Observational Viability and Stability of Nonlocal Cosmology

S. Deser*

California Institute of Technology, Pasadena, CA 91125 and
Department of Physics, Brandeis University, Waltham, MA 02254

and

R. P. Woodard†

Department of Physics, University of Florida, Gainesville, FL 32611

ABSTRACT

We show that a class of nonlocal gravity models, proposed to explain current cosmic acceleration without dark energy, passes two major tests: First, they can be defined so as not to alter the, observationally correct, general relativ-ity predictions for gravitationally bound systems. Second, they are stable, ghost-free, with no additional excitations beyond those of general relativity. In this they differ from their, ghostful, localized versions. The systems’ initial value constraints are the same as in general relativity, and our nonlocal modifications never convert the original gravitons into ghosts.

PACS numbers: 95.36.+x, 04.50.Kd, 11.10.Lm

* e-mail: deser@brandeis.edu
† e-mail: woodard@phys.ufl.edu
1 Introduction

Explaining the current phase of cosmic acceleration is an ongoing challenge [1]. The data are consistent with general relativity operating on a critical energy density whose current composition is about 70% cosmological constant plus about 30% nonrelativistic, and small amounts of relativistic, matter [2, 3]. However, there is no good explanation for why the cosmological constant should be so small, nor why it should recently have come into dominance [4]. Scalar potential models [5, 6] can be devised to reproduce the observed expansion history [7, 8] but they must be fine tuned and are difficult to motivate. Quantum effects from a very light scalar have also been suggested [9].

Various modifications of general relativity that generalize its Lagrangian from $R$ to $f(R)$ [10, 11] represent the only local, metric-based, generally coordinate invariant and stable modification of gravity [12]. But the only model within this class that exactly reproduces the $\Lambda$CDM expansion history is general relativity with $f(R) = R - 2\Lambda$ [13].

More modification freedom is available if locality is abandoned [14], but this novel territory raises the worry of new degrees of freedom (D0F), possibly of instability-negative energy [15]. While we do not believe such models to be fundamental, even if observationally viable in some regime of validity, they must still face the above problems of principle, as well as more phenomenological ones. Their origin should arise as summed quantum graviton corrections that grew non-perturbatively during the primordial inflation epoch [7]. This conjecture, while plausible [16, 17], is as yet unverified [18]. Independent of their ultimate origin, these models have been proposed and studied purely phenomenologically. Our model [19] adds the nonlocal piece, $\Delta L \equiv \frac{1}{16\pi G} R \sqrt{-g} \times f\left(\frac{1}{16\pi G} R\right)$, to the Einstein term $R \sqrt{-g}/16\pi G$. Our signature is $(-++++)$, with the convention $R_{\mu\nu} \sim +\partial_\rho \Gamma^\rho_{\mu\nu}$. The inverse of the (scalar) d’Alembertian $\Box \equiv (-g)^{-\frac{1}{2}} \partial_\mu [\sqrt{-g} g^{\mu\nu} \partial_\nu]$ is the retarded one, with vanishing 0th and 1st time derivatives at the initial time [19]. In addition to simplicity, the great advantage of this class of models is to provide a natural delay for the onset of cosmic acceleration: because the Ricci scalar $R$ vanishes during radiation domination, $\Box^{-1} R$ cannot begin to grow until after the onset of matter domination; thereafter, because of the propagator, its growth becomes logarithmic.
The model’s defining equations take the form, 
\[ G_{\mu\nu} + \Delta G_{\mu\nu} = 8\pi G T_{\mu\nu}, \]
with \[ \Delta G_{\mu\nu} = \left\{ f\left(\frac{1}{\Box} R\right) + \frac{1}{\Box} \left[Rf'\left(\frac{1}{\Box} R\right)\right]\right\} \]
\[ + \left[\delta_{\mu}^{(\rho} \delta_{\nu)}^{\sigma) - \frac{1}{2} g_{\mu\nu} g^{\rho\sigma}\right] \partial_{\rho} \left(\frac{1}{\Box} R\right) \partial_{\sigma} \left[Rf'\left(\frac{1}{\Box} R\right)\right]. \tag{2} \]

The form of the nonlocal distortion function \( f(X) \) can, unlike the local models \[ \text{[13]}, \] be chosen to reproduce the ΛCDM background cosmology exactly \[ \text{[20, 21, 22]}. \] Indeed, there is a simple analytic form for \( f(X) \), effectively equivalent to the numerical solution \[ \text{[21]}, \]
\[ f(X) \approx 0.245 \left[ \tanh\left(0.350 Y + 0.032 Y^2 + 0.003 Y^3\right) - 1 \right], \quad Y \equiv X + 16.5. \tag{3} \]

Like all modified gravity theories, nonlocal cosmology can be differentiated from general relativity with dark energy by how it alters results in the solar system and how it affects structure formation \[ \text{[23]}. \] Koivisto has argued that there are no conflicts with solar system constraints \[ \text{[20]}. \] A recent study of structure formation by Park and Dodelson revealed deviations from general relativity in the 10%-30% range, which are interesting because they are not currently excluded and should be observable by the next generation of large scale structure surveys \[ \text{[24]}. \] While we await for these observations, it is worth examining the theoretical consistency of nonlocal cosmology in its own right. In particular, how does the model behave for gravitationally bound systems, does it possess extra degrees of freedom and is it stable? Those are the questions we will study in sections \[ \text{[2]} \text{, [3]} \text{ and [4]} \] respectively.

2 Screening: Absence of Effects on Bound Systems

In this section we discuss the issue of screening in modified theories; \( f(R) \) models suffer from the major problem that \( R \) typically has the same sign for cosmology, where we want big effects to explain the acceleration data, and for the solar system, where significant deviations from general relativity are excluded by the data. This has prompted the development of elaborate “chameleon mechanisms” in which the extra scalar degree of freedom present
in $f(R)$ models is light in cosmological settings and heavy inside the solar system [25]. Nonlocal cosmology differs from $f(R)$ models in two crucial ways: there are no extra degrees of freedom to mediate new forces; and the factor of $\Box^{-1}$ acting on $R$ allows us to define the nonlocal distortion function so that there are no changes from general relativity at all in a gravitationally bound system, without affecting the model’s predictions for cosmology. The first point will be demonstrated in section 4; it is the second point which concerns us here.

The key fact is that the scalar d’Alembertian $\Box \equiv (-g)^{-\frac{1}{2}} \partial_{\mu} (\sqrt{-g} g^{\mu\nu} \partial_{\nu})$ has OPPOSITE SIGNS when acting on functions of time than on functions of space. In the background cosmology, and perturbations about it, the time dependence of the Ricci scalar is stronger than its space-dependence. This means that $\Box^{-1} R$ is typically negative for cosmology. Indeed, reproducing the $\Lambda$CDM expansion history fixes the nonlocal distortion function $f(X)$ only for negative $X$ [21].

Although gravitationally bound systems are not always static, it is generally true that their space-dependence is stronger than on time. That means $\Box^{-1} R$ is positive inside a gravitationally bound system. Further, reproducing the $\Lambda$CDM expansion history requires $f(0) = 0$ [21]. To completely annul all corrections inside gravitationally bound systems it suffices to define $f(X) = 0$ for all $X > 0$. Hence there is a very simple way for nonlocal models to completely screen inside the solar system, the galaxy, or any other gravitationally bound system, all without affecting the model’s behavior for cosmology.

### 3 Local versus Nonlocal Formulations

Soon after our nonlocal model [19] appeared, a “localized” version, based on two additional scalars, was proposed [26, 27]. Briefly, it replaced the nonlocal terms in (1) by

$$ Rf(\frac{1}{\Box} R) \sqrt{-g} \rightarrow Rf(\Phi) \sqrt{-g} + \Psi(\Phi - R) \sqrt{-g}. \tag{4} $$

The local mechanism then, relied on two new scalars: $\Psi$ is a Lagrange multiplier than enforces $\Phi = \Box^{-1} R$ to recover the original nonlinearity. However, the price is too high, as we now show. The off-diagonal term $\Psi \Box \Phi$ is just

---

1 We thank G. Esposito-Farese for explaining the material in this section to us.
the difference of two diagonal free scalar actions \((S_+, S_-) \equiv \frac{1}{2}(\Phi + \Psi, \Phi - \Psi)\), one of which must therefore necessarily be a ghost.

We now allay the worry that this disease also infects the original system. Clearly, (4) only yields (1) after discarding precisely the homogeneous scalar’s solutions by requiring that they, and their first time derivatives, vanish at the initial time; this precisely discards their DoF! The worry now might arise that perhaps these excitations could somehow appear in the original, purely metric, model. But general relativity itself offer the prime example of how this danger is averted: the “Newtonian”, third (after the two \(g^{TT}_{ij}\) gravitons) mode is indeed dangerous if dynamical — BUT is saved from propagating by general relativity’s constraint equations. This salvation of (1) will indeed be demonstrated in the next section.

4 Nonlocal Stability

We will proceed for concreteness in a particular, synchronous, gauge. There we will see that the nonlocal equations require the same initial data, subject to exactly the same constraints, as general relativity. We will also offer a simple, gauge independent, linearized treatment, and conclude with the demonstration that none of the common DoF is converted to ghost stature by the nonlocal corrections.

4.1 Synchronous gauge

Synchronous gauge is the coordinate frame of a system of timelike, freely falling observers who are released from a spacelike surface with zero initial relative velocities \[ds^2 = -dt^2 + h_{ij}(t, \vec{x})dx^i dx^j.\] (5)

The basic analysis and conclusions should apply in any gauge, as we will see they do at linearized, kinematical level.

In synchronous gauge the covariant scalar d’Alembertian takes the form,

\[
\square = -\partial_t^2 - \frac{1}{2}h^{ij}h_{ij}\partial_t + \frac{1}{\sqrt{h}} \partial_i (\sqrt{h} h^{ij} \partial_j) .
\] (6)

\(^2\)While this gauge has well-known problems with caustics, they are not relevant to our treatment.
Here and henceforth, $h_{ij}$ denotes the inverse of the spatial metric $h_{ij}$, $h$ stands for the determinant of $h_{ij}$, and an overdot represents differentiation with respect to time. The various curvatures we require are,

$$R_{00} = -\frac{1}{2} h^{k\ell} \ddot{h}_{k\ell} + \frac{1}{4} h^{ik} h^{j\ell} \dot{h}_{ij} \dot{h}_{k\ell},$$  

$$R_{ij} = \frac{1}{2} \ddot{h}_{ij} + \frac{1}{4} h^{k\ell} \dot{h}_{ij} \dot{h}_{k\ell} - \frac{1}{2} h^{ik} \dot{h}_{ik} \dot{h}_{j\ell} + 3 \dot{R}_{ij},$$  

$$R = h^{k\ell} \ddot{h}_{k\ell} + \frac{1}{4} h^{ij} h^{k\ell} \dot{h}_{ij} \dot{h}_{k\ell} - \frac{3}{4} h^{ik} h^{j\ell} \dot{h}_{ij} \dot{h}_{k\ell} + 3 \ddot{R},$$

where a superscript “3” means, as usual, that the curvature is intrinsically that of 3-space.

### 4.2 Initial value data and constraints

Let us first see that the nonlocal field equations (2) require the same initial value data as general relativity, namely, the values of the 3-metric and its first time derivative at $t = 0$: $h_{ij}(0, \vec{x})$ and $\dot{h}_{ij}(0, \vec{x})$. The retarded Green’s function associated with $\Box^{-1}$ is defined by the differential equation,

$$\sqrt{h} \Box G[h](t, \vec{x}; t' \vec{x'}) = \delta(t - t') \delta^3(\vec{x} - \vec{x'}),$$  

subject to retarded boundary conditions,

$$G[h](0, \vec{x}; t' \vec{x'}) = 0, \quad \dot{G}[h](0, \vec{x}; t' \vec{x'}) = 0.$$  

Even though we cannot solve equations (10-11) for an arbitrary 3-metric, their form clearly defines the Green’s function $G[h]$ at time $t$ using only the values of $h_{ij}$ and its first time derivative for times less than or equal to $t$.

Because $\Box^{-1} R$ is the integral of $G[h](t, \vec{x}; t' \vec{x'})$ multiplied by the Ricci scalar, we need only consider the second time derivatives of the $\dot{R}$; the lower time derivatives and all spatial derivatives are shielded by the inverse differential operator. From expression (9) we see that these second time derivatives can be written in form,

$$R = \partial_t^2 \ln(h) + \frac{1}{4} \left(h^{ij} h^{k\ell} + h^{ik} h^{j\ell}\right) \dot{h}_{ij} \dot{h}_{k\ell} + (3) \ddot{R}.$$  

Now use relation (6) to express second time derivatives in terms of the scalar d’Alembertian,

$$\partial_t^2 = -\Box - \frac{1}{2} h^{ij} \dot{h}_{ij} \partial_t + \frac{1}{\sqrt{h}} \partial_t \left(\sqrt{h} h^{ij} \partial_j\right).$$
We can obviously combine relation (13) with (12) to conclude,

\[
R = -\Box \ln(h) + \frac{1}{4} \left( h^{ik}h^{\ell j} - h^{ij}h^{k\ell} \right) \dot{h}_{ij} \dot{h}_{k\ell} + h^{ij} \left( \Gamma^k_{ij,k} + \Gamma^k_{ki,j} - \Gamma^k_{k\ell} \Gamma^\ell_{ij} - \Gamma^k_{ki} \Gamma^\ell_{\ell j} \right).
\]

(14)

Here \( \Gamma^k_{ij} \equiv \frac{1}{2} h^{k\ell} (h_{\ell i,j} + h_{j\ell,i} - h_{ij,\ell}) \) is the spatial affine connection and a comma denotes partial differentiation.

With relations (10-11), equation (14) shows that \(-1R\) involves only the usual initial value data of general relativity: \( h_{ij}(0, \vec{x}) \) and \( \dot{h}_{ij}(0, \vec{x}) \). We can show that these initial value data are apportioned, also as in general relativity, between constrained fields and gravitational radiation modes by examining the nonlocal corrections \( \Delta G_{00} \) and \( \Delta G_{0i} \) to the constraint equations. Note first from (11) that \(-1R\) and its first time derivative both vanish at \( t = 0 \). Further, the nonlocal distortion function vanishes at \( t = 0 \). So we need only examine the two terms of (2) in which two covariant derivatives act upon \( f(-1R) + [Rf'(-1R)] \). It is easy to see that neither of the two combinations in the constraint equations contains a second time derivative,

\[
g_{00} \Box - D_0 D_0 = \frac{1}{2} h^{k\ell} \dot{h}_{k\ell} \partial_t - \frac{1}{\sqrt{h}} \partial_k \left( \sqrt{h} h^{k\ell} \partial_\ell \right),
\]

(15)

\[
g_{00} \Box - D_0 D_i = -\partial_t \partial_i + \frac{1}{2} h^{k\ell} \dot{h}_{ik} \partial_\ell.
\]

(16)

Hence we conclude that the nonlocal corrections to the constraint equations vanish at \( t = 0 \),

\[
t = 0 \implies \Delta G_{00} = 0 = \Delta G_{0i}.
\]

(17)

This completes the verification that the nonlocal model and general relativity share the same initial data and constraints.

### 4.3 No ghosts

To see that there are no ghosts it suffices to examine the second derivative terms (in synchronous gauge) of the dynamical equations, \( G_{ij} + \Delta G_{ij} = 8\pi GT_{ij} \). The second derivatives of \( h_{ij}(t, \vec{x}) \) in the Einstein tensor follow from expressions (8-9),

\[
G_{ij} = \frac{1}{2} \ddot{h}_{ij} - \frac{1}{2} h_{ij} h^{k\ell} \ddot{h}_{k\ell} + O(\partial_t).
\]

(18)
Of course it is only the first term, \( \frac{1}{2} \delta_{ij} \), that deals with unconstrained fields; the second term represents completely constrained fields. Because general relativity has no ghosts, we need only check that the nonlocal corrections in \( (2) \) don’t change the sign of the \( \frac{1}{2} \delta_{ij} \) term in \( (18) \).

The work of the previous subsection shows that local second time derivatives can only come from the parts of \( \Delta G_{ij} \) which either multiply \( G_{ij} \) or have two covariant derivatives acting on \( f(\Xi^{-1}R) + \Xi^{-1}[Rf'(\Xi^{-1}R)] \). The latter terms are simple to analyze,

\[
g_{ij} \Xi - D_i D_j = h_{ij} \Xi + O(\partial_t) .
\] (19)

The local second derivative terms are therefore,

\[
G_{ij} + \Delta G_{ij} = \frac{1}{2} \delta_{ij} \times \left[ 1 + f(\frac{1}{\Xi} R) + \frac{1}{\Xi} [Rf'(\frac{1}{\Xi} R)] \right] \\
- \frac{1}{2} \delta_{ij} \delta^{kl} \delta_{kl} \times \left[ 1 + f(\frac{1}{\Xi} R) + \frac{1}{\Xi} [Rf'(\frac{1}{\Xi} R)] - 4f'(\frac{1}{\Xi} R) \right] + O(\partial_t) . \] (20)

Only the first line of expression (20) represents the unconstrained, dynamical part of \( h_{ij} \). By comparing with the approximate analytic form \( (3) \) of the nonlocal distortion function \( f(X) \) we see that the coefficient of the dynamical term is reduced at late times, but never by enough to make it change sign. We therefore conclude that no dynamical graviton mode ever becomes a ghost.

### 4.4 No linearized ghosts

As a complement to our detailed treatment of DoF in the full nonlinear theory, the present subsection is devoted to the linearized (about flat space \( g_{\mu\nu} = \eta_{\mu\nu} + k_{\mu\nu} \)) treatment of the problem. This has several advantages: First, it is of course simpler, yet it retains the main point of the DoF analysis, since their content resides. Second, it allows us to treat the desired results gauge invariantly. Of course, the full treatment is needed to make sure no higher order failure of the critical constraint equations occurs, as notoriously happen in generic massive gravity models \[29\].

We first derive the relevant field equation; varying \( R^{\text{lin}} \partial^{-2} R^{\text{lin}} \) yields

\[
\Delta G^{\text{lin}}_{\mu\nu} = (\eta_{\mu\nu} \partial^2 - \partial_\mu \partial_\nu) \frac{1}{\partial^2} R^{\text{lin}} \equiv \Pi_{\mu\nu} \frac{1}{\partial^2} R^{\text{lin}} , \quad R^{\text{lin}} = \Pi^{\rho\sigma} k_{\rho\sigma} . \] (21)

The transverse projector’s \( 0\mu \) components are respectively of zero and first order in time derivatives: \( \Pi_{00} = -\nabla^2 \), \( \Pi_{0i} = -\partial_0 \partial_i \), already showing these
are constraint components, as in (15-16). For orientation, we revert to synchronous gauge (here \( k_0 = 0 \)), for which

\[ R^{\text{lin}} = \partial^2 k^T - \nabla^2 k^L. \]

Here \( k^T \) and \( k^L \) are components of the usual ADM “TT” decomposition of a symmetric spatial tensor

\[ k_{ij} = k^{TT}_{ij} + \frac{1}{2} (\delta_{ij} - \partial_i \partial_j / \nabla^2) k^T + \frac{1}{2} (\partial_i k^T_j + \partial_j k^T_i) + \partial_i \partial_j k^L; \]

indeed, \( k^T \) is precisely the Newtonian metric of concern, while \( k^L \) is the doubly longitudinal, pure gauge, term [30]. The rest of the story is of course just the linearization of the results of subsections 4.2 and 4.3.

Now we go to the linearized, but gauge invariant \( R^{\text{lin}} \):

\[ R^{\text{lin}} = \partial^2 k^T - C, \quad C \equiv \nabla^2 k_{00} + \nabla^2 k^L - 2k_{0i,0i}. \]  

(22)

The first, gauge invariant, Newtonian term is unchanged, while the additional (also gauge invariant) combination \( C \) differs from its synchronous gauge value only by lower time derivative terms, so the justifications previously exhibited for that gauge simply carry over unchanged to any frame.

5 Discussion

Our nonlocal model [12] exactly reproduces the ΛCDM expansion history with zero cosmological constant [21]. The model has no current problem either with solar system tests [20] or with existing data on structure formation [23]. The small deviations from general relativity it predicts for structure formation should be resolvable with the next generation of large scale structure surveys [24]. In anticipation, we have considered the theoretical issues of screening and stability.

Our first result is that screening inside any gravitationally bound system can be made 100% effective by simply defining the nonlocal distortion function to vanish for positive argument, which has no affect on the cosmological behavior. Our second result is that the “localization” [26] is inequivalent to ours in that it has extra (scalar) excitations, one of which is unavoidably a ghost. Instead, we saw that the nonlocal model has the same DoF as general relativity; the variables of both separate into identical sets of constrained and radiation excitation modes, subject to the identical initial value constraints. Further, despite the (desired!) difference in the evolution equations, explicit and nonperturbative examination of the highest time derivatives shows that no graviton degree of freedom ever becomes a ghost. That ensures the absence of kinetic energy instabilities. The more difficult issue, ruling out instabilities due to possible negative potential energy excitations, seems even more
difficult here than proving the positive energy theorem in general relativity; while bad modes seem unlikely on physical grounds, we have not attempted to exclude them.

Acknowledgements

We thank S. Dodelson and S. Park for the persistent questions which have prompted this study. And it is a pleasure to acknowledge stimulating conversations with J. Khoury and C. Skordis. This work was partially supported by Department of Energy Grant DE-FG02-16492ER40701, by National Science Foundation grants PHY-1266107 and PHY-1205591, and by the Institute for Fundamental Theory at the University of Florida.

References


