A Nonlocal Approach to the Cosmological Constant Problem

Sean M. Carroll and Grant N. Remmen

Walter Burke Institute for Theoretical Physics,
California Institute of Technology, Pasadena, CA 91125

Abstract

We construct a model in which the cosmological constant is canceled from the gravitational equations of motion. Our model relies on two key ingredients: a nonlocal constraint on the action, which forces the spacetime average of the Lagrangian density to vanish, and a dynamical way for this condition to be satisfied classically with arbitrary matter content. We implement the former condition with a spatially-constant Lagrange multiplier associated with the volume form and the latter by including a free four-form gauge field strength in the action. These two features are enough to remove the cosmological constant from the Einstein equation. The model is consistent with all cosmological and experimental bounds on modification of gravity and allows for both cosmic inflation and the present epoch of acceleration.
# Contents

1 Introduction .................................................. 3

2 Theory and Dynamics ........................................... 4
   2.1 The model .................................................. 4
   2.2 Equations of motion ....................................... 5
   2.3 Solutions in general relativity ............................ 7
   2.4 Spacetime averages of the gravity and matter actions 7

3 Implications and Discussion .................................. 9
   3.1 Weinberg’s no-go theorem .................................. 9
   3.2 Locality .................................................... 10
   3.3 Relation to unimodular gravity ........................... 10
   3.4 Four-form quantization ................................... 12
   3.5 Generalization to top forms in $D$ dimensions ............ 13

4 Conclusions .................................................... 15
1 Introduction

The cosmological constant problem is the task of explaining the large hierarchy between the low observed value of the energy density of empty space and the high Planck scale of quantum gravity. One way of thinking about this problem is that, in a quantum field theory with a cutoff scale $\mu$, the energy density of the vacuum $\rho_\Lambda = \Lambda / 8\pi G$ gets a contribution of order $\mu^4$ in the absence of tuning or some symmetry (e.g., supersymmetry). Famously, the discrepancy between the value of the cosmological constant $\Lambda_{\text{obs}}$ inferred from cosmic distance measurements [1,2] and the Planck scale $M_{\text{Pl}} = 1/\sqrt{8\pi G}$ is $\Lambda_{\text{obs}} / M_{\text{Pl}}^2 \sim 10^{-120}$. While the presence of supersymmetry can ameliorate this tuning somewhat, even low-scale supersymmetry at $O(\text{TeV})$ would still leave a hierarchy of $\sim 10^{-60}$.

Many attempts have been made to solve the cosmological constant problem (see Refs. [3–5] for reviews). Approaches have included allowing the cosmological constant to dynamically relax to (nearly) zero [6–10] and anthropic arguments [3,11–13] positing a large multiverse populated with different local values of the vacuum energy. Other notable treatments of the problem include universe multiplication [14], dividing the action by the volume of the universe [15] (see also Ref. [16]), the fat graviton [17], and self-tuning brane-world models [18]. Any proposed solution to the cosmological constant problem that involves classical fields adjusting a bare vacuum energy to some lower value must face up to a no-go theorem due to Weinberg [3], which implies that no local field equations including gravity can have Minkowski solutions for generic values of the parameters.

One idea that has received considerable attention involves the introduction of a three-form gauge field $A_{\mu\nu\rho}$ with four-form field strength $F_{\mu\nu\rho\sigma}$, which in four spacetime dimensions contains no propagating degrees of freedom and contributes to the total vacuum energy [19–22]. The inclusion of such a gauge field in the action effectively makes the cosmological constant a constant of integration, since the value of the momentum of the three-form is completely free. Moreover, the gauge field can induce membrane nucleation, in which the cosmological constant is reduced in a stepwise fashion [7,8]; this mechanism suffers the drawback that the instanton tunneling rate is exponentially suppressed, with the result that by the time the cosmological constant has been neutralized, the universe is devoid of matter and radiation [5]. Other approaches, such as unimodular gravity (which degravitates the cosmological constant) [23–27] and treating gravity as an equation of state [28] similarly make the cosmological constant a constant of integration. Further approaches that are related to, but different from, the model we will present include Ref. [29], which suggested a long-distance nonlocal modification of the field equations, and Ref. [30], which suggested sequestering the cosmological constant from matter contributions by coupling the matter sector to gravity through a rescaled metric whose scaling acts as a Lagrange multiplier.

While making the value of the cosmological constant a free parameter arguably represents progress, it still leaves the open problem of explaining the large hierarchy—why should this free parameter take on such a small value? An additional ingredient is needed. One possible such ingredient comes from Euclidean quantum cosmology [19–21,31]. There, one considers the Euclidean path integral $S_E$ over all of spacetime, positing that the probability of a given field configuration goes as $e^{-S_E}$, setting the wave function for the universe to the exponential of the Euclidean effective action. For universes dominated by a cosmological constant, the action
goes as $-1/\Lambda$, so for $\Lambda > 0$ the Euclidean path integral is dominated by small $\Lambda$, provided one invents a dynamical mechanism whereby $\Lambda$ can achieve different values, rather than being a fixed parameter. In the approach of Refs. [19–21], a four-form field strength constitutes the requisite mechanism for allowing $\Lambda$ to vary. In the approach of Ref. [31], the dynamics of wormholes connecting larger universes was argued to make all constants of nature sample a distribution of possible values [32, 33], effectively making $\Lambda$ a free parameter. Unfortunately, attempts to put the Euclidean path integral for gravity on a firm footing seem to run into problems [34–36].

In this paper, we describe a model for the cancellation of the cosmological constant that is classical, but nonlocal. While our approach is somewhat ad hoc and phenomenological, it may serve as a pointer toward a more comprehensive theory. We use an arbitrary matter Lagrangian with the addition of a four-form field. For the action, we assume that its spacetime average value vanishes, which can be attained using a Lagrange multiplier that is constant over spacetime. Interestingly, this model achieves the removal of the cosmological constant from the gravitational field equations, while still allowing for the present era of accelerated expansion and without running afoul of any tests of general relativity or the equivalence principle. We present the model and its dynamics in Sec. 2, discuss its implications in Sec. 3, and conclude in Sec. 4.

2 Theory and Dynamics

2.1 The model

Any four-manifold with a Lorentzian metric $g_{\mu\nu}$ comes equipped with a natural volume form given by the Levi-Civita tensor $\epsilon = \sqrt{-g} \, d^4x$, where $g = \det(g_{\mu\nu})$. In components, $\epsilon_{\mu\nu\rho\sigma} = \sqrt{-g} \hat{\epsilon}_{\mu\nu\rho\sigma}$, where $\hat{\epsilon}_{\mu\nu\rho\sigma}$ is the Levi-Civita symbol, a density of weight 1 with $\hat{\epsilon}_{0123} = +1$. Since $\epsilon$ is a top form, any other four-form $\mu$ can be written as a scalar field $\eta(x)$ times $\epsilon$ and in principle any such four-form can be used as a volume element.

Our model is based on the assumption that the correct physical volume form used for defining the action of the theory is not necessarily the Levi-Civita tensor, but some other four-form that is also covariantly constant with respect to the metric:

$$\nabla_\lambda \mu_{\alpha\beta\gamma\delta} = \nabla_\lambda (\eta \epsilon_{\alpha\beta\gamma\delta}) = 0.$$  (1)

It is immediate that this requirement is equivalent to the parameter $\eta$ being a constant, rather than a spacetime-dependent field:

$$\nabla_\lambda \eta = 0.$$  (2)

We will discuss possible motivation for this model later, but for now we simply take as input this modification of the standard rules of the action formulation of general relativity. The upshot of this assumption is that there is a Lagrange multiplier $\eta$ in front of the action of the universe:

$$S = \eta \int d^4x \sqrt{-g} \mathcal{L},$$  (3)

enforcing that the total action vanishes,

$$\int d^4x \sqrt{-g} \mathcal{L} = 0.$$  (4)
Had we instead allowed $\eta$ to be a spacetime-dependent field multiplying the Lagrangian density in Eq. (3), this would have been equivalent to removing the requirement that the volume form $\mu$ be covariantly constant with respect to the metric. Doing so would have resulted in an equation of motion for $\eta$ that sets the Lagrangian density to zero, rather than its spacetime integral, and moreover would have introduced propagating degrees of freedom into the four-form terms in the action, allowing for $F_{\mu\nu\rho\sigma}$ to vary with spacetime and hence not act as a cosmological constant.

Let’s see what effect this mechanism has on the dynamics. Consider a general matter Lagrangian $L_m$ plus gravity, with an arbitrary cosmological constant $\Lambda$. As usual, the cosmological constant is defined such that the on-shell Minkowski vacuum value of the action describing the matter and gravitational fields, given by $L_m$ and the Einstein-Hilbert term but not $\Lambda$, vanishes.

To the action of the matter and gravitational fields we add the action for a free three-form $A$ with a boundary term and with the coefficients fixed by canonical normalization,

$$S_F = \frac{1}{2} \int_M F \wedge \star F - \int_{\partial M} A \wedge \star F. \quad (5)$$

Here, $\star$ is the Hodge dual defined with respect to the four-form $\mu$ of Eq. (1), so an overall factor of $\eta$ is implicit. Such $n$-form gauge fields are generic in string theory and a classic item in the toolbox of the theorist wishing to address the cosmological constant. For a universe without boundary, we could imagine regularizing the action by integrating it up to some specified boundary surface. As noted in Ref. [22], Eq. (5) is the appropriate form of the four-form action to use when we are considering labeling the vacua by the value of $\star dA$ on the boundary, which will be appropriate given the equation of motion we eventually derive for $A$. Here, $F = dA$ or, in components, $F_{\mu\nu\rho\sigma} = (dA)_{\mu\nu\rho\sigma} = 4\nabla_{[\mu} A_{\nu\rho\sigma]}$.

The full action for our model is thus\(^1\)

$$S = \eta \int d^4 x \sqrt{-g} \left[ \frac{1}{16\pi G} (R - 2\Lambda) + L_m - \frac{1}{48} F_{\mu\nu\rho\sigma} F^{\mu\nu\rho\sigma} + \frac{1}{6} \nabla_{\mu} (F^{\mu\nu\rho\sigma} A_{\nu\rho\sigma}) \right]. \quad (6)$$

The term $\Lambda$ appearing in Eq. (6) is the total low-energy cosmological constant in the action, other than the contribution from $F$. That is, in vacuum, $L_m$ vanishes on-shell, since the contribution of matter fields to the cosmological constant have already by definition been absorbed into $\Lambda$. It will be useful to define $L_F = -\frac{1}{48} F_{\mu\nu\rho\sigma} F^{\mu\nu\rho\sigma}$ and $L_{DJ} = \frac{1}{6} \nabla_{\mu} (F^{\mu\nu\rho\sigma} A_{\nu\rho\sigma})$, after Duncan and Jensen [22].

The basic effect of the Lagrange multiplier $\eta$ will be to force $F_{\mu\nu\rho\sigma}$ to take on a value such that it cancels the cosmological constant $\Lambda$ in the final equations of motion. We now proceed to see how this happens in practice.

### 2.2 Equations of motion

We will now analyze the equations of motion one obtains from the model (6) and ascertain their implications for gravitation. Let us introduce a regularization procedure for the volume of the

---

\(^1\)We could also have added a boundary term to the gravitational part of the action, the Gibbons-Hawking-York term, which depends on the extrinsic curvature of the boundary of the manifold; however, as this term does not qualitatively affect our results, we drop it.
spacetime $\mathcal{M}$, so that $\int d^4x\sqrt{-g} = V$. The equation of motion for $\eta$ in Eq. (4) can be rewritten as
\[
\frac{1}{V} \int d^4x\sqrt{-g}\mathcal{L} = \langle \mathcal{L} \rangle = 0,
\]
where angle brackets denote spacetime averages of the on-shell value. We can then extend this formalism to consider solutions with an infinite universe by taking the $V \to \infty$ limit in Eq. (7). For our action (6), the dynamics of $\eta$ thus require
\[
\frac{1}{16\pi G} \left( \langle R \rangle - 2\Lambda \right) + \langle \mathcal{L}_m \rangle - \frac{1}{48} \langle F_{\mu\nu\rho\sigma} F^{\mu\nu\rho\sigma} \rangle + \frac{1}{6} \langle \nabla_\mu (F^{\mu\nu\rho\sigma} A_{\nu\rho\sigma}) \rangle = 0.
\]
(8)
The cosmological constant, by definition, equals its own spacetime average value and hence does not require angle brackets.

We could have relaxed the assumption of the appearance of $\eta$ in Eq. (3). To obtain our results, we only need the constraint in Eq. (8), which can be obtained from the weaker assumption that the spacetime average value of the action vanishes. For example, we could allow $S$ to take on some finite value, as long as the spacetime volume of the universe is infinite, and still reproduce Eq. (8).

Having concluded the preliminaries, which established the condition in Eq. (8), we now turn to the problem of deriving the dynamics of this model. The equation of motion for $A$ is
\[
\nabla_\mu F^{\mu\nu\rho\sigma} = 0.
\]
(9)
At this point, we could consider dualizing $F$ to a scalar $\theta$ via
\[
F_{\mu\nu\rho\sigma} = \theta \epsilon_{\mu\nu\rho\sigma}.
\]
(10)
The equation of motion (9) implies that $\theta$ is a constant, $\nabla_\mu \theta = 0$. The addition of $\mathcal{L}_{DJ}$ means that this constant is (for the moment) arbitrary and not determined by the boundary values of $A$. Upon dualizing and using the identity $\epsilon_{\mu\nu\rho\sigma} \epsilon^{\mu\nu\rho\sigma} = -24$, we have the on-shell Lagrangians after substituting in the equation of motion (9):
\[
\mathcal{L}_F = \frac{1}{2} \theta^2,
\]
\[
\mathcal{L}_{DJ} = \frac{1}{6} (\nabla_\mu F^{\mu\nu\rho\sigma}) A_{\nu\rho\sigma} + \frac{1}{24} F_{\mu\nu\rho\sigma} F^{\mu\nu\rho\sigma} = -\theta^2,
\]
(11)
Under the dualization (10), the constraint equation (8) becomes
\[
\frac{1}{16\pi G} \left( \langle R \rangle - 2\Lambda \right) + \langle \mathcal{L}_m \rangle - \frac{1}{2} \theta^2 = 0,
\]
(12)
where Eq. (9) means that we can drop the angle brackets on $\theta$, as it is constant.

For now, we will keep the action for the four-form written in terms of the gauge field $F_{\mu\nu\rho\sigma}$ for the purposes of computing the Einstein equation, i.e., equation of motion for $g_{\mu\nu}$:
\[
\frac{1}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L})}{\delta g^{\mu\nu}} = \frac{1}{16\pi G} \left( R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} \right) - \frac{1}{2} T_{\mu\nu}
\]
\[
+ \frac{1}{96} g_{\mu\nu} F_{\alpha\beta\gamma\delta} F^{\alpha\beta\gamma\delta} - \frac{1}{12} F_{\mu\alpha\beta\gamma} F^{\alpha\beta\gamma} - \frac{1}{12} F^{\alpha\beta\gamma} F^{\alpha\beta\gamma} = 0.
\]
(13)
We used that
\[
\frac{1}{\sqrt{-g}} \frac{\delta}{\delta g^\mu\nu} \left( \sqrt{-g} F_{\alpha\beta\gamma\delta} F^{\alpha\beta\gamma\delta} \right) = -\frac{1}{2} g_{\mu\nu} F_{\alpha\beta\gamma\delta} F^{\alpha\beta\gamma\delta} + 4 F_{\mu\alpha\beta\gamma} F_{\nu}^{\alpha\beta\gamma},
\]
(14)
and we did not need to consider the total derivative term \( \nabla_\mu (\sqrt{-g} F_{\mu\nu\rho\sigma} A^{\nu\rho\sigma}) \), as its variation automatically vanishes. In Eq. (13), \( T_{\mu\nu} = -2 (-g)^{-1/2} \delta(\sqrt{-g} L_m) / \delta g^{\mu\nu} \) as usual.

Let us now dualize the four-form in Eq. (13) according to Eq. (10). Using the identity \( \epsilon^{\alpha\beta\gamma\mu} \epsilon_{\alpha\beta\gamma\nu} = -6 \delta_\mu^\nu \), we have \( F_{\mu\alpha\beta\gamma} F_{\nu}^{\alpha\beta\gamma} = -6 \theta^2 g_{\mu\nu} \) and hence the Einstein equation becomes
\[
\frac{1}{16\pi G} \left( R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} \right) - \frac{1}{2} T_{\mu\nu} + \frac{1}{4} g_{\mu\nu} \theta^2 = 0.
\]
(15)
Finally, we may substitute the spacetime average condition in Eq. (12) into the Einstein equation (15) to cancel the cosmological constant:
\[
R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \frac{1}{2} \langle R \rangle g_{\mu\nu} = 8\pi G \left( T_{\mu\nu} - \langle L_m \rangle g_{\mu\nu} \right).
\]
(16)
This is the important result of our model: a set of gravitational field equations in which the cosmological constant has been dynamically removed.

2.3 Solutions in general relativity

Let us assess the effects of the modified Einstein equation (16). First, we note that all quantities in angle brackets are by definition constants, as they are averages over the entire spacetime. As a result, conservation of \( T_{\mu\nu} \)—i.e., \( \nabla^\nu T_{\mu\nu} = 0 \)—follows in Eq. (16) from the Bianchi identity \( \nabla^\nu R_{\mu\nu} = \nabla_\nu R / 2 \) in the same way as in the usual Einstein equation. Furthermore, due to the constancy of \( \langle R \rangle \) and \( \langle L_m \rangle \), they could not have any local gravitational effect and therefore are immune to, e.g., solar system tests of general relativity and do not induce any equivalence principle violation. Effectively, \( \langle R / 2 + 8\pi G L_m \rangle \) acts as a cosmological constant, but one that would vanish for vacuum configurations of the matter field, since by Eq. (16), we have
\[
\langle \frac{1}{2} R + 8\pi G L_m \rangle = 8\pi G \left( \frac{1}{2} T - L_m \right),
\]
(17)
where \( T = g^{\mu\nu} T_{\mu\nu} \). Thus, Minkowski space is a consistent solution of Eq. (16). Moreover, since this is truly a cosmological constant, the Friedmann equations and other general relativity solutions are unmodified in Eq. (16).

2.4 Spacetime averages of the gravity and matter actions

Let us now consider what values \( \langle R \rangle \) and \( \langle L_m \rangle \) would realistically take.

By definition, \( L_m \) vanishes on-shell in vacuum, since we explicitly pulled out the cosmological constant in the action (6). Moreover, for a universe dominated by fermionic matter and radiation, one has the Lagrangian for quantum electrodynamics (and possibly the gluon kinetic
terms, which we will ignore in this discussion). For radiation, \( \mathcal{L}_m = - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \) and so vanishes on-shell for a thermal bath, in which \( \mathcal{L}_m = \frac{1}{2} (E^2 - B^2) = 0 \). Further, for massive fermionic matter, \( \mathcal{L}_m = \bar{\Psi} (i \slashed{D} - m) \Psi \). Varying the Lagrangian with respect to \( \bar{\Psi} \), we obtain the Dirac equation \( (i \slashed{D} - m) \Psi = 0 \); since this is linear in \( \Psi \) it is straightforward to substitute back into \( \mathcal{L}_m \), from which we find that \( \mathcal{L}_m \) also vanishes on-shell for massive fermionic matter (e.g., baryonic matter). The only situation relevant to cosmological tests of gravity in our universe where \( \mathcal{L}_m \neq 0 \) is a background (non-radiation) electromagnetic field, which is a very subdominant contribution to the energy budget of the universe at the present epoch and would be negligible for a universe that keeps expanding for a long time. Indeed, for matter content that is eventually diluted away by the expansion of the universe (i.e., anything other than a cosmological constant or some phantom-energy-like field), \( \langle \mathcal{L}_m \rangle \) would vanish if the universe keeps expanding forever.

Further, if the on-shell value of \( \mathcal{L}_m \) is nonzero only in a measure-zero portion of the universe, e.g., in a finite region of a spatially-infinite universe, then \( \langle \mathcal{L}_m \rangle \) also vanishes. Therefore ignoring \( \langle \mathcal{L}_m \rangle \), we turn to the question of \( \langle R \rangle \), which for vanishing \( \langle \mathcal{L}_m \rangle \) satisfies \( \langle R \rangle = 8\pi G \langle T \rangle \). For a perfect fluid, \( T = -\rho + 3p \). As long as \( w \leq \frac{1}{3} \) and \( \rho \geq 0 \), at least averaged over all of spacetime, then \( \langle T \rangle \leq 0 \) and we effectively have a negative cosmological constant in Eq. (16), which corresponds to an effectively negative vacuum energy density. Of course, if \( T \) is only nonzero over a finite spacetime region and if the universe’s spacetime volume is infinite, then \( \langle R \rangle \) vanishes and we simply have the Einstein equation with no cosmological constant, \( R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu} \). This would not be in conflict with the observed acceleration of the universe, since the present acceleration could be driven by a quintessence field that will eventually (though at arbitrary late time) turn off, leaving a finite \( \langle R \rangle \) despite exponentially expanding the universe for an arbitrarily long, but not infinite, time. In exactly the same way, a finite period of inflation is completely possible in our model and further would manifest no differences from inflation in standard general relativity.

Moreover, if the universe undergoes phase transitions in a given cosmology, the associated exponential expansion associated with the false vacuum still occurs in our model via Eq. (16), since these temporary effects do not impact spacetime-averaged quantities. The present weak apparent cosmological constant observed in our own universe then has two possible interpretations in our model. Either the universe is in its true vacuum and the accelerated expansion is being driven by some quintessence field, the dynamics of an as-yet-undetected ultralight sector, or the universe is in a false vacuum with (fine-tuned) energy very close to the true vacuum. Even in the latter case, our model has the merit of changing the cosmological constant problem from the twofold question of why we are in a false vacuum with very low energy and why the true vacuum has vanishing energy, to the single fine-tuning problem of a false vacuum with energy very close to that dictated by the generic cosmological constant \( \Lambda \) in the action. Note that all of these statements require that the classical dynamics of our model, in particular the constraint equation (12), can be solved exactly, a requirement that engenders subtleties that we will discuss in Sec. 3.4.
3 Implications and Discussion

We now address some of the subtleties of our model, including locality, quantization, and the dimensionality of spacetime. We further consider its implications for the evolution of the universe and its relation to unimodular gravity.

3.1 Weinberg’s no-go theorem

An important hurdle to be faced by any model addressing the cosmological constant problem is Weinberg’s celebrated no-go theorem [3]. Briefly, the no-go theorem forbids the existence of any solution to the cosmological constant problem within local quantum field theory that does not contain tuning. Let us briefly sketch the proof of the Weinberg no-go theorem as well as observe how the model presented in this paper evades the theorem.

The proof of the theorem proceeds by first assuming a theory described entirely by an action $S = \int d^4x \hat{L}$, where $\hat{L}$ is a local functional that encodes the dynamics of the metric $g_{\mu\nu}$ and $N$ matter fields $\psi_i$. Assuming the theory solves the cosmological constant problem, one then posits the existence of a translation-invariant field configuration (so that the spacetime is Minkowski and the fields are constant), for which the field equations satisfy

$$\frac{\partial \hat{L}}{\partial \psi_i} = 0 \quad \text{and} \quad \frac{\partial \hat{L}}{\partial g_{\mu\nu}} = 0.$$  \hspace{1cm} (18)

Weinberg then argues that despite the $N + 6$ equations of motion and the same number of unknown field values, one cannot obtain a generic (untuned) solution to Eq. (18). In particular, after imposing translation invariance, diffeomorphism symmetry is reduced to $GL(4)$, under which the Lagrangian must transform as $\hat{L} \rightarrow \det M \hat{L}$ for $M \in GL(4)$. With Eq. (18), one then finds that the Lagrangian must satisfy

$$\hat{L} = \sqrt{-g} V_0,$$  \hspace{1cm} (19)

where $V_0$ is a constant independent of the metric. Setting $V_0$ to zero is not accomplished by any equation of motion and the cosmological constant problem thus cannot be solved, within the hypotheses of the theorem, without explicitly assuming a tuning.

In essence, the Weinberg no-go theorem is the statement that the equations of motion of the metric and local matter fields do not set the zero point of the potential within quantum field theory. However, the model presented in this paper evades the hypotheses of the no-go theorem by including a nonlocal parameter, $\eta$, which couples to the entire action as in Eq. (3). The equation of motion for this field, Eq. (4), constitutes an additional constraint. Importantly, however, it does not add an additional unknown, since as we have seen the value of $\eta$ itself is arbitrary. The equation of motion for $\eta$ thus provides the constraint necessary to set $V_0$ to zero for translation-invariant solutions.

The inclusion of $\eta$ in the action is different than simply tuning $V_0$ to zero, since $\eta$ does more than merely act on the cosmological constant. Indeed, $\eta$ couples to the entire action in the egalitarian manner of Eq. (3), which in principle could have induced observable changes to
cosmological predictions. However, as we have seen, the presence of $\eta$ proves to be compatible with cosmological observations and in fact yields predictions for the nature of the present epoch of accelerated expansion, as noted in Sec. 2.4. In essence, one should view $\eta$ as a convenient mechanism for enforcing the vanishing of the action, integrated over all of spacetime. If the vanishing of the action is indeed to be a fundamental principle, its origin must come from some deeper mechanism, about which we remain agnostic in this paper.

3.2 Locality

The question of locality arises naturally in a model that contains averages of fields over all of spacetime. Any constraint on the action that does not come from a local field is necessarily nonlocal, in the sense that the constraint is an integral equation over all of spacetime. Whether this is an issue for causality should in practice be a question of whether one can use this fact to construct a causal paradox (i.e., form a closed signal trajectory or have noncommuting operators outside the lightcone). As we have shown, however, the theory obeys energy conservation, leaves tests of gravity unmodified, and, for an infinite universe, when spacetime averages over the matter fields vanish, reduces to the standard Einstein equation with zero cosmological constant. Hence, the model seems to yield no ability to produce a causal paradox or indeed even give an observable test. In particular, the only place that acausality/nonlocality is present in our model is in setting the value of the effective cosmological constant equal to $\langle R/2 + 8\pi G L_m \rangle$.

While the presence of spacetime averages in equations of motion is unusual, field equations containing spacetime averages, as in Eq. (16), have been obtained in other contexts, such as Ref. [29] (which proposed addressing the cosmological constant problem by effectively turning Newton’s constant into a high-pass filter), Ref. [30] (which replaced the metric in the matter sector with one rescaled by a Lagrange multiplier akin to our $\eta$), and Ref. [15] (which replaced the action to be varied over by an effective action that was divided by the spacetime volume).

Interestingly, in a universe with positive curvature, matter, and nothing else, our model predicts an observed negative cosmological constant with magnitude equal to $4\pi G$ times the matter density averaged over the life of the universe, which for universes that recollapse soon enough would be observable; however, our universe, by its observed expansion and vanishing curvature, does not fall into this category.

3.3 Relation to unimodular gravity

It is useful to consider the relationship between our model and another paradigm for addressing the cosmological constant problem, namely, unimodular gravity [23–27]. The similarity between our model and unimodular gravity lies in the fact that both rely on the idea of treating the volume form differently than the spacetime metric.

In our model, the volume form is dynamical and is simply required to be a four-form $\mu$ that is covariantly constant with respect to the spacetime metric. As we have seen, the extra freedom in the overall scale of the volume form produces an equation of motion that acts as a constraint on the spacetime average of the Lagrangian. Due to the presence of the four-form gauge field $F$, this constraint equation can be straightforwardly satisfied without appreciably interfering with the usual behavior of gravity on cosmological and astrophysical scales, as we saw in Secs. 2.3
and 2.4. In particular, the constraint sets the effective value of the cosmological constant to zero, while still allowing for finite periods of inflation.

In contrast, in unimodular gravity, one requires that the volume form is entirely nondynamical. Thus, in computing the equation of motion by varying the metric by $\delta g_{\mu\nu}$, one must restrict to traceless variations for which $g^{\mu\nu}\delta g_{\mu\nu} = 0$. As a result, the equation of motion one obtains is the trace-subtracted Einstein equation

$$R_{\mu\nu} - \frac{1}{4} R g_{\mu\nu} = 8\pi G \left( T_{\mu\nu} - \frac{1}{4} g_{\mu\nu} T \right). \tag{20}$$

The cosmological constant term in the action, which couples to gravity only through its multiplication by the volume form, thus entirely decouples in unimodular gravity. Furthermore, unlike in general relativity, conservation of energy-momentum $\nabla_{\mu} T^{\mu\nu} = 0$ does not follow automatically from the Bianchi identity $\nabla^{\mu} R_{\mu\nu} = \nabla_{\nu} R / 2$, but rather is imposed independently; taking the derivative of both sides of Eq. (20), we have $\nabla_{\mu} R + 8\pi G \nabla_{\mu} T = 0$, so one can identify $R + 8\pi G T$ as a constant of integration that can be suggestively labeled $4\Lambda_{\text{int}}$. Then Eq. (20) reduces to the Einstein equation,

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda_{\text{int}} g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{21}$$

but where, instead of containing the bare cosmological constant in the action, simply contains a constant of integration. One could then apply the typical anthropic or Euclidean path-integral arguments for why $\Lambda_{\text{int}}$ should take a particular value.

In contrast, our model selects an effective zero cosmological constant via the dynamics of the volume form itself, so the cosmological constant as an arbitrary constant of integration does not arise. Interestingly, taking the trace of the field equation for our model (16) and then taking the spacetime average, we have

$$\langle L_m \rangle = \frac{1}{4} \left( \langle T \rangle - \frac{1}{8\pi G} \langle R \rangle \right), \tag{22}$$

so plugging this back into Eq. (16), we have

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \frac{1}{4} \langle R \rangle g_{\mu\nu} = 8\pi G \left( T_{\mu\nu} - \frac{1}{4} \langle T \rangle g_{\mu\nu} \right). \tag{23}$$

Taking the trace once more, we find that

$$R + 8\pi G T = \langle R + 8\pi G T \rangle, \tag{24}$$

so we are able to eliminate the spacetime averages from Eq. (23), recovering the equation

$$R_{\mu\nu} - \frac{1}{4} R g_{\mu\nu} = 8\pi G \left( T_{\mu\nu} - \frac{1}{4} T g_{\mu\nu} \right). \tag{25}$$

That is, our model predicts the trace-free Einstein equation (20) and is thus just as consistent with observations as unimodular gravity. However, rather than finding that the cosmological constant is a constant of integration as in Eq. (21), our model predicts that it is constrained by the dynamics to equal $\langle R + 8\pi G T \rangle / 4 = \langle R / 2 + 8\pi G L_m \rangle$. That is, in addition to the trace-free Einstein equation, our model yields the additional equations of motion given by Eqs. (12), (23), and (24).
3.4 Four-form quantization

The preceding discussion has been entirely about the classical theory. We now turn to the question of quantization of the four-form, as discussed in Ref. [37], and its relation to the dynamics of our model. In a spacetime of purely four dimensions (i.e., four-dimensional on all length scales), the four-form $F = F_4$ is nondynamical to arbitrarily short wavelengths and its dual $-\theta = F_0 = *_4 F_4$ can take on a continuum of values, allowing the constraint equation (8) to be satisfied exactly. Here, we denote the Hodge dual in $D$ dimensions as $*_D$, which in this section for clarity we define with respect to the conventional volume form $\epsilon$, not $\mu$. A spacetime of exactly $D = 4$ is of course compatible with what we know about the world, in which case the predictions of our model regarding the vacuum discussed at the end of Sec. 2.4 are in force.

However, if there are more than four dimensions, as for example in string theory, then the story changes somewhat. In eleven spacetime dimensions, one has a four-form $F_4$ and its dual $F_7 = -*_11 F_4$. If the spacetime geometry is $K_7 \otimes M_4$, where $M_4$ is a four-dimensional noncompact geometry and $K_7$ is a spacelike seven-dimensional compact manifold, then the eleven-dimensional four-form can remain a four-form in the effective field theory on $M_4$, while the seven-form becomes a zero-form in the compactified theory. Moreover, string theory naturally produces 5-branes and 2-branes, to which the gauge fields $A_6$ and $A_3$ respectively couple, where $F_n = dA_n$. As a result, the usual Dirac quantization condition applies, meaning that if we look at solutions in which the seven-form wraps $K$, we require

$$\int_K F_7 = 2\pi n \text{ for some } n \in \mathbb{Z},$$

(26)

where in this subsection we absorb the charge into the definition of the gauge field for simplicity. The equation of motion for $F_7$ implies that $F_7 = f_7 \epsilon_7$, where $\epsilon_7$ is the seven-dimensional volume form, so $f_7 V_K = 2\pi n$, where $V_K$ is the volume of $K$. The four-form becomes

$$F_4 = *_{11} F_7 = f_7 \epsilon_4 = \frac{2\pi n}{V_K} \epsilon_4.$$  

(27)

Thus, the value of the four-form is quantized, in units set by the volume of the compactified space.

Naively, this quantization condition on the allowed classical values of the four-form is an obstacle to our mechanism, in which the four-form adjusts to cancel the cosmological constant exactly. On closer examination, however, such quantization can be completely compatible with our approach.

In Ref. [37] it was shown that by adding multiple four-forms to the theory, there existed solutions in which the effective step size of the quantization, and hence the effective minimal size of the cosmological constant for fixed $V_K$, could be decreased. Let us suppose that in our model we have $O(10^2)$ different four-form fluxes, which Ref. [37] argued is both plausible in string theory and sufficient to cancel the bare cosmological constant to within the value suggested by observations, and let $\Delta \Lambda$ be the difference, i.e., the minimal effective cosmological constant in the action. The equations of motion in our model, in particular the constraint equation (8) coming from the dynamics of $\eta$, come from the usual assumption that the principle of least action dictates that the classical dynamics must be a saddle point of the action as a functional
of the fields. While this is valid classically, it does not take into account the possibility explored in this section that quantum mechanics may dictate a discretuum of field values, in this case for the four-form flux. This means that the principle of least action, as applied to this system, selects field configurations in which the variation of \( S \) is nonvanishing, but minimized.

Typically, the dominant contribution to the path integral comes from trajectories for which the action is stationary, so that the phases for nearby paths do not destructively interfere. However, if quantization of the four-form flux makes it impossible for the equation of motion for \( \eta \) in Eq. (8) to be exactly satisfied, then these saddle points are forbidden and the path integral will be dominated by field configurations for which nearby paths interfere least destructively, i.e., those for which \( \delta S \) is minimized. That is, the equation of motion for \( \eta \) requires field strengths for the four-forms that minimize the effective cosmological constant. We then have \( \delta S \propto \Delta \Lambda \) when the fields are varied. Replacing the zero on the right side of Eq. (12) with \( \mathcal{O}(\Delta \Lambda) \) adds a cosmological constant of order \( \Delta \Lambda \) to the field equation (16). For a positive cosmological constant, the matter Lagrangian and energy-momentum vanish as the universe expands, so taking the spacetime average of Eq. (16) leads to \( \langle R \rangle \sim \Delta \Lambda \) and thus de Sitter space is a consistent solution.

### 3.5 Generalization to top forms in \( D \) dimensions

We can generalize our model to a spacetime \( M \) of arbitrary dimension \( D > 4 \), which will eventually be compactified on a manifold \( K \) of \( D - 4 \) dimensions. Consider a model with a top form \( F_D = dA_{D-1} \) coupled to gravity, with bare \( D \)-dimensional cosmological constant \( \Lambda \) and an overall Lagrange multiplier field \( \eta \):

\[
S = \eta \int d^Dx \sqrt{-g} \frac{1}{16\pi G} (R - 2\Lambda) + \frac{1}{2}\eta \int_M F_D \wedge \star_D F_D - \eta \int_{\partial M} A_{D-1} \wedge \star_D F_D, \tag{28}
\]

where \( \star_D \) is defined here with respect to the volume form associated with the metric, \( \epsilon \). The equation of motion for \( F_D \) is \( \nabla_a F_{a_1 \cdots a_D} = 0 \), where we use Roman indices for the \( D \)-dimensional spacetime. Hence, \( F_{a_1 \cdots a_D} = \theta \epsilon_{a_1 \cdots a_D} \) for \( \theta \) = constant. We have \( F_{a_1 \cdots a_D} = D \nabla [a_1 A_{a_2 \cdots a_D}] \).

The equation of motion for \( \eta \) sets the spacetime average of the Lagrangian density to zero. The canonical normalization of the \( D \)-form in terms of its components is

\[
\mathcal{L}_F + \mathcal{L}_{DJ} = -\frac{1}{2D!} F_{a_1 \cdots a_D} F^{a_1 \cdots a_D} + \frac{1}{(D-1)!} \nabla_a (F^{a_1 \cdots a_D} A_{a_2 \cdots a_D}), \tag{29}
\]

so dualizing, we have the on-shell Lagrangians just as in Eq. (11):

\[
\mathcal{L}_F = \frac{1}{2} \theta^2
\]

\[
\mathcal{L}_{DJ} = \frac{1}{(D-1)!} \nabla_a (F^{a_1 \cdots a_D} A_{a_2 \cdots a_D}) + \frac{1}{D!} F_{a_1 \cdots a_D} F^{a_1 \cdots a_D} = -\theta^2 \tag{30}
\]

where we used the fact that \( \epsilon_{a_1 \cdots a_D} \epsilon^{a_1 \cdots a_D} = -D! \) in mostly-plus signature. Thus, the constraint from \( \eta \) becomes

\[
\frac{1}{16\pi G} (\langle R \rangle - 2\Lambda) - \frac{1}{2} \theta^2 = 0. \tag{31}
\]
Meanwhile, the equation of motion for $g_{ab}$ is
\[
\frac{1}{16\pi G} \left( R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} \right) + \frac{1}{4(\text{D})!} g_{ab} F_{a_1 \cdots a_D} F^{a_1 \cdots a_D} - \frac{1}{2(\text{D} - 1)!} F_{a_2 \cdots a_D} F^{a_2 \cdots a_D} = 0, \tag{32}
\]
so dualizing using the identity $\epsilon_{a_2 \cdots a_D} \epsilon^{b a_2 \cdots a_D} = -(\text{D} - 1)! \delta^b_a$, we have
\[
\frac{1}{16\pi G} \left( R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} \right) + \frac{1}{4} g_{ab} \theta^2 = 0. \tag{33}
\]
Hence, putting together Eq. (33) with Eq. (31), we end up with the same equation of motion as in Eq. (16), which for vacuum solutions is
\[
R_{ab} - \frac{1}{2} R g_{ab} + \frac{1}{2} \langle R \rangle g_{ab} = 0, \tag{34}
\]
Adding matter would proceed in exactly the same way as in Sec. 2.

Let us now compactify the spacetime on $K$. We define $V_K = \int_K d^{D-4}x \sqrt{-g}$. We assume a product metric for simplicity and ignore the degrees of freedom associated with the components that mix compact and noncompact directions, so that $R_{ab}$ is block-diagonal and thus we can define $R = (4)R + R_K$, where $R_K$ goes as the contraction of the components of the Ricci tensor along $K$. The low-energy gravitational sector action becomes:
\[
S_G = \frac{V_K}{16\pi G} \int d^4x \sqrt{-(4)g} \left( \langle R_K \rangle + (4)R - 2\Lambda \right), \tag{35}
\]
where $\langle R_K \rangle$ is the value of $R_K$ averaged over $K$, which is equivalent to its value averaged over all spacetime, since $R_K$ is by definition independent of the noncompact directions. The on-shell value of the Lagrangian for the $D$-forms in the four-dimensional effective theory is $V_K \theta^2/2$. That is, everything except the matter action is rescaled by $V_K$, so the cancellation of the bare cosmological constant works the same as in the purely four-dimensional model of Sec. 2. However, there is now an additional term to the cosmological constant, namely, $\langle R_K \rangle$, the average of the Ricci scalar on the compact space, which to the four-dimensional effective theory appears as a constant. This additional contribution is canceled along with the bare constant $\Lambda$ by the equation of motion for $\eta$.

We can arrive at this result in another way, by considering the $D$-dimensional Einstein equation (34), evaluated on the noncompact four-dimensional subspace (which we will parameterize with Greek indices):
\[
(4)R_{\mu\nu} - \frac{1}{2} (4)R + \langle 4)R \rangle g_{\mu\nu} + \frac{1}{2} \langle (4)R + \langle 4)R \rangle \rangle g_{\mu\nu} = 0, \tag{36}
\]
where we consider vacuum solutions for simplicity. (If we added matter, the coupling would go as $G_4 = G/V_K$.) That is,
\[
(4)R_{\mu\nu} - \frac{1}{2} (4)R g_{\mu\nu} + \frac{1}{2} \langle (4)R \rangle g_{\mu\nu} = 0. \tag{37}
\]
Now, the value of $\theta$ is not quantized, unlike the situation in Sec. 3.4, since the top form $F_D$ can take on arbitrary values (as its dual, a zero-form, does not wrap any manifold).
4 Conclusions

We have presented a simple model in which a Lagrange multiplier parameter \( \eta \) allows the energy density of a four-form gauge field strength to exactly cancel the cosmological constant. The model is unusual in that \( \eta \) is not a dynamical field and therefore its effects are nonlocal. In particular, the equations of motion involve averages over all of spacetime. Perhaps most obviously, there is no especially good motivation for including such a Lagrange multiplier, other than that it gives us the answer we want.

On the other hand, the model has a number of attractive features. It cancels the low-energy cosmological constant exactly, at least up to possible corrections due to quantization of the four-form field strength. It does so while remaining compatible with Weinberg’s no-go theorem and in a way that seems compatible with all known observations. It has the consequence that the late-time vacuum energy is exactly zero, so that our current period of acceleration is necessarily temporary. We therefore think the model is worth considering on purely phenomenological grounds.

Models that adjust the cosmological constant to zero are often viewed with suspicion, with some justification. Even if such a mechanism sets the vacuum energy to zero classically, it is generally difficult to protect such a value against radiative corrections. The situation here is somewhat unusual in that respect. The heavy lifting in our model is done by the Lagrange multiplier \( \eta \), which is nondynamical. We therefore do not expect its dynamics to be subject to loop corrections.

Perhaps a more pressing question is that of the naturalness of the form we chose for how \( \eta \) enters the action in Eq. (3). We motivated that choice by deriving it from the requirement that the volume form \( \mu \) be covariantly constant, which is admittedly ad hoc. One could certainly imagine, for example, more general actions of the form \( f(\eta) \int d^4 x \sqrt{-g} \mathcal{L} \), in which case the constraint that comes from varying with respect to \( \eta \) could be satisfied either by setting the conventional action to zero (as in the version we’ve been considering) or by setting \( df/d\eta = 0 \). The latter choice is simply equivalent to having no constraint in the first place. Alternatively, one could imagine different functions of \( \eta \) multiplying different terms in the Lagrangian:

\[
S = \int d^4 x \sqrt{-g} \left[ \sum_i f_i(\eta) \mathcal{L}_i \right].
\]

Then the constraint from varying \( \eta \) would generically not match the condition that the vacuum energy vanish and we would still have a cosmological constant problem.

There is therefore undoubtedly a choice that we made while constructing the model: that the \( \eta \) constraint enforce the vanishing of an otherwise conventional action (with a four-form gauge field). In our view, this is best understood as a phenomenological approach to a true dynamical mechanism that is yet to be understood, rather than as a complete theory in its own right. For example, given that the Feynman path integral sums over terms of the form \( \exp(-iS/h) \), perhaps our action-minimization procedure could be derived from a principle that treated Planck’s constant \( h \) as a Lagrange multiplier. We leave exploration of this and other possible underlying principles for future work.
Acknowledgments

We thank Cliff Cheung and John Schwarz for helpful comments, as well as Monica Guica and Eugene Lim for discussions and collaboration at an earlier stage of this work. This research is funded in part by the Walter Burke Institute for Theoretical Physics at Caltech, by DOE grant DE-SC0011632, and by the Foundational Questions Institute. G.N.R. is supported by a Hertz Graduate Fellowship and a NSF Graduate Research Fellowship under Grant No. DGE-1144469.

References


