One gravitational potential or two?
Forecasts and tests

BY EDMUND BERTSCHINGER*

Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

The metric of a perturbed Robertson–Walker space–time is characterized by three functions: a scale-factor giving the expansion history and two potentials that generalize the single potential of Newtonian gravity. The Newtonian potential induces peculiar velocities and, from these, the growth of matter fluctuations. Massless particles respond equally to the Newtonian potential and to a curvature potential. The difference of the two potentials, called the gravitational slip, is predicted to be very small in general relativity, but can be substantial in modified gravity theories. The two potentials can be measured, and gravity tested on cosmological scales, by combining weak gravitational lensing or the integrated Sachs–Wolfe effect with galaxy peculiar velocities or clustering.

Keywords: gravitation; cosmological tests; general relativity

1. Introduction

Cosmic acceleration remains as mysterious today as it was when its discovery was announced in 1998 [1,2]. In the cosmological standard model based on general relativity (GR), a homogeneous, isotropic and uniformly expanding universe accelerates if and only if the pressure is large and negative such that \( p + 3\rho < 0 \). Dark energy is the name given to a class of exotic fluids having this property. Dark energy may be either constant (the cosmological constant \( \Lambda \)) or have varying density and pressure (dynamical dark energy). At present, the chief goal of observational cosmology is to measure the expansion history \( H(a) \) in order to determine the time dependence of the energy density \( \rho(a) \) from the Friedmann equation

\[
H^2(a) = \frac{8\pi}{3} G \rho(a) - Ka^{-2}.
\]

Measurements of \( \rho(a) \) yield the dark energy equation of state \( p(\rho) \) through energy conservation, \( d\rho/d\ln a = -3(\rho + p) \).

The conclusion that dark energy exists depends on the assumptions of the cosmological standard model. For example, if the galaxy distribution is strongly inhomogeneous while preserving near-isotropy around us, it is possible for galaxy motions along the past light cone to mimic the acceleration of a homogenous

*edbert@mit.edu

One contribution of 16 to a Theo Murphy Meeting Issue ‘Testing general relativity with cosmology’.
universe without dark energy [3–5]. This possibility requires very special initial conditions that break the fundamental translation-invariance of the standard model; moreover, gravitational instability might convert large radial gradients into large angular gradients that are not seen. The breaking of translation invariance would be problematic for initial conditions generated by inflation, among other challenges. This article will, therefore, retain the assumption of large-scale homogeneity and isotropy present in the standard cosmological model.

There is one key element of the standard cosmological model that is poorly tested: GR itself. Since Einstein’s 1917 paper introduced the cosmological constant, GR has been so fundamental to cosmological models that it is not immediately obvious how to modify it. Clearly, any modifications great enough to eliminate the need for dark energy must preserve the successes of GR in the solar system and in binary pulsars [6]. Thus, the modifications of the field equations or their solution must be scale dependent: negligible on short length scales and appreciable on the Hubble scale.

Measurements of the expansion history alone cannot distinguish GR from alternative theories. An example of a modified gravity theory having cosmic acceleration without dark energy is the braneworld model of Dvali et al. (DGP) [7], implemented in cosmology by Deffayet et al. [8]. This is a higher dimensional theory in which our Universe lives on a brane of three spatial dimensions. The field equations are modified at long wavelengths. In the so-called self-accelerating branch of the theory, the Friedmann equation is modified to become

\[ H(H - H_{\infty}) = \frac{8\pi}{3} G \rho(a) - K a^{-2}, \]

where \( H_{\infty} \) is a constant. At early times \( H > H_{\infty} \). As the universe expands, \( \rho(a) \) decreases so \( H \) decreases until it asymptotically approaches \( H_{\infty} \). A constant Hubble parameter \( H = \frac{d \ln a}{dt} \) implies exponential expansion, the same as a cosmological constant-dominated universe in GR. Such a model has acceleration without dark energy.

The DGP model makes a specific prediction for \( H(a) \) that can be tested and falsified; it appears to be in conflict with observations [9,10]. However, there are other classes of modified gravity theories that are not, and cannot be, ruled out based on the expansion history. In particular, models in which the Einstein–Hilbert action for gravity is modified by an arbitrary function of the Ricci scalar \( R \) can reproduce any \( H(a) \) [11,12]. The action for the theory is

\[ S[g_{\mu\nu}(x)] = \int \frac{d^4x \sqrt{-g}}{16\pi G} [R + f(R)], \]

where \( f(R) = 0 \) for GR. There is no fundamental theory for \( f(R) \); instead, there exist parametrized forms (e.g. [13,14]) for which predictions can be made and compared with observations.

These results imply that GR cannot be tested on cosmological scales using the expansion history. One must look beyond the homogeneous and isotropic, uniformly expanding Robertson–Walker models to include perturbations. Moreover, it is most helpful to identify those features of the data that most distinguish modified gravity theories from GR, and use them to construct cosmological tests. Theorists and observers have begun to do this (e.g. [15,16]).
As we show, these tests involve two functions of space and time, which we call potentials. We also show that observations of galaxies or cold dark matter, in particular the growth of structure, are insufficient to test GR.

What does it mean to test GR? GR makes a number of assumptions that may also hold in alternative theories of gravity [17]:

— space–time is a four-dimensional pseudo-Riemannian manifold;
— special relativity holds locally, in particular, energy–momentum is locally conserved;
— the weak equivalence principle holds, i.e. freely falling bodies follow space–time geodesics; and
— the metric is the solution to the Einstein field equations subject to appropriate initial and boundary conditions; this is uniquely true in GR.

The first three assumptions will be adopted here for modified gravity theories as well as for GR. Modifying gravity then implies modifying the field equations; the fourth assumption is to be tested. Specifically, testing GR requires measuring the metric (or the transfer functions and power spectra for perturbations in the metric) and comparing measurements with the solution of the field equations.

Section 2 introduces the two potentials used to characterize cosmological perturbations and shows that non-relativistic matter (atoms and cold dark matter) are sensitive to only one of them, while the deflection and redshift of photons are sensitive to both potentials. Section 3 introduces the primary tools for measuring the potentials: peculiar velocities, the growth of structure, the integrated Sachs–Wolfe (ISW) effect and gravitational lensing. While full reconstruction of the potentials remains a long-term goal, current efforts emphasize measurements of correlation functions and power spectra, which also provide valuable constraints on the two potentials.

Section 4 summarizes the field equations for the potentials in GR and in one class of modified gravity theories, the $f(R)$ theories. It is shown that a key discriminating feature between theories is the ‘gravitational slip’, the difference of the two potentials. In GR, the slip is generated only by relativistic shear stress and is negligible in the matter-dominated era. In $f(R)$ theories, the slip is generated by a new gravitational field called the scalaron. The scalaron also modifies the growth of structure, while avoiding solar system constraints. Section 5 briefly summarizes current tests.

More extensive reviews are given by Caldwell & Kamionkowski [18], Silvestri & Trodden [19], Uzan [20] and Jain & Khoury [21].

2. Two potentials

The most general metric of a weakly perturbed Robertson–Walker space–time, regardless of whether GR is valid, takes the form

$$\text{d}s^2 = a^2(\tau)\{-\left(1 + 2\Phi\right)\text{d}\tau^2 + 2w_i\text{d}x^i\text{d}\tau + [(1 - 2\Psi)\gamma_{ij} + 2s_{ij}]\text{d}x^i\text{d}x^j\},$$

(2.1)

where $\gamma_{ij}$ is the three-metric of constant-curvature spaces, e.g. $\gamma_{ij} = \delta_{ij}$ for Cartesian coordinates in a Euclidean space. The inverse spatial metric is $\gamma^{ij}$. For an unperturbed Robertson–Walker space–time, $\Phi = \Psi = w_i = s_{ij} = 0$. Without
loss of generality, we can take $s_{ij}$ to be symmetric and traceless, i.e. $\gamma^{ij} s_{ij} = 0$. The time variable $\tau$ is called conformal time and is related to the proper time for comoving observers (those at fixed spatial coordinates) in the unperturbed background by $dt = a d\tau$.

Equation (2.1) is too general in that the same space–time is described by infinitely many sets of solutions for the functions $(\Phi, \Psi, w_i, s_{ij})$. This ambiguity arises because of general covariance, i.e. the invariance of the gravitational action under coordinate transformations. Like the vector potential $A_m$ of electromagnetism, the metric perturbations change under gauge transformations (here, infinitesimal coordinate transformations). We are free to remove the ambiguity by gauge fixing, which means choosing a particular coordinate system. We choose the Poisson gauge [22], a coordinate system defined by transverse gauge conditions

$$\nabla^i w_i = 0 \quad \text{and} \quad \nabla^i s_{ij} = 0,$$

(2.2)

where spatial divergences are covariant derivatives taken with respect to the background three-metric $\gamma_{ij}$. The function $w_i$ is called the gravomagnetic potential and is associated with vector perturbations. The function $s_{ij}$ is called the gravitational wave strain and is associated with tensor perturbations. The Poisson gauge generalizes the transverse-traceless gauge widely used for gravitational waves to include all possible perturbations of the metric (scalar, vector and tensor).

Although gravomagnetic and gravitational wave perturbations are expected to be present in the universe, they have not been directly detected and are known to be much smaller in amplitude than the other terms in the metric. Therefore, we neglect them in what follows.

The potential $\Phi$ is called the Newtonian potential. Being part of the time–time part of the metric, it is important for the motion of non-relativistic particles. For slowly moving particles, $dx^i/d\tau$ is small, and the spatial part of the metric makes a negligible contribution to the proper time and therefore to the equations of motion that extremize proper time. The second potential $\Psi$ is called the curvature potential because it is associated with spatial curvature. It is important for the motion of relativistic particles.

With this coordinate choice and with the neglect of gravomagnetic and gravitational wave perturbations, the metric simplifies to the conformal Newtonian metric [23,24]

$$ds^2 = a^2(\tau)[- (1 + 2\Phi) d\tau^2 + (1 - 2\Psi) \gamma_{ij} dx^i dx^j].$$

(2.3)

The potentials are assumed to be small enough ($\Phi \sim \Psi \sim 10^{-5}$) that their products can be neglected. For a flat background, with $\gamma_{ij} = \delta_{ij}$, and with $\Psi = \Phi$, this metric is conformal to (i.e. identical to aside from an overall multiplicative factor) the weak-field metric used in the Newtonian limit [25]. The conformal Newtonian coordinate system used here has the advantage of making motion look simple. The reader should beware that different choices of names and signs for the two potentials appear in the literature.

Freely falling particles move along geodesics of metric (2.3) and obey equations of motion

$$\frac{dx^i}{d\tau} = (1 + \Phi + \Psi) v^i$$

(2.4a)
and

\[ \frac{1}{\gamma a(1 - \Psi)} \frac{d}{d\tau} [\gamma a(1 - \Psi)v^i] = -\nabla^i(\Phi + v^2\Psi) - (1 + \Phi + \Psi)\gamma^i_{jk}v^jv^k. \tag{2.4b} \]

Here, \( v^i \) is the proper (physical) three-velocity measured by a comoving observer (note that \( \tau \) is not the proper time). Specifically, it is the peculiar velocity with respect to uniform Hubble flow. Other terms in the equations of motion have their usual definitions: \( v^2 \equiv \gamma_{ij}v^iv^j, \gamma \equiv (1 - v^2)^{-1/2} \) is the special relativistic Lorentz factor and \( \gamma^i_{jk} \) is the spatial connection needed when non-Cartesian coordinates are used (giving rise to terms such as centripetal acceleration). The factors \( 1 + \Phi + \Psi \) convert proper velocity to coordinate velocity. Hereafter, we will ignore the approximately \( 10^{-5} \) corrections from proper to coordinate velocities.

The equations of motion simplify in the limits \( v \to 0 \) (non-relativistic particles) and \( v \to 1 \) (massless particles). In the first case, denoted cold dark matter (CDM),

\[ \frac{1}{a} \frac{d}{d\tau} (av) = -\nabla\Phi, \quad v^2 \ll 1 \) (CDM). \tag{2.5} \]

In the second case,

\[ \frac{dv}{d\tau} = -\nabla_{\perp}(\Phi + \Psi), \quad v^2 = 1 \) (photons), \tag{2.6} \]

where the gradient is taken in the plane perpendicular to the photon trajectory,

\[ \nabla_{\perp}^i \equiv \nabla^i - v^i v^j \nabla^j. \tag{2.7} \]

Equations (2.5) and (2.6) are very important and are more general than GR. They have a crucial implication: the dynamics of galaxies and CDM cannot fully test GR because they depend only on one of the two potentials.

Gravitational lensing measurements are sensitive to the sum of the two potentials. Another dependence arises for the redshift of light,

\[ \frac{d\ln(aE)}{d\tau} = \frac{\partial\Psi}{\partial\tau} - v \cdot \nabla\Phi, \quad v^2 = 1 \) (photons). \tag{2.8} \]

The first term on the right-hand side is less familiar than the second term (gravitational redshift). It becomes more familiar when equation (2.8) is integrated to give the accumulated change in energy of a photon between emission at redshift \( z \) and observation,

\[ (1 + z) \frac{E_{\text{obs}}}{E_{\text{em}}} = 1 + \Phi_{\text{em}} - \Phi_{\text{obs}} - \int_0^z (\dot{\Phi} + \dot{\Psi}) \left( \frac{d\tau}{dz} \right) dz, \tag{2.9} \]

where a dot denotes \( \partial/\partial\tau \). Equation (2.9) gives the familiar Sachs–Wolfe effect [26]; the last term, arising from time variation of the potentials along the past light cone, is called the ISW effect. Thus, cosmic microwave background (CMB) anisotropy responds to both the Newtonian potential \( \Phi_{\text{em}} \) and to the time derivative of the sum of the two potentials.
3. Methods to measure the potentials

Testing gravity on cosmological scales requires making distinct measurements for each of the three functions appearing in the metric. The most direct way is:

— measure \( a(\tau) \) using redshift and distance measurements to infer the cosmic expansion history \( H(a) \); integrate \( \frac{d\tau}{da} = \frac{1}{Ha^2} \) to get \( \tau(a) \), which can then be inverted;
— measure \( \Phi(x, \tau) \) using equation (2.5) with the peculiar velocities of galaxies; and
— measure \( (\Phi + \Psi)(x, \tau) \) using equation (2.6) with gravitational lensing or equation (2.9) with the ISW effect.

Although the second two measurements give derivatives of the potentials rather than their absolute values, with appropriate initial or boundary conditions, the potentials can be determined up to irrelevant constants. These measurements are then compared with the relationships between the three functions predicted by GR or modified gravity theories.

In practice, these direct methods are difficult at best. Peculiar velocities are difficult to measure with sufficient precision for gravitational tests; while the radial velocity is easy to measure from redshifts, the Hubble velocity is proportional to distance. Even with the precision given by type Ia supernovae, uncertainties in distance measurements exceed the peculiar velocities for distances beyond a few tens of Megaparsec. Gravitational lensing is promising, however, the deflections are small for structures larger than a few Megaparsec. The ISW signal is concentrated at the lowest multipoles of the CMB where primary anisotropy and cosmic variance (the variance arising from Gaussian random processes with a very small number of terms contributing to the signal) are large.

(a) Growth of structure

To increase the power of galaxies to measure \( \Phi \), cosmologists commonly turn to the growth of density perturbations. Density perturbations are part of the stress–energy tensor, which is decomposed for scalar perturbations as

\[
T^{00}_0 = -(\bar{\rho} + \delta\rho), \\
T^{i0}_i = -(\bar{\rho} + \bar{p}) \nabla_i u \\
T^{ij}_j = \delta_j^i (\bar{\rho} + \delta p) + \frac{1}{2}(\bar{\rho} + \bar{p})(\nabla^i \nabla_j - \frac{1}{3} \delta^i_j \Delta) \pi,
\]

where \( \Delta = \gamma^{ij} \nabla_i \nabla_j \) is the spatial Laplacian, \( \bar{\rho} \) and \( \bar{p} \) are the background density and pressure and \( u \) and \( \pi \) are velocity and shear stress potentials, respectively. The perturbations are gauge dependent. A convenient gauge-invariant density perturbation is the physical number density perturbation in the fluid rest frame [27], obtained from conformal Newtonian gauge variables as

\[
v \equiv \frac{\delta \rho + 3 \mathcal{H}}{\bar{\rho} + \bar{p}} \quad \text{and} \quad \mathcal{H} = \frac{\dot{a}}{a} = aH.
\]
For a non-relativistic fluid on scales much less than the Hubble length with negligible shear stress, linear perturbations of energy–momentum conservation gives

$$\frac{1}{a} \frac{\partial}{\partial \tau} \left( a \frac{\delta \rho}{\partial \tau} \right) = \Delta (\Phi + c_s^2 \nu + \sigma).$$

Equation (3.3) says that the density perturbation growth is driven by the Newtonian potential $F$, it is opposed by adiabatic pressure perturbations through the Jeans term $c_s^2 \nu$, and responds to pressure perturbations produced by entropy perturbations if they are present. It is important to note that equation (3.3) is restricted to linear density perturbations. It may be generalized to incorporate nonlinear perturbations on scales much smaller than the Hubble length, as is commonly done in cosmological simulation codes (see [28] for a review).

If pressure forces are unimportant on linear scales—as is thought to be valid on scales much larger than 10 Mpc, then atomic gas, galaxies and cold dark matter particles all approximately obey equation (3.3) with $c_s^2 = \sigma = 0$. Two key assumptions are implicit:

- all matter couples universally to gravity; there are no additional long-range forces and
- luminous atoms trace dark matter and so do galaxies, or one can correct for ‘biased galaxy formation’ [29,30].

The first assumption involves fundamental physics and is a restatement of the weak equivalence principle. It is important to test this assumption on cosmological scales, but that is beyond the scope of this article [21]. The second assumption is astrophysical and represents a serious concern for cosmological tests.

If these assumptions can be justified, then growth of structure offers an alternative means to determine the Newtonian potential $\Phi$. A measurement of density perturbations $\nu(a)$ substituted into equation (3.3) gives $\Delta \Phi$.

(b) Weak gravitational lensing

Weak gravitational lensing results when the displacements of photon trajectories implied by equation (2.6) accumulate over the past light cone so as to deform the images of distant sources. The angular deflection arising between $z_L$ and $z_L + \delta z_L$ is

$$\delta \alpha = -\nabla_\perp (\Phi + \Psi) \frac{d\tau}{dz_L} \delta z_L,$$

where $\alpha$ is a two-dimensional vector perpendicular to the photon trajectory. The gravitational lens equation [31]

$$\beta = \theta - \frac{\eta_S}{r_S} \delta \alpha$$

Phil. Trans. R. Soc. A (2011)
relates the observed direction of a ray \( \theta \) to the deflection and the direction to the source \( \beta \) in the absence of deflection. The two distances \( r_S \) and \( r_{LS} \) are the comoving angular distances \( r(\chi) \) appearing in the spatial line element \( \gamma_{ij} dx^i dx^j = d\chi^2 + r^2(\chi)(d\theta^2 + \sin^2 \theta d\phi^2) \).

The absolute deflection of rays is unobservable since we have no way of knowing the true direction of sources in the absence of deflection. However, the gradient of deflection results in image magnification and shear that are observable and are described by the inverse magnification matrix

\[
M^{-1} = \frac{\partial \beta}{\partial \theta} = 1 + \int_0^{\infty} \frac{r_{LS}}{r_S} \nabla_\perp \nabla_\perp (\Phi + \Psi') \frac{d\tau}{d\mathbf{z}_L} d\mathbf{z}_L. \tag{3.7}
\]

The two-by-two magnification matrix is parametrized by a convergence \( \kappa \) and shear parameters \( \gamma_1 \) and \( \gamma_2 \),

\[
M = \begin{pmatrix}
1 + \kappa + \gamma_1 & \gamma_2 \\
\gamma_2 & 1 + \kappa - \gamma_1
\end{pmatrix}. \tag{3.8}
\]

This description is most useful in the weak lensing limit \( \kappa^2 \ll 1, \gamma_i^2 \ll 1 \). Although the convergence \( \kappa \) cannot be measured directly because the unlensed size and flux of sources are generally unknown, the shear components \( \gamma_i \) can be measured by averaging galaxy image distortions over fields of view containing hundreds of distant galaxies \[32\]. Thus, weak lensing offers a probe of the two-by-two traceless Hessian matrix of the sum of the two potentials,

\[
\Psi_{ij} = (\nabla_i \nabla_j - \frac{1}{2} \delta_{ij} \Delta_2)(\Phi + \Psi), \tag{3.9}
\]

where \( \Delta_2 = \nabla_\perp \cdot \nabla_\perp \). Given this matrix as a function of direction, and imposing appropriate boundary conditions, it is possible to invert the partial derivative operator to solve for \( \Phi + \Psi \) or the convergence \( \kappa \) \[33\].

\( \text{(c) Power spectra and transfer functions} \)

An ideal test of gravitation theories would determine \( \Phi(\mathbf{x}, \tau) \) and \( \Psi(\mathbf{x}, \tau) \) along the past light cone giving (with the expansion history) the solution to the gravitational field equations over as much of space–time as is accessible to our view. Indeed, reconstruction should be the goal. However, many problems impede this effort: galaxies (the main cosmological probe) are sparsely distributed through space, their properties and small-scale clustering are strongly affected by nonlinear dynamics and the astrophysics of star formation (e.g. \[34\]), and building up complete samples is an ongoing project for the field of observational cosmology. In the case of CMB anisotropy, accurate measurements are possible but the Sachs–Wolfe and ISW effects cannot be separated from other sources of anisotropy generated during recombination and before. Even if they could be separated, the ISW effect is most important at low multi-poles where cosmological theories predict that the power spectrum has a \( \chi^2 \)-distribution with \( 2l + 1 \) d.f. (cosmic variance).

For these reasons, it is useful to measure the statistical properties of the galaxy density and peculiar velocity fields, weak lensing shear and the ISW effect by measuring correlation functions. Correlation functions can be regarded as a lossy data compression technique; they eliminate information about non-Gaussianity, but facilitate tests of the time- and length-scale dependence of fluctuations.
and thereby of the metric potentials. Each fluctuating field (\( \Phi, \Psi, \nu, u, \gamma_i \), etc.) has an autocorrelation with itself and cross-correlations with other fields. The Fourier transform of a spatial correlation function is a power spectrum or cross-power spectrum, which itself is related to a correlation in Fourier space. For a Euclidean background,

\[
\xi_{\nu\nu}(|x_1 - x_2|) \equiv \langle \nu(x_1)\nu(x_2) \rangle = \int \frac{d^3k}{(2\pi)^3} e^{ik(x_1 - x_2)} P_{\nu\nu}(k),
\]

where

\[
\langle \nu(k)\nu(k') \rangle = (2\pi)^3 P_{\nu\nu}(k)\delta^3(k + k').
\]

These equations assume that the fluctuations are statistically homogeneous and isotropic, which holds for quantum fluctuations superposed on a translationally and rotationally invariant background.

The time dependence of the fields and correlations has been suppressed in equations (3.10)–(3.11). In linear perturbation theory, all fluctuation fields are given by convolutions of a primordial scalar field or possibly a superposition of several such convolutions. The simplest category is curvature fluctuations characterized by the gauge-invariant spatial curvature perturbation

\[
\mathcal{R} = \Psi + H u,
\]

where \( u \) is the total velocity potential (a weighted average over all species) defined by equation (3.1b). The field \( \mathcal{R} \) is constant on scales larger than the Hubble and curvature lengths, irrespective of the expansion history [35]. It is even constant on super-horizon scales in modified gravity theories [17].

Every other field is given, in Fourier space, by a rotationally invariant transfer function multiplying this primary variable

\[
\nu(k, \tau) = \nu(k)\mathcal{R}(k).
\]

The transfer functions for all matter and metric variables are given by solutions of the appropriate field equations with initial conditions \( \mathcal{R}(k) = 1 \). The power spectrum of \( \mathcal{R} \) is conventionally made non-dimensional by defining

\[
\Delta^2_{\mathcal{R}}(k) = \frac{4\pi k^3 P_{\mathcal{R}\mathcal{R}}(k)}{(2\pi)^3}.
\]

Wilkinson Microwave Anisotropy Probe 7 year results [36] give \( \Delta_{\mathcal{R}} = (4.93 \pm 0.11) \times 10^{-5} \) for \( k = 0.002 \) Mpc\(^{-1}\).

Cross-correlation techniques can be used to extract information about the ISW effect that is otherwise hidden by other sources of low-multipole CMB anisotropy. Galaxy (mass) density fluctuations at modest redshift correlate with the potential fluctuations giving rise to the ISW effect. The cross-correlation of CMB anisotropy with galaxy density can therefore extract the ISW signal [37,38].

Cross-correlation techniques can also be used to minimize the systematic effects of galaxy bias. Zhang et al. [16] defined a ratio of weak lensing shear to galaxy peculiar velocities based on cross-correlations with the galaxy density. The resulting statistic, \( E_G \), can be measured on large scales using weak lensing combined with redshift-space distortions in galaxy redshift surveys.
4. Potentials in general relativity and in $f(R)$ theories

To test a theory of gravity, test the gravitational field equations. As argued in §1, the field equations for the unperturbed background are insufficient. Therefore, we examine the linear perturbations of the field equations. In GR, the field equations are $G_{\mu}^{\nu} = 8\pi G T_{\mu}^{\nu}$, with first-order perturbations giving

$$ (\Delta + 3K)\Psi = \alpha v, \quad (4.1a) $$
$$ \dot{\Psi} + H\Phi = \alpha u, \quad (4.1b) $$
$$ \Psi - \Phi = \alpha \pi \quad (4.1c) $$

and

$$ \ddot{\Psi} - K\Psi + H(2\dot{\Psi} + \dot{\Phi}) + (2H + H^2)\Phi + \frac{1}{3} \Delta(\Phi - \Psi) = \frac{\alpha \delta p}{\dot{\rho} + \dot{p}}, \quad (4.1d) $$

where

$$ \alpha \equiv 4\pi G a^2 (\bar{\rho} + \bar{p}). \quad (4.2) $$

The right-hand side of equations (4.1) uses the stress–energy perturbations given in equations (3.1) and (3.2). An alternative form of equation (4.1d) is [17]

$$ \frac{\alpha}{H} \frac{\partial}{\partial \tau} \left[ \frac{H^2}{\alpha a^2} \frac{\partial}{\partial \tau} \left( \frac{a^2}{H} \Psi \right) \right] - c_s^2 \Delta \Psi = \alpha S \quad \text{and} \quad S \equiv \sigma + \frac{1}{H} \frac{\partial}{\partial \tau} (H^2 \pi) + \frac{1}{3} \Delta \pi. \quad (4.3) $$

This equation is exact in GR with linear perturbation theory for open, curved or flat backgrounds, and for arbitrary matter fields and dynamics.

If we know nothing about the mass–energy content of the universe, then the Einstein equations provide no possible test; we cannot test GR without making assumptions about the mass–energy content. We can simply measure the potentials and claim there exist matter fields obeying the Einstein field equations. Because of the Bianchi identities, we are even guaranteed that the ‘stress–energy’ tensor inferred this way is locally conserved.

This approach is unreasonable. We aspire to a complete physical characterization of mass–energy in the laboratory as well as the cosmos. Although we do not yet know the composition of the dark matter and dark energy, the most natural models share an important property: the shear stress is small compared with the energy density. For example, for linear perturbations of a scalar field or a perfect fluid, $\pi = 0$. For free-streaming particles, $\Delta \pi \sim c_s^2 \nu$ where $c_s$ is the sound speed. The largest source of shear stress in the standard cosmological model is relativistic neutrinos after neutrino decoupling. On scales larger than the Hubble length, and while the neutrinos are relativistic ([23], with some change of notation),

$$ \Delta \pi = \frac{2}{5} \frac{(\bar{\rho} + \bar{p})_r}{(\bar{\rho} + \bar{p})_{\text{tot}}} \nu, \quad (4.4) $$

leading to [39]

$$ \frac{\Psi - \Phi}{\Phi} = \frac{2}{5} \frac{(\bar{\rho} + \bar{p})_r}{(\bar{\rho} + \bar{p})_{\text{tot}}}. \quad (4.5) $$

This ratio is negligible in the matter-dominated era. In order for shear stress to be gravitationally important, there must exist a relativistic component that dominates the mass–energy density of the universe and which has nearly maximal
free-streaming. Suppose that it is the dark energy. Where such a field is present, $S$ in equation (4.3) would be substantial and would modify the evolution of the gravitational potentials, and thereby would change the evolution of cold dark matter perturbations. The observed structure formation could be achieved only if the shear stress becomes important after galaxies form (e.g. stressed dark energy, [40]). Even in this case, the ISW effect would be large.

The combination $\Psi - \Phi$ is called gravitational slip [41,42]. GR with the standard cosmological parameters predicts negligible gravitational slip today.

In alternative theories of gravity, the slip is non-zero. For example, in $f(R)$ theories [43,44], an additional scalar degree of freedom is present in the gravitational sector, called the scalaron $\chi$ [45]. In equation (1.3), the modification is the term $f(R)$; the scalaron $\chi$ is given by the fluctuations in its derivative,

$$f_R \equiv \frac{df}{dR} \quad \text{and} \quad \chi \equiv \frac{df_R}{dR}\delta R.$$ (4.6)

The scalaron obeys the wave equation of a field with mass

$$m_s^2 = \frac{1 + f_R}{3d f_R/dR}.$$ (4.7)

The field equations (4.1) are modified by the addition of terms proportional to $\chi$ and derivatives of $f_R$. For the present purposes, the most important modification is to the gravitational slip,

$$(1 + f_R)(\Psi - \Phi) = \chi + \alpha \pi.$$ (4.8)

Modified gravity theories, and $f(R)$ in particular, modify the gravitational coupling and they introduce a non-zero gravitational slip independently of the matter sector.

How large is the gravitational slip predicted to be if $\pi = 0$? On scales smaller than the Hubble length, a quasi-static solution to the $f(R)$ field equations exists with [46–48]

$$\chi \approx -\frac{(1 + f_R)k^2}{k^2 + a^2 m_s^2} \left( \Phi + \Psi \right) \frac{3}{3}.$$ (4.9)

In addition, the Poisson equation (4.1a) is modified on scales much smaller than the Hubble length to become

$$\Delta \left( \frac{\Phi + \Psi}{2} \right) = \frac{\alpha \nu}{1 + f_R}.$$ (4.10)

Combining equations (4.8) and (4.9) gives

$$\frac{\Psi}{\Phi} = \frac{2k^2 + 3a^2 m_s^2}{4k^2 + 3a^2 m_s^2}.$$ (4.11)

On such scales, not only is the gravitational slip non-zero, but also the effective gravitational coupling in the standard Poisson equation $\Delta \Phi = 4\pi G_{eff} a^2 \delta \rho$ is
increased to

\[ G_{\text{eff}} = \frac{G}{1 + f_{R}} \left( \frac{4k^2 + 3a^2m_s^2}{3k^2 + 3a^2m_s^2} \right). \]  

(4.12)

On scales much larger than the scalaron Compton wavelength \( m_s^{-1} \), gravity is unmodified, aside from an overall reduction factor \( 1 + f_R \). However, on smaller scales, the gravitational coupling increases by a factor \( 4/3 \) and the gravitational slip becomes \( \Psi - \Phi = -\frac{1}{2} \Phi \).

This is a substantial change in the behaviour of gravity on small scales, suggesting that the \( f(R) \) theories might be ruled out based on solar system tests. However, the scalaron mass and the coupling factor \( 1 + f_R \) depend on curvature and hence on local density. The nonlinearity of the field equations allow for a gravitational ‘chameleon’ effect \([49]\) that shields gravity in the solar system \([13]\) by restricting the rapid change of \( f_R \) to regions of high curvature. It remains to be seen whether star formation and time-dependent stellar evolution are compatible with such modifications of gravity in stars.

5. Current status of tests

Strong limits on the gravitational slip are obtained in the solar system using the Shapiro time delay \([50]\), an effect in the propagation of light that is sensitive to \( \Phi + \Psi \). The round-trip time of signals from the Cassini spacecraft near Saturn as the signals pass close to the Sun is longer than one would expect in flat space–time. The result is an extremely tight limit on the slip \([51]\),

\[ \frac{|\Psi - \Phi|}{\Phi} < 2 \times 10^{-5} \text{ within the solar system.} \]  

(5.1)

On galaxy scales, Schwab et al. \([52]\) have combined stellar dynamics with strong gravitational lensing (e.g. the production of Einstein rings) for lenses in the Sloan Lens Advanced Camera for Surveys to limit the slip on kiloparsec scales,

\[ \frac{\Psi - \Phi}{\Phi} = 0.01 \pm 0.05 \text{ within galaxies on kpc scales.} \]  

(5.2)

On scales of a few to a few tens of Megaparsec, cosmologists have recently been combining structure formation measurements with weak lensing \([15,53,54]\) and lensing plus the ISW effect \([55–57]\). The Reyes et al. \([15]\) measurement was based on the \( E_G \) parameter of Zhang et al. \([16]\), which has the virtue of relative insensitivity to galaxy bias and other systematics. Their results of these tests show consistency to date with GR on length scales up to about 50 Mpc. Future surveys such as Large Synoptic Survey Telescope \([58]\) and Euclid \([59]\) will be able to improve the limits substantially.

The author thanks Alessandra Silvestri for her collaboration and helpful discussion. This work was supported by NSF grant AST-0708501 and the Kavli Institute for Astrophysics and Space Research.

References

One gravitational potential or two?


One gravitational potential or two?


