.

The function $f(R)$ in (4.20) is not a completely free function. There are various constraints on its form coming both from theoretical consistency and phenomenological viability. One well-studied choice for the form of the function $f(R)$ is the *Hu–Sawicki* model [16]

$$f(R) = -\frac{aM^2}{1 + \left(\frac{R}{M^2}\right)^{-\alpha}}. \tag{4.27}$$

### 4.2.4 Galileons and beyond

A particularly interesting version of a scalar tensor theory known as the galileon [17] has recently been in vogue. The novelty of these theories is that they are derivatively coupled and can become classically nonlinear; they are in some sense a scalar analogue of Einstein gravity. The defining features of the galileon are invariance under the shift symmetry

$$\phi \longmapsto \phi + c + b_\mu x^\mu, \tag{4.28}$$

and second-order equations of motion. This second property is nontrivial: something like $(\Box\phi)^2$ will certainly have the required symmetry, but will generically have higher-order equations of motion. The second-order equations of motion are desirable in order to ensure that the theory is free from

[9]In field theory language, we are integrating out $\Phi$ at tree level.

what is known as the Ostrogradsky instability. It turns out that there are only a finite number of terms in any given dimension (5 in $D = 4$) which satisfy both of these requirements. The simplest nontrivial galileon theory is described by

$$\mathcal{L} = -\frac{c_2}{2}(\partial\phi)^2 - \frac{c_3}{\Lambda^3}\Box\phi(\partial\phi)^2. \tag{4.29}$$

Though it may not look it, both of these terms are invariant under the symmetry (4.28)—up to a total derivative.

The galileons have a number of interesting properties. Maybe the most intriguing is that they obey a non-renormalization theorem: the coefficient of the cubic term above receives no quantum corrections at any order in perturbation theory (even from external fields, provided that the couplings respect the galileon symmetry). Therefore, if the galileon can be made to drive cosmic acceleration, its stress energy contribution will be radiatively stable.[10]

In fact, the galileon can act like an effective cosmological constant, to see this couple the theory minimally to gravity. The Friedmann equations then read:

$$3M_{\rm Pl}^2H^2 = \frac{c_2}{2}\dot{\phi}^2 - \frac{6Hc_3}{\Lambda^3}\dot{\phi}^3 \tag{4.30}$$

$$3M_{\rm Pl}^2H^2 + 2M_{\rm Pl}^2\dot{H} = -\frac{c_2}{2}\dot{\phi}^2 - \frac{2c_3}{3\Lambda^3}\frac{\mathrm{d}}{\mathrm{d}t}\dot{\phi}^3\ , \tag{4.31}$$

For simplicity, let's first look for exact de Sitter solutions, for which we expect that

$$\dot{\phi} = \alpha H_0 M_{\rm Pl} \qquad \text{and} \qquad H = H_0 \tag{4.32}$$

Plugging this ansatz into (4.31), we obtain

$$\alpha^2 = -\frac{6}{c_2}\ . \tag{4.33}$$

In order for $\alpha$ to be real, we have to take $c_2 = -1$, indicating that the galileon is a ghost around Minkowski space. We can then plug this into (4.30) to find $H_0$ in terms of $\Lambda$:

$$\Lambda^3 = -\alpha^3 c_3 H_0^2 M_{\rm Pl} \tag{4.34}$$

This particular model seems to be in tension with experimental data but people have generalized the above construction significantly. One of the slogans which was abstracted from the galileon is that scalar-tensor theories which have second-order equations of motion are interesting. This led to a search for the most general scalar tensor theory with this property—in fact the answer had been derived long ago by Horndeski, but had gone mostly unnoticed in the literature. Horndeski's theory provides a very general framework (perhaps too general to be useful) to investigate a variety of cosmological questions.

[10] Note that this by itself does *not* solve the CC problem—it does not explain why the large contributions from the SM fields we saw above are canceled to exquisite precision.

More recently, it has been appreciated that second-order equations of motion are not the only way to ensure the absence of ghost-like instabilities in theories with more than one degree of freedom (of which scalar-tensor theories are a special case). Instead, there can be complicated relations between the fields which enforce constraints that remove the dangerous modes. This has led to the investigation of "beyond Horndeski" theories, which are an even more general theory of a scalar degree of freedom interacting with a spin-2.

#### 4.2.5 Partially massless gravity

Massive higher spin fields on de Sitter space possess gauge symmetries at certain values of their masses. The first example of a field with more than one distinguished mass value is a massive spin-2, $h_{\mu\nu}$, on a $D$ dimensional de Sitter space of radius $H^{-1}$, which has the action

$$S = \int \mathrm{d}^D x\sqrt{-g}\left[-\frac{1}{2}\nabla_\lambda h_{\mu\nu}\nabla^\lambda h^{\mu\nu} + \nabla_\lambda h_{\mu\nu}\nabla^\nu h^{\mu\lambda} - \nabla_\mu h\nabla_\nu h^{\mu\nu} + \frac{1}{2}\nabla_\mu h\nabla^\mu h \right. \\ \left. + (D-1)\,H^2\left(h^{\mu\nu}h_{\mu\nu} - \frac{1}{2}h^2\right) - \frac{1}{2}m^2(h_{\mu\nu}h^{\mu\nu} - h^2)\right]. \tag{4.35}$$

When the mass takes the value

$$m^2 = (D-2)\,H^2\,, \tag{4.36}$$

the theory develops a scalar gauge symmetry

$$\delta h_{\mu\nu} = \left(\nabla_\mu\nabla_\nu + H^2 g_{\mu\nu}\right)\phi\,, \tag{4.37}$$

where $\phi(x)$ is a scalar gauge parameter. This is the partially massless (PM) graviton. There is of course another distinguished value of the mass, $m = 0$, corresponding to the ordinary massless graviton, which is invariant under linear diffeomorphism invariance. These are the only two values of the mass of a spin-2 for which a gauge symmetry appears. In four dimensions, a generic massive spin-2 field propagates five degrees of freedom, a massless spin-2 field propagates two degrees of freedom, and a partially massless spin-2 lies in-between, propagating 4 degrees of freedom.[11]

This partially massless theory has been of interest as a possible theory of gravity because the symmetry-enforced relation (4.36) links the value of the cosmological constant to the graviton mass. A small graviton mass is in turn technically natural due to the enhanced diffeomorphism invariance of general relativity at the value $m = 0$. This offers a tantalizing possible avenue towards solving the cosmological constant problem (see [18] and references therein). Unfortunately, there are obstructions to realizing a complete two-derivative non-linear theory that maintains the gauge symmetry and propagates the same number of degrees of freedom as the linear theory.

## References

[1] A. Einstein, "Cosmological Considerations in the General Theory of Relativity," Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1917**, 142 (1917).

[11] More generally, in $D$ dimensions, the PM graviton propagates $N_{\rm dof} = \frac{(D+1)(D-2)}{2} - 1$ degrees of freedom.

[2] A. G. Riess *et al.* [Supernova Search Team Collaboration], "Observational evidence from supernovae for an accelerating universe and a cosmological constant," Astron. J. **116**, 1009 (1998) doi:10.1086/300499 [astro-ph/9805201].

[3] S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], "Measurements of Omega and Lambda from 42 high redshift supernovae," Astrophys. J. **517**, 565 (1999) doi:10.1086/307221 [astro-ph/9812133].

[4] S. Weinberg, "The Cosmological Constant Problem," Rev. Mod. Phys. **61**, 1 (1989). doi:10.1103/RevModPhys.61.1

[5] S. Nobbenhuis, "The Cosmological Constant Problem, an Inspiration for New Physics," gr-qc/0609011.

[6] J. Martin, "Everything You Always Wanted To Know About The Cosmological Constant Problem (But Were Afraid To Ask)," Comptes Rendus Physique **13**, 566 (2012) doi:10.1016/j.crhy.2012.04.008 [arXiv:1205.3365 [astro-ph.CO]].

[7] S. Weinberg, "Anthropic Bound on the Cosmological Constant," Phys. Rev. Lett. **59**, 2607 (1987). doi:10.1103/PhysRevLett.59.2607

[8] S. Weinberg, "Photons and gravitons in perturbation theory: Derivation of Maxwell's and Einstein's equations," Phys. Rev. **138**, B988 (1965). doi:10.1103/PhysRev.138.B988

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

[10] S. Deser, "Selfinteraction and gauge invariance," Gen. Rel. Grav. **1**, 9 (1970) doi:10.1007/BF00759198 [gr-qc/0411023].

[11] A. Joyce, B. Jain, J. Khoury and M. Trodden, "Beyond the Cosmological Standard Model," Phys. Rept. **568**, 1 (2015) doi:10.1016/j.physrep.2014.12.002 [arXiv:1407.0059 [astro-ph.CO]].

[12] A. Joyce, L. Lombriser and F. Schmidt, "Dark Energy vs. Modified Gravity," arXiv:1601.06133 [astro-ph.CO].

[13] S. Capozziello, "Curvature quintessence," Int. J. Mod. Phys. D **11**, 483 (2002) doi:10.1142/S0218271802002025 [gr-qc/0201033].

[14] S. Capozziello, S. Carloni and A. Troisi, "Quintessence without scalar fields," Recent Res. Dev. Astron. Astrophys. **1**, 625 (2003) [astro-ph/0303041].

[15] S. M. Carroll, V. Duvvuri, M. Trodden and M. S. Turner, "Is cosmic speed - up due to new gravitational physics?," Phys. Rev. D **70**, 043528 (2004) doi:10.1103/PhysRevD.70.043528 [astro-ph/0306438].

[16] W. Hu and I. Sawicki, "Models of f(R) Cosmic Acceleration that Evade Solar-System Tests," Phys. Rev. D **76**, 064004 (2007) doi:10.1103/PhysRevD.76.064004 [arXiv:0705.1158 [astro-ph]].

[17] A. Nicolis, R. Rattazzi and E. Trincherini, "The Galileon as a local modification of gravity," Phys. Rev. D **79**, 064036 (2009) doi:10.1103/PhysRevD.79.064036 [arXiv:0811.2197 [hep-th]].

[18] C. de Rham, K. Hinterbichler, R. A. Rosen and A. J. Tolley, "Evidence for and obstructions to nonlinear partially massless gravity," Phys. Rev. D **88**, no. 2, 024003 (2013) doi:10.1103/PhysRevD.88.024003 [arXiv:1302.0025 [hep-th]].