

PHYSICS OF OUR DAYS

PACS numbers: 04.20.Dw, **04.62.+v**, 98.80.Bp

The null energy condition and its violation

V A Rubakov

DOI: 10.3367/UFNe.0184.201402b.0137

Contents

Abstract. **A brief review is given of scalar field theories with second-derivative Lagrangians yielding second-order field equations. Some of these theories permit solutions that violate the null energy condition but otherwise show no obvious inconsistencies. The use of these theories in constructing cosmological scenarios and in the context of a laboratory-created universe is illustrated with examples.**

1. Introduction

Among various energy conditions discussed in the context of general relativity, the null energy condition (NEC) plays a special role. This condition states that the matter energy–momentum tensor $T_{\mu\nu}$ obeys the inequality [1]

$$T_{\mu\nu} n^\mu n^\nu > 0 \tag{1}$$

for any null (light-like) vector n^μ, i.e., for any vector satisfying $g_{\mu\nu} n^\mu n^\nu = 0$. The reason the NEC is particularly interesting is twofold. First, the NEC is quite robust; we illustrate this point in Section 2. In fact, until rather recently, the common lore was that the NEC could not be violated in a healthy theory, with the possible exception of a scalar field nonminimally coupled to gravity [1]. The developments that refuted this viewpoint are the main emphasis of this mini-review.

Second, the NEC is a crucial assumption of the Penrose singularity theorem [2], valid in general relativity. The theorem assumes that (i) the NEC holds and (ii) the Cauchy hypersurface is noncompact. The theorem states that once there is a trapped surface in space, there will be a singularity in the future. A trapped surface is a closed surface on which outward-pointing light rays actually converge (moving inwards). In a spherically symmetric situation, this means the following. Let R be a coordinate that measures the area of a sphere, $S(R) = 4\pi R^2$. Then the sphere is a trapped surface if R decreases along any future null direction; all light rays emanating from this sphere in this sense move toward its center (see Appendix A for the details). An example is a sphere inside the horizon of a Schwarzschild black hole, or in the case of a contracting, spatially flat homogeneous isotropic Universe, a sphere greater than $|H|^{-1}$ in size, where H is the Hubble parameter. Thus, for matter obeying the NEC, there is always a singularity that is formed inside the black hole horizon, and any contracting universe ends up in a singularity, if its spatial curvature is dynamically negligible (which is often the case). By time reversal, an expanding universe has a singularity in the past. All this is true in classical general relativity; things are different in other classical theories of gravity, and probably very different in quantum gravity.

Inter alia, the Penrose theorem almost forbids, within classical general relativity, a bouncing Universe scenario, in which the Universe contracts at early times, the contraction terminates at some instant of time, and the Universe enters the expansion epoch, which continues today. We show

[1] The case of the cosmological constant, $T_{\mu\nu} = \Lambda g_{\mu\nu}$, is special. We then have $T_{\mu\nu} n^\mu n^\nu = 0$. In what follows, we do not exclude the possibility that the cosmological constant is nonzero, but we assume that some other matter is also present in the system. Our entire argument then stays in force.

V A Rubakov Institute for Nuclear Research,
Russian Academy of Sciences,
prosp. 60-letiya Oktyabrya 7a, 117312 Moscow, Russian Federation
E-mail: rubakov@ms2.inr.ac.ru
Faculty of Physics, Lomonosov Moscow State University,
Leninskie Gory, 119991 Moscow, Russian Federation

Received 15 January 2014
Uspekhi Fizicheskikh Nauk **184** (2) 137 – 152 (2014)
DOI: 10.3367/UFNr.0184.201402b.0137
Translated by V A Rubakov; edited by A M Semikhatov

explicitly that the NEC is crucial for that ban. We consider a homogeneous isotropic Universe with the Friedmann–Lemaitre–Robertson–Walker metric

$$ds^2 = dt^2 - a^2(t)\gamma_{ij}\, dx^i\, dx^j\,, \tag{2}$$

where γ_{ij} is a time-independent metric of the unit 3-sphere (a parameter $\kappa = +1$ is assigned to this case) or the unit 3-hyperboloid ($\kappa = -1$) or the Euclidean 3-dimensional space ($\kappa = 0$). Matter governing the evolution of this universe must also be homogeneous and isotropic, meaning that the only nonvanishing components of the energy–momentum tensor are

$$T_{00} = \rho\,,$$
$$T_{ij} = a^2\gamma_{ij}\,p\,,$$

where ρ and p are the energy density and effective pressure. The (00) and (ij) components of the Einstein equations then give

$$H^2 = \frac{8\pi}{3}\,G\rho - \frac{\kappa}{a^2}\,, \tag{3a}$$

$$2\dot{H} + 3H^2 = -8\pi G p - \frac{\kappa}{a^2}\,, \tag{3b}$$

where $H \equiv \dot{a}/a$ is the Hubble parameter. A combination of these equations determines how it changes with time:

$$\dot{H} = -4\pi G(\rho + p) + \frac{\kappa}{a^2}\,. \tag{4}$$

Now, we choose the null vector n^μ entering Eqn (1) as $n^\mu = (1, a^{-1}v^i)$, where $\gamma_{ij}v^iv^j = 1$, and find that the NEC is equivalent in the cosmological setting to the condition

$$\rho + p > 0\,.$$

Hence, if the second, spatial curvature, term in the right-hand side of Eqn (4) is negative ($\kappa < 0$, open Universe), zero ($\kappa = 0$, spatially flat Universe), or negligible, the Hubble parameter decreases in time. If it is negative (contraction), it remains negative. Therefore, a bouncing Universe is almost impossible. A loophole is that the bounce is possible for a closed Universe ($\kappa = +1$) if the energy density and pressure grow slower than a^{-2} as the Universe shrinks [3].[2] We note that the Penrose theorem does not apply in the last case because the Cauchy hypersurface is compact in a closed Universe (3-sphere).

Applied to the present-day Universe (which is spatially flat to an excellent accuracy), the NEC implies that the Hubble parameter cannot grow today. Observational evidence for the growing Hubble parameter would mean that either dark energy violates the NEC or general relativity is not valid at the present-day cosmological scales. This would, of course, be highly nontrivial.

Another facet of the NEC shows up through the covariant energy–momentum conservation $\nabla_\mu T^{\mu\nu} = 0$. In the cosmological setting, it takes the form

$$\frac{d\rho}{dt} = -3H(\rho + p)\,. \tag{5}$$

[2] According to Eqn (5), for matter with the equation of state $p = w\rho$, this requires $w < -1/3$.

Thus, the NEC implies that the energy density always decreases in an expanding Universe. Modulo the loophole mentioned above, the Penrose theorem states that the expansion starts from a singularity (infinite energy density, infinite expansion rate).

One more consequence of the NEC is an obstruction to the creation of a universe in the laboratory. The question of whether one can *in principle* create a universe in the laboratory was raised in [4–6] soon after the invention of the inflation theory [7–12]. Indeed, inflation — nearly exponential expansion of the Universe at a high expansion rate — is capable of stretching, in a fraction of a second, a tiny region of space into a region of a huge size, possibly exceeding the size of the presently observable Universe. It therefore appears at first sight that it is not impossible to artificially create a region in our present Universe in which the physical conditions are similar to those at the onset of inflation, and then this region would automatically expand to a very large size and become a universe like ours. In theories obeying the NEC and within classical general relativity, this is impossible [6, 13] because of the Penrose theorem. By definition, a universe 'like ours' is a nearly homogeneous patch in space whose size exceeds the Hubble distance H^{-1}. The Hubble sphere is then an anti-trapped surface, and hence there had to be a singularity in the past. Because we cannot create an appropriate singularity (and control the evolution through any singularity), we cannot create a universe 'like ours'. Widely discussed ways out are to invoke tunneling [14–23] or other quantum effects [24–27] and modify gravity [28–30], but it is certainly of interest to stay within general relativity and invoke NEC violation instead. There have been several attempts in this last direction [31–34], but many of them are problematic because of instabilities.

Finally, the NEC also forbids the existence, within general relativity, of throats in space, both static [35–37] and time-dependent [38]. Such a throat could join asymptotically flat regions of space, forming a Lorentzian wormhole [35–37, 39, 40] (Fig. 1). Alternatively, it could serve as a bridge between a large but finite region of space and an asymptotically flat region, forming a semiclosed world [24] (Fig. 2). Again, it is of

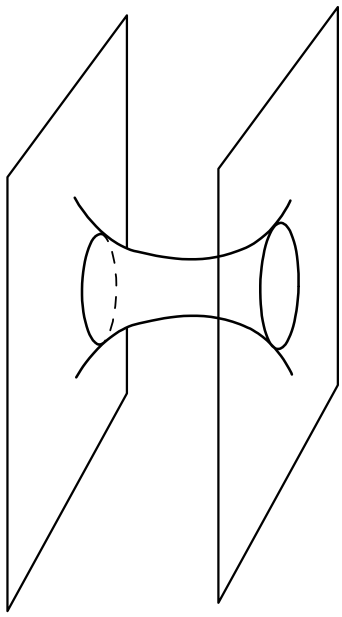

Figure 1. Spatial geometry of a Lorentzian wormhole.

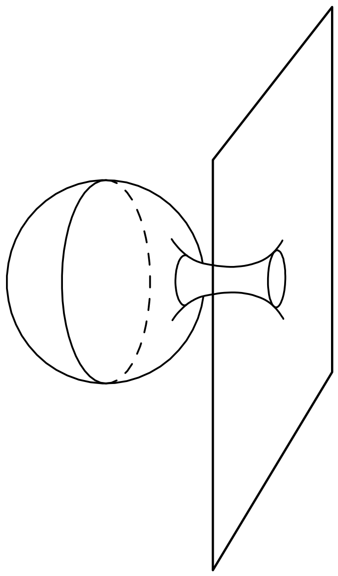

Figure 2. Spatial geometry of a semiclosed world.

interest to construct healthy NEC-violating theories possessing wormhole solutions.

All this motivates the search for healthy NEC-violating theories. For a field theorist, it is natural to start with scalar field theories. However, as we discuss in Section 2, solutions in theories of scalar fields minimally coupled to gravity and described by Lagrangians containing only first derivatives either obey the NEC or have pathologies (modulo a loophole, which we briefly discuss in Section 2). Because of that, one either turns to vector fields (and, indeed, there are examples of acceptable NEC-violating solutions in rather contrived theories involving vector, but not gauge, fields [41, 42]) or considers higher-derivative Lagrangians. It is commonly believed, however, that theories with Lagrangians containing second and higher-order derivatives are unacceptable (unless the higher-derivative terms are treated as perturbations in the sense of an effective low-energy theory) because their field equations involve more than two derivatives and hence these theories have pathological degrees of freedom. This is not the case, however: there is a class of scalar field theories with second-derivative Lagrangians and second-order field equations. These theories were found in an unnoticed paper by Horndeski [43], rediscovered in a rather different context by Fairlie, Govaerts, and Morozov [44–46] and relatively recently became popular in their various reincarnations, such as the Dvali–Gabadadze–Porrati [47] model in the decoupling limit [48, 49], Galileon theory [50] and its generalizations [51–56], k-mouflage [57], kinetic gravity brading [58–60], Fab-Four [61, 62], etc. As we discuss in Section 3, there are finitely many classes of second-derivative Lagrangians yielding second-derivative field equations [50, 63–65]. At least some of these Lagrangians allow NEC violation [66–71], with the NEC-violating solutions and their neighborhoods being perfectly healthy. In Section 4, we give a few examples of using these theories to construct fairly nontrivial cosmological models and to give a proof-of-principle construction for creating a universe in the laboratory.

We conclude in Section 5 by pointing out some potentially problematic features of the NEC-violating second-derivative theories that have yet to be understood.

2. NEC violation and instabilities

2.1 Tachyons, gradient instabilities, and ghosts

In this paper, we are mainly interested in the weak gravity regime, which occurs when $M_{\rm Pl} = G^{-1/2}$ is the largest parameter in the problem. To the lowest order, this corresponds to switching off the dynamics of the metric and considering other fields in the Minkowski background. In many cases, the relevant solutions of the field equations are spatially homogeneous, and we stick to this case here. We ask whether the NEC can be violated in this situation.

In a theory of one scalar field π, a spatially homogeneous classical solution $\pi_{\rm c}(t)$ may or may not be pathological. The pathology, if any, shows up in the behavior of small perturbations about this background, $\pi = \pi_{\rm c} + \chi$. Assuming that the linearized field equation for χ is of the second order in derivatives, the quadratic Lagrangian for χ is always given by

$$L_\chi^{(2)} = \frac{1}{2}\, U\dot{\chi}^2 - \frac{1}{2}\, V(\partial_i\chi)^2 - \frac{1}{2}\, W\chi^2\,, \qquad (6)$$

where U, V, and W depend on time. We consider the high-momentum regime, meaning that variations of χ in space and time occur at scales much shorter than the time scale characteristic of the background $\pi_{\rm c}(t)$. Then, at a given time, the time dependence of U, V, and W can be neglected, and the following possibilities exist:

(1) Stable background,

$$U > 0\,, \qquad V > 0\,, \qquad W \geqslant 0\,.$$

The dispersion relation is

$$U\omega^2 = V\mathbf{p}^2 + W\,, \qquad (7)$$

which is the dispersion relation for conventional excitations, while the energy density for perturbations

$$T_{00}^{(2)} = \frac{1}{2}\, U\dot{\chi}^2 + \frac{1}{2}\, V(\partial_i\chi)^2 + \frac{1}{2}\, W\chi^2 \qquad (8)$$

is positive, as it should be. For $V < U$, the χ-waves travel at subluminal speed; for $V = U$, they travel at the speed of light, while for $V > U$, the χ-waves are superluminal. While superluminal propagation is probably less of a problem, it does signal that the theory cannot be UV-completed in a Lorentz-invariant way [72] (meaning that it cannot be a low-energy theory of some Lorentz-invariant quantum theory valid at all scales) (see, however, Ref. [73], which debates this point). We would therefore like to avoid superluminality. The case $U = V$ is also potentially problematic, since there may or may not be backgrounds *in the neighborhood* of $\pi_{\rm c}$ on which the perturbations are superluminal. Hence, the safe case is

$$U > V > 0\,.$$

(1a) Special case:

$$U > 0\,, \qquad V = 0\,.$$

To understand how to treat this case, we think of the original scalar theory as an effective field theory with a UV cutoff Λ. The Lagrangian of such a theory generically has corrections of a higher order in derivatives, which are suppressed by powers of Λ^{-1} and are therefore normally negligible. For

$V = 0$, however, these corrections cannot be neglected, because only these corrections give the terms in the Lagrangian for perturbations that involve spatial gradients [74]. The dominant higher-derivative terms in the Lagrangian for perturbations involve second-order derivatives, and hence the Lagrangian is given by [3]

$$L_\chi^{(2)} = \frac{1}{2}\, U\dot{\chi}^2 + \frac{1}{2\Lambda^2}\left[a\ddot{\chi}^2 + b\dot{\chi}^2(\partial_i\chi)^2 + c(\partial_i\partial_i\chi)^2\right],$$

where we set $W = 0$ for simplicity, as in the ghost condensate theory [74]. The dispersion relation, modulo corrections suppressed by Λ^{-1} even more strongly, is now

$$U\omega^2 = \frac{c}{\Lambda^2}\,\mathbf{p}^4\,,$$

which is healthy for $c > 0$. Other roots of the dispersion equation obey $|\omega| \gg \Lambda$, and hence cannot be trusted in the low-energy effective theory.

(1b) Tachyonic instability:

$$U > 0\,, \qquad V > 0\,, \qquad W < 0\,.$$

Formally, dispersion relation (7) yields imaginary ω for sufficiently *low* momenta, $V\mathbf{p}^2 < |W|$, and hence there are growing perturbations $\chi \propto \exp\left(\int|\omega|\,\mathrm{d}t\right)$ with $|\omega| \leqslant |W|^{1/2}$. This is indeed a problem, if the time scale $|W|^{-1/2}$ is much shorter than the time scale characteristic of the background $\pi_c(t)$. In the opposite case, we cannot use the approximation of slowly varying $U(t)$, $V(t)$, and $W(t)$ and hence cannot conclude that the background π_c is unstable. Instead, the background is stable at short time scales, and to see what is going on at long time scales, we have to perform a full stability analysis. We note in passing that tachyonic instabilities are inherent in some NEC-violating models of dark energy [41, 42], and they may have interesting observational consequences [75, 76].

(2) Gradient instability:

$$U > 0\,, \quad V < 0 \quad \text{or} \quad U < 0\,, \quad V > 0\,.$$

According to Eqn (7), the 'frequencies' $\omega(\mathbf{p})$ are imaginary at high momenta, and there are perturbations that grow arbitrarily fast. This means that the background π_c is unstable, and hence not healthy. Considering the original scalar theory as an effective low-energy theory valid below a certain UV scale Λ does not help: for consistency, the rate of variation of the background $\pi_c(t)$ must be well below Λ, while the rates of development of the instabilities extend up to Λ; the background is ruined at short time scales.

(3) Ghost instability:

$$U < 0\,, \quad V < 0\,.$$

In *classical* field theory, the background is stable against high-momentum perturbations: Eqn (7) shows that the frequencies are real at high momenta. Yet the background is *quantum mechanically* unstable. Indeed, energy (8) is negative at high momenta, and the χ-particles acquire negative energies upon quantization; they are ghosts. Energy conservation does not forbid pair creation from the vacuum of ghosts together with other, normal particles (say, via graviton exchange, since gravitons definitely interact with χ-quanta); the vacuum is quantum mechanically unstable. Energies and momenta of the created particles can take values up to the UV scale Λ below which one can trust the theory, and hence the available phase space is generically large, and the time scale of the instability is short. Unless Λ is low enough, this instability is unacceptable. Therefore, backgrounds with ghosts are generally considered to be pathological. We note in passing that in a Lorentz-invariant theory and for the Lorentz-invariant background $\pi_c = \text{const}$, the ghost instability is truly catastrophic: if particles can be created from the vacuum with some energies and momenta, then the same, but Lorentz-boosted, process is also allowed; the available phase space is proportional to the volume of the Lorentz group, i.e., it is infinite; the time scale of the instability is infinitesimally short. Put differently, ghosts in the present Universe are allowed only if Lorentz invariance is violated in the ghost sector in such a way that the energies of ghost particles cannot exceed 3 MeV [77].

The above discussion is straightforwardly generalized to a theory with several scalar fields π^I, $I = 1, \ldots, N$. The Lagrangian for perturbations χ^I is now

$$L_\chi^{(2)} = \frac{1}{2}\, U_{IJ}\,\dot{\chi}^I\dot{\chi}^J - \frac{1}{2}\, V_{IJ}\,\partial_i\chi^I\partial_i\chi^J - \frac{1}{2}\, W_{IJ}\,\chi^I\chi^J\,, \qquad (9)$$

and the energy density is

$$T_{00}^{(2)} = \frac{1}{2}\, U_{IJ}\,\dot{\chi}^I\dot{\chi}^J + \frac{1}{2}\, V_{IJ}\,\partial_i\chi^I\partial_i\chi^J + \frac{1}{2}\, W_{IJ}\,\chi^I\chi^J\,.$$

Barring the case of a degenerate matrix V_{IJ}, similar to (1a) above, the matrix V_{IJ} can be diagonalized by a field redefinition. If it has negative eigenvalue(s), the energy is unbounded from below [78]: we can construct an initial configuration with $\dot{\chi}^I = 0$ with an arbitrarily high momentum and $\mathbf{p}^2 V_{IJ}\chi^I\chi^J < 0$. This is a pathological situation: there are either ghosts or gradient instabilities, or both. For a positive definite diagonal V_{IJ}, we can rescale χ^I to transform V_{IJ} into a unit matrix, $V_{IJ} = \delta_{IJ}$. We can then diagonalize U_{IJ} by an orthogonal transformation, and the derivative terms in the Lagrangian become $\sum_I[\lambda_I(\dot{\chi}^I)^2 - (\partial_i\chi^I)^2]$. If U_{IJ} has negative eigenvalues λ_I, there are gradient instabilities. Hence, the requirement of the absence of gradient instabilities and ghosts gives the necessary condition

$$\text{Stable background:} \quad \text{positive definite } U_{IJ} \text{ and } V_{IJ}\,. \qquad (10)$$

Whether there are tachyons at sufficiently low momenta depends now on the positive definiteness of W_{IJ}.

2.2 Scalar theories with first-derivative Lagrangians

The first attempt to construct a NEC-violating theory is to consider the Lagrangian involving first derivatives only,

$$L = F(X^{IJ}, \pi^I)\,, \qquad (11)$$

where

$$X^{IJ} = \partial_\mu\pi^I\,\partial^\mu\pi^J\,.$$

[3] A possible lower-derivative term $(\alpha(t)/\Lambda)\dot{\chi}(\partial_i\chi)^2$, upon integration by parts, reduces to $(\dot{\alpha}/2\Lambda)(\partial_i\chi)^2$; the pertinent transformation is $(\alpha/\Lambda)\dot{\chi}(\partial_i\chi)^2 \to -(\alpha/\Lambda)\,\partial_i\dot{\chi}\,\partial_i\chi = -(\alpha/2\Lambda)\,\partial_0(\partial_i\chi)^2 \to (\dot{\alpha}/2\Lambda)(\partial_i\chi)^2$. It is subdominant at $\mathbf{p}^2 \gg \dot{\alpha}\Lambda$, but becomes relevant at lower momenta. It is healthy for $\dot{\alpha} < 0$.

If we assume minimal coupling to gravity, then the energy–momentum tensor for this theory is

$$T_{\mu\nu} = 2\,\frac{\partial F}{\partial X^{IJ}}\,\partial_\mu \pi^I \partial_\nu \pi^J - g_{\mu\nu} F\,.$$

Therefore, for a homogeneous background,

$$T_{00} \equiv \rho = 2\,\frac{\partial F}{\partial X^{IJ}}\,X^{IJ} - F\,,$$

$$T_{11} = T_{22} = T_{33} \equiv p = F\,,$$

and

$$\rho + p = 2\,\frac{\partial F}{\partial X^{IJ}}\,X^{IJ} = 2\,\frac{\partial F}{\partial X^{IJ}}\,\dot\pi^I \dot\pi^J\,. \qquad (12)$$

We see that NEC violation requires that the matrix $\partial F/\partial X_{\rm c}^{IJ}$, evaluated for the background $\pi_{\rm c}^I$ be nonpositive definite. On the other hand, we expand Lagrangian (11) to the second order in perturbations, $\pi^I = \pi_{\rm c}^I + \chi^I$, and obtain the Lagrangian for perturbations in form (9) with

$$U_{IJ} = 2\,\frac{\partial F}{\partial X_{\rm c}^{IJ}} + 4\,\frac{\partial^2 F}{\partial X_{\rm c}^{IK}\,\partial X_{\rm c}^{JL}}\,\dot\pi_{\rm c}^K \dot\pi_{\rm c}^L\,,$$
$$V_{IJ} = 2\,\frac{\partial F}{\partial X_{\rm c}^{IJ}}\,. \qquad (13)$$

Hence, the stability of the background — positive definiteness of V_{IJ} [see Eqn (10)] — is inconsistent with NEC violation [78].

A loophole here is related to case (1a) above [79]. To see this, we consider a ghost condensate theory with a small potential added [79, 80],

$$L = M^4 (X^2 - 1)^2 - V(\pi)\,,$$

where π is the ghost condensate field [of dimension (mass)$^{-1}$], $X = \partial_\mu \pi\, \partial^\mu \pi$, and M is the energy scale. In the absence of the potential, there is the solution $\pi_{\rm c} = t$, for which $F \equiv M^4 (X^2 - 1)^2 = 0$ and $\partial F/\partial X = 0$. This is on the borderline of NEC violation. A higher-derivative term of an appropriate sign renders this background stable. Now, upon adding a small potential $V(\pi)$ with a positive slope, we make $(\dot\pi_{\rm c} - 1)$ slightly negative. According to Eqns (12) and (13), this leads to NEC violation, and at the same time to the gradient instability. However, with the higher-derivative terms present, that instability occurs at low momenta $\mathbf{p}$ only, and can be made harmless [80] by a careful choice of parameters and of the form of higher-derivative corrections. This construction was used in Ref. [80], in particular, to design a viable cosmological scenario similar to what is now called Genesis. We discuss a less contrived Genesis model in Section 4. Also, the ghost condensate idea was used to construct consistent bouncing Universe models [81, 82], which start from the ekpyrotic contraction stage [83, 84]. Again, the consistency of the bounce requires a careful choice of parameters in these models. We consider a simpler version of this scenario in Section 4.

3. Second-derivative Lagrangians

The main emphasis of this mini-review is on scalar field theories with Lagrangians involving second-order derivatives, whose equations of motion do not contain third or fourth-order derivatives. Although the nomenclature has not yet been settled, we call them (generalized) Galileons. We concentrate on theories of one scalar field π in Minkowski space and write the Euler–Lagrange equation for a theory with the Lagrangian $L(\pi, \partial_\mu\pi, \partial_\mu\partial_\nu\pi)$:

$$\frac{\partial L}{\partial \pi} - \partial_\mu\,\frac{\partial L}{\partial \pi_\mu} + \partial_\mu \partial_\nu\,\frac{\partial L}{\partial \pi_{\mu\nu}} = 0\,, \qquad (14)$$

where

$$\pi_\mu = \partial_\mu \pi\,, \qquad \pi_{\mu\nu} = \partial_\mu \partial_\nu \pi\,.$$

Because of the last term in Eqn (14), the field equation is generically of the fourth order in derivatives. However, there are exceptions, which are precisely Galileons. The simplest exceptional second-derivative Lagrangian is

$$L_{(1)} = K^{\mu\nu}(\pi, \partial_\lambda \pi)\,\partial_\mu \partial_\nu \pi\,. \qquad (15)$$

It would seem that the corresponding field equation is of the third order, but in fact it is not. Indeed, the second term in Eqn (14) gives rise to the third-order contribution

$$-\frac{\partial K^{\mu\nu}}{\partial \pi_\lambda}\,\partial_\lambda \partial_\mu \partial_\nu \pi\,, \qquad (16)$$

while the third term in Eqn (14) is

$$\partial_\mu \partial_\nu K^{\mu\nu}(\pi, \pi_\lambda) = \frac{\partial K^{\mu\nu}}{\partial \pi_\lambda}\,\partial_\mu \partial_\nu \partial_\lambda \pi + \ldots\,, \qquad (17)$$

where the omitted terms do not contain third derivatives. Hence, third-order terms cancel, and the field equation is of the second order.

It is instructive to make the following observation. There seem to exist two terms of the general form (15) with different Lorentz structures:

$$K(\pi, X)\,\Box\pi \quad \text{and} \quad H(\pi, X)\,\partial^\mu \pi\, \partial^\nu \pi\, \partial_\mu \partial_\nu \pi\,,$$

where $\Box = \partial_\lambda \partial^\lambda$ and, as before, $X = \partial_\lambda \pi \partial^\lambda \pi$. But the second structure can be reduced to the first one by integrating by parts (which we denote by an arrow):

$$H(\pi, X)\,\partial^\mu \pi\, \partial^\nu \pi\, \partial_\mu \partial_\nu \pi = \frac{1}{2}\,H\,\partial^\mu \pi\, \partial_\mu X$$
$$= \frac{1}{2}\,\partial_\mu Q\, \partial^\mu \pi \rightarrow -\frac{1}{2}\,Q\,\Box\pi\,,$$

where the function $Q(\pi, X)$ is such that $H = \partial Q/\partial X$. Hence, the only remaining term in the Lagrangian is

$$L_{(1)} = K_1(\pi, X)\,\partial_\mu \partial^\mu \pi\,. \qquad (18)$$

We note that this term cannot be reduced by integration by parts to any Lagrangian involving first derivatives only.

We consider a more complicated example of the Lagrangian quadratic in second-order derivatives. There are five possible Lorentz structures:

$$L_{(2)} = F_1\,\partial^\mu \pi\, \partial^\nu \pi\, \partial^\lambda \pi\, \partial^\rho \pi\, \partial_\mu \partial_\nu \pi\, \partial_\rho \partial_\lambda \pi$$
$$+ F_2\,\partial^\mu \pi\, \partial^\nu \pi\, \partial_\mu \partial_\lambda \pi\, \partial_\nu \partial^\lambda \pi$$
$$+ F_3\,\partial^\mu \pi\, \partial^\nu \pi\, \partial_\mu \partial_\nu \pi\, \Box\pi$$
$$+ F_4\,\partial_\mu \partial_\nu \pi\, \partial^\mu \partial^\nu \pi$$
$$+ F_5 (\Box\pi)^2\,, \qquad (19)$$

where $F_a = F_a(\pi, X)$, $a = 1, \ldots, 5$. The resulting field equation has the following *fourth-order* terms:

$$\begin{aligned} & F_1\, \partial^\mu\pi\, \partial^\nu\pi\, \partial^\lambda\pi\, \partial^\rho\pi\, \partial_\mu\partial_\nu\partial_\rho\partial_\lambda\pi \\ & + F_2\, \partial^\mu\pi\, \partial^\nu\pi\, \partial_\mu\partial_\nu\Box\pi \\ & + F_3\, \partial^\mu\pi\, \partial^\nu\pi\, \partial_\mu\partial_\nu\Box\pi \\ & + F_4\Box\Box\pi \\ & + F_5\Box\Box\pi\,. \end{aligned} \tag{20}$$

We see that the fourth order terms cancel if $F_1 = 0$, $F_2 = -F_3$, and $F_4 = -F_5$, such that the Lagrangian takes the form

$$\begin{aligned} L_{(2)} &= H\,\partial^\mu\pi\,\partial^\nu\pi(\partial_\mu\partial_\nu\pi\,\partial_\lambda\partial^\lambda\pi - \partial_\mu\partial_\lambda\pi\,\partial_\nu\partial^\lambda\pi) \\ &\quad + K(\partial^\nu\partial_\nu\pi\,\partial_\mu\partial^\mu\pi - \partial^\nu\partial_\mu\pi\,\partial_\nu\partial^\mu\pi) \\ &= H\,\partial^\mu\pi\,\partial^\nu\pi\,\partial_\mu\partial_{[\nu}\pi\partial_{\lambda]}\partial^\lambda\pi + K\,\partial^\nu\partial_{[\nu}\pi\partial_{\mu]}\partial^\mu\pi\,, \end{aligned} \tag{21}$$

where square brackets denote antisymmetrization (our definition is $A_{[\mu\nu]} = A_{\mu\nu} - A_{\nu\mu}$ without the numerical prefactor). We now understand the reason for the cancelation of the fourth-order terms in the field equation: it occurs because, e.g., $\partial_\mu\partial_{[\nu}\partial_{\lambda]}\partial^\mu\pi = 0$. Again, the first term in (21) can be transformed to the form of the second term:

$$\begin{aligned} & H\,\partial^\mu\pi\,\partial^\nu\pi(\partial_\mu\partial_\nu\pi\,\partial_\lambda\partial^\lambda\pi - \partial_\mu\partial_\lambda\pi\,\partial_\nu\partial^\lambda\pi) \\ &\qquad = \frac{1}{2}\,H(\partial_\mu X\,\partial^\mu\pi\Box\pi - \partial_\lambda X\,\partial_\mu\pi\,\partial^\lambda\partial^\mu\pi) \\ &\qquad = \frac{1}{2}\,\partial_\mu Q(\partial^\mu\pi\Box\pi - \partial_\nu\pi\,\partial^\nu\partial^\mu\pi) \\ &\qquad \to -\frac{1}{2}\,Q(\Box\pi\Box\pi - \partial_\nu\partial_\mu\pi\,\partial^\nu\partial^\mu\pi)\,. \end{aligned}$$

Hence, there again remains one term,

$$L_{(2)} = K_2(\pi, X)(\partial^\nu\partial_\nu\pi\,\partial_\mu\partial^\mu\pi - \partial^\nu\partial_\mu\pi\,\partial_\nu\partial^\mu\pi)\,.$$

It is now straightforward to verify that the third-order terms in the field equation also cancel: the terms with $(\partial K_2/\partial\pi_\lambda)\,\partial_\lambda(\partial^\nu\partial_\nu\pi\,\partial_\mu\partial^\mu\pi - \partial^\nu\partial_\mu\pi\,\partial_\nu\partial^\mu\pi)$ cancel automatically in the same way as in Eqns (16) and (17), while the remaining terms like

$$\partial^\nu K_2(\pi, X)(\partial_\nu\partial_\mu\partial^\mu\pi - \partial_\mu\partial_\nu\partial^\mu\pi)$$

also vanish.

The story repeats itself in cubic and higher orders in the second derivatives. The only exceptional nth-order term in D dimensions is [50, 63]

$$L_{(n)} = K_n(\pi, X)\,\partial^{\mu_1}\partial_{[\mu_1}\pi\ldots\partial^{\mu_n}\partial_{\mu_n]}\pi\,. \tag{22}$$

That the corresponding field equation is of the second order is verified trivially; the proof that no other terms exist is not so simple [63]. We note that $L_{(n)}$ can be written as

$$\begin{aligned} L_{(n)} &= \frac{1}{(D-n)!} \\ &\times K_n(\pi, X)\epsilon^{\nu_1\ldots\nu_{D-n}\mu_1\ldots\mu_n}\,\epsilon_{\nu_1\ldots\nu_{D-n}\lambda_1\ldots\lambda_n}\partial_{\mu_1}\partial^{\lambda_1}\pi\ldots\partial_{\mu_n}\partial^{\lambda_n}\pi\,. \end{aligned} \tag{23}$$

Indeed, any antisymmetric tensor $A_{\mu_1\ldots\mu_n}$ can be written as

$$A_{\mu_1\ldots\mu_n} = \epsilon_{\nu_1\ldots\nu_{D-n}\,\mu_1,\ldots,\mu_n}B^{\nu_1\ldots\nu_{D-n}}\,, \tag{24}$$

where

$$B^{\nu_1\ldots\nu_{D-n}} = \frac{1}{n!(D-n)!}\,\epsilon^{\nu_1\ldots\nu_{D-n}\,\mu_1,\ldots,\mu_n}A_{\mu_1\ldots\mu_n} \tag{25}$$

is the dual tensor. The expression in the right-hand side of Eqn (22) is antisymmetric in both upper and lower indices. Applying transformations (24) and (25) to the upper indices and lower indices separately, we arrives at the form in (23).

We note that there are $D+1$ allowed classes of Lagrangians, if we count the class without second derivatives,

$$L_{(0)} = K_0(\pi, X)\,. \tag{26}$$

In particular, there are five classes in four dimensions. A general Galileon Lagrangian is a sum of all these terms.

This completes the discussion of the exceptional theories of one scalar field, Galileon, in Minkowski space. Theories with *multiple* scalar fields are considered in Refs [64, 65] (see also Refs [85–88]). The minimal generalization of $L_{(1)}$ to a *curved* space–time is simple:

$$L_{(1)} = K_1(\pi, X)\nabla_\mu\nabla^\mu\pi\,,$$

where $X = g^{\mu\nu}\partial_\mu\pi\,\partial_\nu\pi$; it is straightforward to verify that the resulting field equation is still of the second order. The energy–momentum tensor, and hence the Einstein equations, are also of the second order in derivatives. The generalizations of $L_{(2)}$ and higher-order Lagrangians are, on the other hand, nontrivial [43, 51, 63]. Finally, we note that some Galileon Lagrangians have an interesting interpretation as describing a three-brane evolving in five-dimensional space–time [52, 53].

4. Examples of NEC-violation

In this section, we consider an example of a simple NEC-violating solution and its use for constructing rather nontrivial cosmological scenarios. We also discuss the possibility of creating a universe in the laboratory by using the Galileon models in Section 3. Our set of illustrations is, of course, personal and by no means complete.

4.1 Rolling background

The analysis is particularly simple in models exhibiting scale invariance

$$\pi(x) \to \pi'(x) = \pi(\lambda x) + \ln\lambda\,. \tag{27}$$

It is sufficient for our purposes to consider the Lagrangian involving only the terms $L_{(0)}$ and $L_{(1)}$ [see Eqns (26) and (18)]. In the scale-invariant case and in Minkowski space, we write

$$L_\pi = F(Y)\,\mathrm{e}^{4\pi} + K(Y)\Box\pi\,\mathrm{e}^{2\pi}\,, \tag{28}$$

where

$$Y = \mathrm{e}^{-2\pi}(\partial\pi)^2\,, \qquad (\partial\pi)^2 \equiv \partial_\mu\pi\,\partial^\mu\pi\,, \tag{29}$$

and the functions F and K are not yet specified. Assuming that K is analytic near the origin, we set

$$K(Y=0) = 0\,. \tag{30}$$

Indeed, upon integrating by parts, the constant part of K can be absorbed into the F-term in Eqn (28). We need the expression for the energy–momentum tensor. For this, we consider minimal coupling to the metric, i.e., set $Y = e^{-2\pi} g^{\mu\nu}\partial_\mu\pi\,\partial_\nu\pi$ and $\Box\pi = \nabla^\mu\nabla_\mu\pi$ in a curved space–time. To calculate the energy–momentum tensor, we note that in a curved space–time, the K-term in $\sqrt{-g}\,L_\pi$ can be written, by integrating by parts, as $\sqrt{-g}\,g^{\mu\nu}\,\partial_\mu\pi\,\partial_\nu(K e^{2\pi})$. Then the variation with respect to $g^{\mu\nu}$ is straightforward, and we obtain

$$T_{\mu\nu} = 2F' e^{2\pi}\partial_\mu\pi\,\partial_\nu\pi - g_{\mu\nu}F e^{4\pi} + 2\Box\pi\,K'\partial_\mu\pi\,\partial_\nu\pi - \partial_\mu\pi\,\partial_\nu\big(K e^{2\pi}\big) - \partial_\nu\pi\,\partial_\mu\big(K e^{2\pi}\big) + g_{\mu\nu}g^{\lambda\rho}\partial_\lambda\pi\,\partial_\rho\big(K e^{2\pi}\big)\,.$$

This expression is also valid in curved space–times.

In what follows, we consider homogeneous backgrounds in Minkowski space, $\pi = \pi(t)$. For a homogeneous field, the field equation is

$$4e^{4\pi}F + F'e^{2\pi}(-6\dot\pi^2 - 2\ddot\pi) - 2e^{2\pi}\dot\pi F''\dot Y + K e^{2\pi}(4\dot\pi^2 + 4\ddot\pi) + 4e^{2\pi}\dot\pi K'\dot Y + K''\dot Y(-2\dot\pi^3) + K'(-12\dot\pi^2\ddot\pi + 4\dot\pi^4) = 0\,, \qquad (31)$$

while the energy density and pressure are

$$\rho = e^{4\pi} Z\,, \qquad (32a)$$

$$p = e^{4\pi}\big(F - 2YK - e^{-2\pi}K'\dot\pi\dot Y\big)\,, \qquad (32b)$$

where

$$Z = -F + 2YF' - 2YK + 2Y^2K'\,.$$

It is easy to see that for $\dot\pi \neq 0$, Eqn (31) is equivalent to the energy conservation condition $\dot\rho = 0$.

It is instructive to calculate the quadratic Lagrangian for perturbations near the homogeneous background. It has the form (6) with

$$\frac{1}{2}\,U = e^{2\pi_c}(F' + 2YF'' - 2K + 2YK' + 2Y^2K'') = e^{2\pi_c}Z'\,, \qquad (33a)$$

$$\frac{1}{2}\,V = e^{2\pi_c}(F' - 2K + 2YK' - 2Y^2K'') + (2K' + 2YK'')\ddot\pi_c\,. \qquad (33b)$$

We do not need the general expression for W. We note that U is proportional to the derivative with respect to Y of the same function Z that determines the energy density [Eqn (32a)].

With $F(0) = 0$, the theory has the constant solution $\pi_c = \text{const}$, $Y = 0$, and $T_{\mu\nu} = 0$. In the absence of other forms of energy, this solution corresponds to Minkowski space. Equations (33a) and (33b) show that the Minkowski background is stable for $F'(0) > 0$ and that perturbations travel with the speed of light [we recall that we set $K(0) = 0$]. This is easy to understand: it follows from Eqn (28) that perturbations with respect to a constant π_c are governed by the first term there, and $L^{(2)} = e^{2\pi_c}F'(0)(\partial\chi)^2$, which is the Lagrangian for a massless scalar field. In the neighborhood of the Minkowski background, i.e., for small $\partial\pi_c$, perturbations are not superluminal [70] if $K'(0) = 0$, $F''(0) > 0$.

In a wide range of the functions F and K, Eqn (31) also has a rolling solution,

$$e^\pi = \frac{1}{\sqrt{Y_*}(t_* - t)}\,, \qquad (34)$$

where t_* is an arbitrary constant. For this solution, $Y = Y_* = \text{const}$, and Y_* is determined from the equation

$$Z(Y_*) \equiv -F + 2Y_*F' - 2Y_*K + 2Y_*^2K' = 0\,, \qquad (35)$$

where F, F', etc., are evaluated at $Y = Y_*$. For this solution, we have $T_{00} = \rho = 0$ and

$$p = \frac{1}{Y_*^2(t_* - t)^4}(F - 2Y_*K)\,. \qquad (36)$$

Thus, the rolling background violates the NEC if

$$\text{NEC violation}: \quad 2Y_*K - F > 0\,. \qquad (37)$$

The quadratic Lagrangian for perturbations (6) reduces in this background to

$$L^{(2)} = \frac{A}{Y_*(t_* - t)^2}\big[\dot\chi^2 - (\partial_i\chi)^2\big] + \frac{B}{Y_*(t_* - t)^2}\,\dot\chi^2 + \frac{C}{Y_*^2(t_* - t)^4}\,\chi^2\,, \qquad (38)$$

where

$$A = \frac{1}{2}\,e^{-2\pi_c}V = F' - 2K + 4Y_*K'\,,$$

$$B = \frac{1}{2}\,e^{-2\pi_c}(U - V) = 2Y_*F'' - 2Y_*K' + 2Y_*^2K''\,,$$

$$C = 8F - 12Y_*F' + 8Y_*^2F'' + 8Y_*K - 8Y_*^2K' + 8Y_*^3K''$$

are time-independent coefficients. As a cross check, we can derive the equation for a homogeneous perturbation $\chi(t)$ near the rolling background from the last Lagrangian and see that $\chi = \partial_t\pi_c = (t_* - t)^{-1}$ obeys this equation, as it should. Indeed, using Eqn (35), we find that the coefficients of $\dot\chi^2$ and χ^2 in Eqn (38) are related in a simple way:

$$4(A + B) = \frac{C}{Y_*}\,.$$

Hence, the homogeneous perturbation obeys a universal equation,

$$-\frac{d}{dt}\left(\frac{\dot\chi}{(t_* - t)^2}\right) + 4\,\frac{\chi}{(t_* - t)^4} = 0\,,$$

whose solutions are $\chi = (t_* - t)^{-1}$ and $\chi = (t_* - t)^4$. This shows that the rolling background is an attractor and that it is stable against low-momentum perturbations: the growing perturbation $\chi = (t_* - t)^{-1}\chi_0(\mathbf{x})$ with a slowly varying $\chi_0(\mathbf{x})$ can be absorbed into a slightly inhomogeneous time shift.

We consider the stability of the rolling background and subluminality of the perturbations on it. The spatial gradient term in (38) has the correct (negative) sign if

$$\text{No gradient instability:} \quad A = F' - 2K + 4Y_*K' > 0\,. \qquad (39)$$

The speed of perturbations with respect to the rolling background is less than the speed of light if the coefficient of $\dot\chi^2$ is greater than that of $-(\partial_i\chi)^2$, i.e.,

$$\text{Subluminality}: \quad B = 2Y_*F'' - 2Y_*K' + 2Y_*^2K'' > 0\,. \qquad (40)$$

We require that this inequality hold in the strong sense; then the perturbations with respect to the rolling solution are strictly subluminal and hence the perturbations with respect to backgrounds neighboring the rolling solution are also subluminal. When both inequalities (39) and (40) are satisfied, there are no ghosts either. Conditions (37), (39), and (40), together with Eqn (35), can be satisfied at $Y = Y_*$ by a judicious choice of the functions F and K in the neighborhood of this point, such that the NEC violation is stable and subluminal. This can be seen as follows. Equation (35) can be used to express $F(Y_*)$ in terms of $F'(Y_*)$, $K(Y_*)$, and $K'(Y_*)$, namely, $F = 2Y_*F' - 2Y_*K + 2Y_*^2K'$. Then inequalities (37) and (39) are satisfied if $2K - 4Y_*K' < F' < 2K - Y_*K'$, which is possible for positive K'. Condition (40) can be satisfied by an appropriate choice of F'' and K''.

To tie in with existing studies, we note that a particular Lagrangian like (28) considered in Ref. [67] is

$$L_\pi = -f^2 e^{2\pi}(\partial\pi)^2 + \frac{f^2}{2\Lambda^3}(1+\alpha)(\partial\pi)^4 + \frac{f^3}{\Lambda^3}(\partial\pi)^2\Box\pi\,, \qquad (41)$$

which corresponds to

$$F = -f^2 Y + \frac{f^2}{2\Lambda^3}(1+\alpha)Y^2\,, \qquad K = \frac{f^3}{\Lambda^3}\,Y\,.$$

Here, the parameters f and Λ have the dimension of mass, and the parameter α is dimensionless. The solution of Eqn (35) is

$$Y_* = \frac{2}{3(1+\alpha)}\,\frac{\Lambda^3}{f}\,.$$

The energy scale $\sqrt{Y_*}$ associated with this solution is required to be lower than Λ, which is interpreted as the UV cutoff scale. This gives

$$f \gg \Lambda\,. \qquad (42)$$

From Eqn (37), we find that the background $Y = Y_*$ violates the NEC if

$$2Y_*K - F = 2f^2 Y_*\,\frac{3+\alpha}{1+\alpha} > 0\,,$$

while the stability and subluminality conditions, Eqns (39) and (40), give

$$A = \frac{3-\alpha}{3(1+\alpha)}\,f^2 > 0\,, \qquad B = \frac{4\alpha}{3(1+\alpha)}\,f^2 > 0\,.$$

All these conditions are satisfied for [67]

$$0 < \alpha < 3\,.$$

The case $\alpha = 0$ corresponds to luminal propagation of perturbations around the background solution $Y = Y_*$. In fact, in this case, theory (41) is invariant under conformal symmetry [50, 66]. However, the case $\alpha = 0$ is problematic because there are backgrounds in the neighborhood of $Y = Y_*$ on which the propagation of perturbations is superluminal [67]. We also note that Lagrangian (41) does not allow a stable Minkowski background, since $F'(0) < 0$. A conformally invariant theory with a stable Minkowski background and subluminal propagation in the rolling solution $Y = Y_*$ and in its neighborhood was constructed in Ref. [69] building upon Ref. [52], and is known as the DBI conformal Galileon theory.

To end this section, we consider, following [70], the structure of the configuration space $(\pi, \dot\pi)$ of spatially homogeneous Galileons in an arbitrary Galileon theory with scale invariance (27). The Lagrangian can contain all terms discussed in Section 3. We noted above that for $\dot\pi \neq 0$, the field equation is equivalent to the energy conservation condition $\dot\rho = 0$. This is not accidental. The Noether theorem states that the Noether energy–momentum tensor (which coincides with the metric energy–momentum tensor for a scalar field minimally coupled to gravity) obeys

$$\partial_\mu T^\mu_\nu = -(\text{E.O.M.})\,\partial_\nu\pi\,,$$

where (E.O.M.) stands for the equation of motion. Therefore, the equation of motion for a spatially homogeneous $\pi = \pi(t)$ is

$$(\text{E.O.M.}) = -\frac{1}{\dot\pi}\,\dot\rho\,. \qquad (43)$$

Because the field equation is of the second order, $\rho = \rho(\pi, \dot\pi)$ does not contain $\ddot\pi$ or higher derivatives, and by scale invariance it has the form

$$\rho = e^{4\pi}\,Z(Y)\,,$$

where $Y = \dot\pi^2 \exp(-2\pi)$ [cf. Eqn (29)] and Z is a model-dependent function. Now we can understand in more general terms that the rolling background with $Z = 0$ and $\dot\pi > 0$ is an attractor in the class of homogeneous solutions. For this, we use the energy conservation condition $\dot\rho = 0$ and, for any homogeneous solution, write

$$e^{4\pi}Z = \text{const}\,. \qquad (44)$$

As π increases, $|Z|$ decreases, and hence the solution tends to a configuration with $Z \to 0$. The configuration space of homogeneous Galileons with $\dot\pi > 0$ is thus divided into basins of attraction of solutions with $Z = 0$.

We also noted above that the coefficient U entering the quadratic action for perturbations is proportional to Z'. This is not accidental either. To see this, we again use Eqn (43), valid for any homogeneous Galileon. It follows from this equation that the equation of motion for a homogeneous perturbation about the background $\pi_c(t)$ has the form

$$-\frac{1}{\dot\pi_c}\,\frac{\partial\rho}{\partial\dot\pi_c}\,\ddot\chi + \ldots = 0\,,$$

where the omitted terms do not contain $\ddot\chi$. Hence, the Lagrangian for the perturbations has the form

$$L^{(2)} = \frac{1}{2\dot\pi_c}\,\frac{\partial\rho}{\partial\dot\pi_c}\,\dot\chi^2 + \ldots = e^{2\pi_c}\,Z'(Y)\dot\chi^2 = \ldots\,,$$

where the omitted terms do not contain $\dot\pi$. We conclude that $\rho = e^{4\pi_c}Z(Y_c)$ and $U = e^{2\pi_c}Z'(Y_c)$ for any point $(\pi_c, \dot\pi_c)$ in the configuration space of a homogeneous Galileon in any scale-invariant Galileon theory.

We finally recall that a configuration space point $(\pi_c, \dot\pi_c)$ at which $U < 0$ is unstable: there is either a ghost or a gradient instability among perturbations about this point. The above results therefore mean that any path in the space of homogeneous configurations $(\pi, \dot\pi)$ that connects two zero-energy attractor solutions, $Z = 0$, passes through an unstable region: indeed, Z' is negative somewhere on this path. This property creates difficulties in using scale-invariant Galileons, as we discuss in what follows. Here, we note that it implies that there is no evolution without pathologies that connects the Minkowski and rolling backgrounds, even if this evolution is driven by a source (as long as this source does not couple to $\dot{\pi}$).

The above analysis heavily uses scale invariance. Once one gives up scale invariance, this analysis and its conclusions are no longer valid. In particular, evolution from a nearly Minkowski regime to the rolling regime can occur without pathologies [71].

4.2 Genesis scenario

As the first example of using the solution discussed in Section 4.1, we consider Galilean Genesis [66] — a cosmological scenario alternative to inflation.[4] We assume that at early times $t \to -\infty$, the space–time is Minkowskian, the energy and pressure vanish, and the Universe is empty. At that time, the only relevant form of matter is the Galileon field π described by Lagrangian (28) (other Galileon Lagrangians are considered in Refs [69, 89] with similar results). Once conditions (37), (39), and (40) are satisfied, the solution $Y = Y_*$ is stable and violates the NEC. At the initial stage of the evolution, i.e., at large enough $t_* - t$, the energy density and pressure are small, and we can use the perturbation theory in $G \equiv M_{\rm Pl}^{-2}$. Equation (4) with $\kappa = 0$ determines the Hubble parameter, and to the lowest nontrivial order in $M_{\rm Pl}^{-2}$, we use the Minkowski expressions for the energy density, $\rho = 0$, and pressure, Eqn (36),

$$p = -\frac{P}{(t_* - t)^4}\,, \tag{45}$$

where $P = (2Y_* K - F)Y_*^{-2}$. It then follows that

$$H = \frac{4\pi P}{3M_{\rm Pl}^2 (t_* - t)^3}\,.$$

Equation (3a) is then used to find the energy density in the first order in $M_{\rm Pl}^{-2}$:

$$\rho = \frac{3}{8\pi} M_{\rm Pl}^2 H^2 = \frac{3\pi}{8} \frac{P^2}{M_{\rm Pl}^2 (t_* - t)^6}\,.$$

We see that as the field $\pi_{\rm c}$ evolves, the energy density increases, and the cosmological expansion is accelerated. The weak gravity approximation (the expansion in $M_{\rm Pl}^{-2}$) is valid when $\rho \ll p$, i.e.,

$$(t_* - t)^2 \gg \frac{P}{M_{\rm Pl}^2}\,. \tag{46}$$

The parameter P can be large: in the example with Lagrangian (41), we have $P \sim f^3/\Lambda^3 \gg 1$ in view of Eqn (42). Still, if P is not exceedingly large, the weak gravity regime holds almost to the Planck scale.

As a cross check, we consider the field equation for a homogeneous π in an expanding spatially flat Universe. It is given by

$$\begin{aligned}
&4{\rm e}^{4\pi}F + F'{\rm e}^{2\pi}(-6\dot{\pi}^2 - 2\ddot{\pi}) + 4F''(-\dot{\pi}^2\ddot{\pi} + \dot{\pi}^4)\\
&\quad + 4K{\rm e}^{2\pi}(\dot{\pi}^2 + \ddot{\pi}) - 4K'(\dot{\pi}^2\ddot{\pi} + \dot{\pi}^4)\\
&\quad + 4{\rm e}^{-2\pi}K''(-\dot{\pi}^4\ddot{\pi} + \dot{\pi}^6) - 6H{\rm e}^{2\pi}\dot{\pi}F' + 12H{\rm e}^{2\pi}\dot{\pi}K\\
&\quad - 6K'(2H\dot{\pi}^3 + 2H\dot{\pi}\ddot{\pi} + \dot{H}\dot{\pi}^2 + 3H^2\dot{\pi}^2)\\
&\quad + 12{\rm e}^{-2\pi}HK''(-\dot{\pi}^3\ddot{\pi} + \dot{\pi}^5) = 0\,.
\end{aligned} \tag{47}$$

We see that gravitational corrections here are small if $H \ll \dot{\pi}$, which again gives condition (46). Discussing the weak gravity regime is sufficient for our purposes, but of course one can follow the evolution after the end of this regime, with gravity effects fully accounted for. This is done in Ref. [66] in model (41) with $\alpha = 0$.

So far, we have seen that the theory allows a cosmological scenario in which the Universe is initially empty and Minkowskian and evolves into the stage of rapid expansion and high energy density. This evolution is precisely the Genesis epoch. There are two other ingredients in the Genesis scenario. First, at some late stage, the Galileon energy density should be converted into heat, and the standard hot epoch should begin. A possible mechanism of 'defrosting' is suggested in Ref. [90]. At the end of the defrosting stage, whatever it is, the Galileon should settle to its Minkowski value, $\dot{\pi} = 0$. In a scale-invariant Galileon theory, this is problematic, because of our observations at the end of Section 4.1. The violation of scale invariance at defrosting can probably cure this problem.

The second ingredient is a mechanism of the generation of density perturbations, eventually responsible for CMB anisotropies and structure formation. These perturbations are Gaussian (or nearly Gaussian) random fields with a nearly flat power spectrum. Perturbations in the Galileon field itself cannot do the job [66]. A simple extension of the Galileon theory can work quite well, however [66]. We assume that the theory is scale invariant in the Genesis epoch and add a new field θ that transforms trivially under scale transformations, $\theta(x) \to \theta(\lambda x)$. By scale invariance, the kinetic term in its Lagrangian is

$$L_\theta = \frac{1}{2}\,{\rm e}^{2\pi}(\partial\theta)^2\,.$$

If other interactions of the new field are negligible in the Genesis epoch, the Lagrangian in rolling background (34) is

$$L_\theta = \frac{1}{2}\frac{1}{Y_*(t_* - t)^2}(\partial\theta)^2\,.$$

This coincides with the Lagrangian of a scalar field minimally coupled to gravity, evolving in the inflationary epoch with the Hubble parameter $\sqrt{Y_*}$, if we identify t with the conformal time at inflation. Thus, we borrow the well-known result of the inflationary theory: vacuum fluctuations of the field θ develop into a Gaussian random field with the power spectrum

$$\mathcal{P}_{\delta\theta} = \frac{Y_*}{(2\pi)^2}\,. \tag{48}$$

The field perturbations $\delta\theta$, which are entropy fluctuations in the Genesis epoch, are assumed to be reprocessed into adiabatic perturbations some time after the Genesis epoch, e.g., by a curvaton [91–95] or modulated decay [96–98]

[4] The term Genesis, corresponding to the English name of the first book of the Old Testament, was introduced by the authors of [66].

mechanism. The adiabatic perturbations ζ inherit the properties of perturbations $\delta\theta$ (modulo non-Gaussianities that may be produced in the process of conversion of entropy to the adiabatic perturbation); in particular, their power spectrum is $\mathcal{P}_\zeta = \mathrm{const} \cdot \mathcal{P}_\theta$. Spectrum (48) is flat; a small tilt, required by observations [99, 100], can emerge due to weak explicit breaking of scale invariance (cf. Ref. [101]).

To conclude this section, we note that the Genesis scenario, especially its version with the conformal Galileon, is an example of what is now called (pseudo-)conformal cosmology [66, 68, 102–104]. In general terms, this class of scenarios assumes that the Universe is initially effectively Minkowskian, and matter is in a conformally invariant state. Conformal invariance is then spontaneously broken by a rolling background, similar to (34). The mechanism of the generation of density perturbations is similar to the one just discussed. The conformal scenario makes a number of model-independent predictions that potentially distinguish it from inflation. These include non-Gaussianities and a statistical anisotropy of scalar perturbations [105–109]. Another property is the absence of tensor perturbations.

4.3 Bouncing Universe

Galileon theories can also be used to construct models of a bouncing universe [66, 110–114]. Before we discuss a concrete model of this sort, we make the following comment. A contracting universe can easily become strongly inhomogeneous and anisotropic because of the Belinsky–Lifshitz–Khalatnikov phenomenon [115–119]. This creates a consistency problem for the entire bouncing scenario. A way to solve this problem is to assume that the dominant matter at the contracting stage has a super-stiff equation of state, $p > \rho$ [120]. This is what is generically called the ekpyrotic Universe [83, 84]. We discuss this point in Appendix B. We note that for matter with the equation of state $p = w\rho$, $w = \mathrm{const}$, Eqn (5) gives $\rho \propto a^{-3(1+w)}$, and we then find from Eqn (3a) with $\kappa = 0$ that the scale factor evolves as

$$a(t) \propto |t|^{\alpha}\,, \qquad t < 0\,,$$

where

$$\alpha = \frac{2}{3(1+w)}\,.$$

The super-stiff equation of state, $w > 1$, hence corresponds to

$$\alpha < \frac{1}{3}\,. \tag{49}$$

An example of super-stiff matter is a scalar field with a negative exponential potential,

$$L_\phi = \frac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi - V(\phi)\,, \qquad V(\phi) = -V_0\,\mathrm{e}^{\phi/M}\,, \tag{50}$$

where V_0 and M are positive parameters. The equation for the homogeneous field $\phi(t)$ and Friedmann equation (3a) have the solution

$$a(t) = |t|^{\alpha}\,, \qquad \phi(t) = \mathrm{const} - 2M\ln|t|\,,$$
$$V\big[\phi(t)\big] = -\frac{2M^2(1-3\alpha)}{t^2}\,, \qquad t < 0\,, \tag{51}$$

where

$$\alpha = 16\pi\,\frac{M^2}{M_{\mathrm{Pl}}^2}\,. \tag{52}$$

This is an attractor in the case of collapse. According to (49) and (50), the effective equation of state is indeed super-stiff, $w \gg 1$, for $M \ll M_{\mathrm{Pl}}$. We note that the energy density is positive and increases as the Universe collapses,

$$\rho = \frac{1}{2}\,\dot\phi^2 + V(\phi) = \frac{6M^2\alpha}{t^2}\,.$$

This leaves open the possibility that the potential $V(\phi)$ becomes positive at large ϕ, and the field moves out of the negative potential at some late epoch.

It is worth noting that for $M \ll M_{\mathrm{Pl}}$, this solution is always in the weak gravity regime similar to that studied in Section 4.2. In the weak gravity limit, we neglect gravity in the field equation for ϕ and obtain the solution in Minkowski space:

$$\phi(t) = M\ln\left(\frac{2M^2}{t^2V_0}\right), \qquad V(t) = -\frac{2M^2}{t^2}\,. \tag{53}$$

The energy density vanishes in this limit, while the pressure is

$$p = \frac{1}{2}\,\dot\phi^2 - V = \frac{4M^2}{t^2}\,. \tag{54}$$

The weak gravity approximation is valid at all times for $M \ll M_{\mathrm{Pl}}$.

To construct models where the ekpyrotic epoch ends with a bounce, it suffices in principle to restrict to theories with a single scalar field [112, 114]. But it is much simpler [113] to extend the model in (50) by adding a new Galileon field with Lagrangian (28), such that the total matter Lagrangian becomes

$$L = L_\pi + L_\phi\,. \tag{55}$$

In the weak gravity limit, the fields ϕ and π do not interact with each other, the Galileon rolls as in Eqn (34), while $\phi(t)$ is given by (53). The energy density is zero, and the pressure is the sum of (45) and (54):

$$p = \frac{4M^2}{t^2} - \frac{P}{(t_*-t)^4}\,. \tag{56}$$

The Hubble parameter is found from Eqn (4):

$$H = -\frac{16\pi M^2}{M_{\mathrm{Pl}}^2|t|} + \frac{4\pi P}{3M_{\mathrm{Pl}}^2(t_*-t)^3}\,. \tag{57}$$

At early times, the field ϕ dominates and the universe contracts ($H < 0$); later, the Galileon takes over, at least for $t_* < 0$, the contraction terminates ($H = 0$, bounce), and the expansion epoch begins and proceeds as in the Genesis scenario ($H > 0$). It is not difficult to see that the bounce indeed occurs in the weak gravity regime $H \ll \dot\pi$ if the following mild inequality holds: $|t_*| \gg P^{1/2}M^2/M_{\mathrm{Pl}}^3$, $t_* < 0$ (the case $t_* > 0$ is considered in Ref. [113] with the result that the bounce always occurs, but not necessarily in the weak gravity regime).

To make this toy model more realistic, one modifies the potential $V(\phi)$ at large ϕ and adds the potential to the

Galileon to ensure that the cosmological constant vanishes at late times. Depending on the parameters, the system may or may not enter the late time inflationary regime [113]. The ingredients discussed at the end of Section 4.2 have to be present in this model as well.

4.4 Creating a universe in the laboratory

Our last example is an attempt to design a model for the creation of a universe in the laboratory [70]. The idea is to construct the initial condition in a Galileon-type theory such that inside some large sphere, the field π is nearly homogeneous and behaves as it does at the initial stage of Genesis, whereas outside this sphere, this field tends to a constant and the space–time is asymptotically Minkowskian. For this initial data, the energy density and pressure are initially small everywhere, and the entire space–time is nearly Minkowskian, and hence the required field configuration can in principle be prepared in the laboratory. As the field $\pi(t, \mathbf{x})$ evolves from this initial state according to its equation of motion, the energy density inside the large sphere increases, the space undergoes accelerated expansion there, and the region inside the sphere eventually becomes a human-made universe. Outside this sphere, the energy density remains small and asymptotically tends to zero at large distances; the space–time is always asymptotically Minkowskian.

It is tempting to implement this idea in a simple way, by considering the initial field $\pi(t, \mathbf{x})$ that slowly varies in space and interpolates between the rolling solution (34) inside the large sphere and the Minkowski vacuum $\partial\pi = 0$ at spatial infinity. By slow variation in space, we mean that the spatial derivatives of π are negligible compared to temporal ones, such that at each point in space, π evolves in the same way as in the homogeneous case.

An advantage of this quasi-homogeneous approach is its simplicity; a disadvantage is that it does not actually work in the class of scale-invariant models in Section 4.1. The obstruction comes from the property discussed at the end of Section 4.1: if the evolution of π is effectively homogeneous everywhere, then the analysis in Section 4.1 applies, and because $Z(Y)$ vanishes both inside the large sphere (Genesis region) and far away from it (Minkowski region), there is a region in between where $Z' < 0$ and the system is unstable.

One way to bypass this obstruction would be to insist on the slow spatial variation of the initial field configuration, but give up the requirement that the field inside the large sphere be in the Genesis regime (34). Instead, we could consider the field with a nonzero energy density inside the sphere, such that there would be a smooth and stable configuration that interpolates, as r increases, between this field and the asymptotic Minkowski vacuum. This can hardly lead to the creation of a universe, however, since, as we discussed in Section 4.1, the Minkowski point $Y = 0$ is an attractor, and the field in the interior of the sphere would relax to it.

Other possibilities are to consider field configurations with nonnegligible spatial gradients or give up scale invariance of the action (the latter possibility has been successfully explored in Ref. [71] in the cosmological context). In either case, the above no-go argument would be irrelevant, but the analysis would be more complicated. It is simpler to follow another route, and complicate the model instead.

For this, the functions F and K are allowed to depend explicitly on spatial coordinates. This can be the case if there is another field, φ, that determines the couplings entering these functions, and this field acts as a time-independent background, $\varphi = \varphi(\mathbf{x})$. In this case, we can consider a field configuration $\pi(t, \mathbf{x})$ that at any point in space is approximately given by the rolling solution (34), but with Y_* depending on $\mathbf{x}$. The background $\varphi(\mathbf{x})$ is prepared in such a way that $Y_*(\mathbf{x})$ is constant inside the large sphere (to evolve into a human-made universe) and gradually approaches zero as $r \to \infty$. It is straightforward to verify that with an appropriate choice of the functions $F(Y; \varphi)$ and $K(Y; \varphi)$, this construction does not encounter pathologies anywhere.

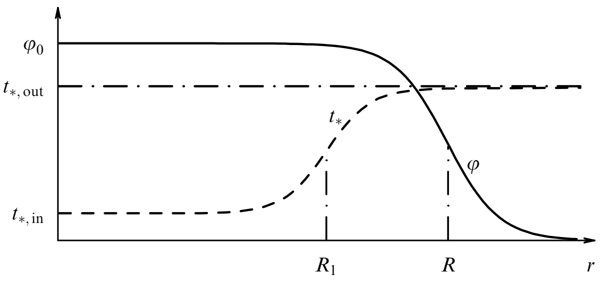

Figure 3. The setup. Dashed and solid lines respectively show $t_*(r)$ and $\varphi(r)$. The behavior of the function $Y_*(r) = Y_*(\varphi(r))$ is similar to that of $\varphi(r)$.

We now sketch a concrete construction. We assume that the field φ is a usual scalar field and has two vacua, $\varphi = 0$ and $\varphi = \varphi_0$. We prepare a spherical configuration of this field with $\varphi = \varphi_0$ inside a sphere of a sufficiently large radius R and $\varphi = 0$ outside this sphere (Fig. 3). We assume for definiteness that there is a source for the field φ that keeps this configuration static. Let $L \ll R$ be the thickness of the wall separating the two vacua; L is also kept time-independent by the source. We require that the mass of this ball be small enough, such that $R \gg R_s$, where R_s is the Schwarzschild radius. The mass is of the order of $\mu^4 R^2 L$, where μ is the mass scale characteristic of the field φ. Hence, the last requirement becomes $\mu^4 RL \ll M_{\mathrm{Pl}}^2$. For small enough μ, both R and L can be large.

Let the function $Y_*(\varphi)$ be such that $Y_*(0) = 0$ and $Y_*(\varphi_0) = Y_0$. We prepare the initial configuration of π at $t = 0$ in such a way that it initially evolves as

$$\mathrm{e}^{\pi} = \frac{1}{\sqrt{Y_0}\, t_*(r) - \sqrt{Y_*(r)}\, t} \,, \tag{58}$$

where we allow the parameter t_* in (34) to vary in space, and choose a convenient parameterization. We choose $t_*(r) = t_{*,\,\mathrm{in}}$ inside a somewhat smaller sphere of a radius $R_1 < R$ (but $R_1 \sim R$) and $t_*(r) = t_{*,\,\mathrm{out}} \gg t_{*,\,\mathrm{in}}$ at $r > R_1$ (hereafter, subscripts 'in' and 'out' refer to the respective regions $r < R_1$ and $r > R_1$), as shown in Fig. 3, with the transition region, e.g., of the same thickness L. We take $t_{*,\,\mathrm{out}} \ll L$; then the characteristic time scales are smaller than the smallest length scale L inherent in the setup, and therefore the spatial derivatives of π are indeed negligible compared to the time derivatives. This ensures that the field π is in the quasi-homogeneous regime. As $r \to \infty$, we have $Y_*(r) \to 0$ and $t_* \to \mathrm{const}$, and hence π tends to the Minkowski vacuum $\pi = \mathrm{const}$.

At the initial stage of evolution, the pressure inside the radius-R_1 sphere is

$$p_{\mathrm{in}} = \frac{M^4}{Y_0^2 (t_{*,\,\mathrm{in}} - t)^4} \,,$$

where M is the mass scale characteristic of the field π. We require that $p_{\text{in}} R^3/M_{\text{Pl}}^2 \ll R$, then the gravitational potentials are small everywhere, and gravity is initially in the linear regime. Thus, we impose the constraint

$$\frac{M^4 R^2}{Y_0^2 t_{*,\text{in}}^4} \ll M_{\text{Pl}}^2 , \tag{59}$$

which is consistent with the above conditions for $M \ll M_{\text{Pl}}$ and $Y_0 \gtrsim M^2$. In complete analogy with Section 4.2, the Hubble parameter inside the sphere of the radius R_1 shortly after the beginning of evolution is

$$H_{\text{in}} = \frac{4\pi M^4}{3 M_{\text{Pl}}^2 Y_0^2 (t_{*,\text{in}} - t)^3} . \tag{60}$$

In view of (59) and $t_{*,\text{in}} \ll R$, the Hubble length scale is large for some time, $H^{-1} \gg R$. This is also true for $r > R_1$, and hence there are no anti-trapped surfaces initially.

As t approaches $t_{*,\text{in}}$, the pressure in the Genesis region $r < R_1$ increases, and the Hubble length shrinks there to $R_1 \sim R$. The anti-trapped surfaces are formed inside the sphere of the radius R_1, and a new universe is created and enters the Genesis regime there. This occurs when $H_{\text{in}} \sim R^{-1}$, i.e., at a time t_1 such that

$$(t_{*,\text{in}} - t_1) \sim \left(\frac{M^4 R}{M_{\text{Pl}}^2 Y_0^2}\right)^{1/3} .$$

We note that at that time, the energy density $\rho_{\text{in}} \sim M_{\text{Pl}}^2 H_{\text{in}}^2$ is still relatively small,

$$\frac{\rho_{\text{in}}}{p_{\text{in}}} \sim \left(\frac{M^4}{Y_0^2 R^2 M_{\text{Pl}}^2}\right)^{1/3} \ll 1 .$$

This implies that at the time t_1, space–time is locally nearly Minkowskian. Another manifestation of this fact is that the scale factor is close to unity:

$$a_{\text{in}}(t_1) = 1 + \frac{2\pi M^4}{3 M_{\text{Pl}}^2 Y_0^2 (t_{*,\text{in}} - t_1)^2} , \tag{61}$$

where the correction to unity is of the order of $\rho_{\text{in}}/p_{\text{in}}$. Hence, our approximate solution (58), (60) is legitimate.

Because $t_{*,\text{out}} \gg t_{*,\text{in}}$, the field e^{π} at the time t_1 is still small for $r > R_1$, and the Hubble length scale exceeds R there. Gravity is still weak at $r > R_1$, and it is therefore consistent to assume that the configuration of φ is not modified by that time. We also note that a black hole is not formed by then either.

This completes the construction of the initial configuration and the analysis of the early epoch of a human-made universe. We verify in Appendix C that this analysis does not contradict the general results in Ref. [13].

Of course, the construction discussed here is merely a sketch. To make the scenario complete, one has to specify a way to design the configuration of the field φ and keep it static (or consider an evolving field φ instead). Also, one has to understand the role of spatial gradients. Finally, one would like to trace the dynamics of the system to longer times, with gravity effects included, and see what geometry develops towards the end of the Genesis epoch occuring at $r < R_1$. In particular, it is of interest to see whether a black hole is formed.

5. Conclusion

The theories of (generalized) Galileons offer an interesting possibility of consistent and controllable NEC violation. Still, open issues remain. One of them is the danger of superluminality. While the background we consider in Section 4.1 may be safe in this respect, it is not impossible that other backgrounds are not, especially when gravity generated by some other matter is relevant. An example of this sort is given in Ref. [121]. The superluminality issue is closely related to the possibility of UV completion [72]. Another issue is the stability against radiative corrections. While the simplest Galileon theories have enough symmetries to guarantee stability, generic Galileon Lagrangians (22) do not. There are also largely unexplored areas where NEC-violating theories may lead to surprises, like black hole thermodynamics [122] and the absence/existence of closed time-like curves [123].

Of course, the most intriguing question is whether NEC-violating fields exist in Nature. Needless to say, no such fields have been discovered. The situation is not entirely hopeless, however: we may learn at some point in the future that the Universe went through the bounce or Genesis epoch, and that would be an indication that NEC violation indeed took place in the past.

The author is indebted to S Demidov, D Levkov, M Libanov, I Tkachev, and M Voloshin for the helpful discussions and S Deser, Y-S Piao, and A Vikman for the useful correspondence. The work was supported in part by a grant from the President of the Russian Federation, NS-5590.2012.2, and the Ministry of Education and Science, contract 8412.

Appendix A

We consider a general spherically symmetric metric, which we choose in diagonal form,

$$ds^2 = N^2 dt^2 - a^2 dr^2 - R^2 (d\theta^2 + \sin^2\theta \, d\varphi^2) , \tag{62}$$

where $N = N(t,r)$, $a = a(t,r)$, and $R = R(t,r)$. Our purpose is to show that a trapped sphere is such that $R(t, r(t))$ decreases along an outgoing null geodesic for which r increases.

The formal definition of a trapped sphere is that

$$\nabla_\mu l^\mu < 0$$

for a vector $l^\mu = dx^\mu/d\lambda$ tangent to an outgoing radial null geodesic, where λ is the affine parameter. The vector l^μ is null,

$$g_{\mu\nu} l^\mu l^\nu = 0 , \tag{63}$$

and obeys the geodesic equation

$$\frac{dl^\mu}{d\lambda} + \Gamma^\mu_{\nu\rho} l^\nu l^\rho = 0 . \tag{64}$$

For metric (62), Eqn (63) gives

$$l^0 = u(t) , \qquad l^r = u(t) \, \frac{N(t, r(t))}{a(t, r(t))} ,$$

where we have chosen to parameterize the geodesic by the time t, such that the null world line is $(t, r(t), 0, 0)$; the sign of l^r corresponds to the outgoing geodesic. The normalization factor $u(t)$ is to be determined from Eqn (64). To find this factor, we write $\mathrm{d}l^\mu/\mathrm{d}\lambda = \mathrm{d}l^\mu/\mathrm{d}t \cdot l^0$ and, for the 0th component of Eqn (64), obtain

$$\frac{\mathrm{d}u}{\mathrm{d}t} + \Gamma^0_{00} u + 2\Gamma^0_{0r} u \frac{N}{a} + \Gamma^0_{rr} u \left(\frac{N}{a}\right)^2 = 0 \,. \qquad (65)$$

The relevant Christoffel symbols are

$$\Gamma^0_{00} = \frac{\dot N}{N}\,, \qquad \Gamma^0_{0r} = \frac{N'}{N}\,, \qquad \Gamma^0_{rr} = \frac{a\dot a}{N^2}\,,$$

where the dot and the prime denote *partial* derivatives, and Christoffel symbols entering Eqn (65) are to be taken at $r = r(t)$. Thus, the function $u(t)$ obeys

$$\dot u + \left(\frac{\dot N}{N} + 2\,\frac{N'}{a} + \frac{\dot a}{a}\right) u = 0\,, \qquad (66)$$

where, again, the terms in parenthesis are partial derivatives evaluated at $r = r(t)$. As a cross check, we write the r-component of geodesic equation (64),

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[u\,\frac{N\big(t, r(t)\big)}{a\big(t, r(t)\big)}\right] + \Gamma^r_{00} u + 2\Gamma^r_{0r} u \frac{N}{a} + \Gamma^r_{rr} u \left(\frac{N}{a}\right)^2 = 0\,. \qquad (67)$$

Using the relations

$$\Gamma^r_{00} = \frac{NN'}{a^2}\,, \qquad \Gamma^0_{0r} = \frac{\dot a}{a}\,, \qquad \Gamma^0_{rr} = \frac{a'}{a}\,,$$
$$\frac{\mathrm{d}r(t)}{\mathrm{d}t} = \frac{N}{a}\,,$$

we find that Eqn (67) coincides with Eqn (66).

We now calculate

$$\nabla_\mu l^\mu = \frac{1}{\sqrt{-g}}\,\partial_\mu\big(\sqrt{-g}\,l^\mu\big)$$
$$= \frac{1}{aNR^2}\left[\partial_0(aNR^2 u) + \partial_r\left(aNR^2 u\,\frac{N}{a}\right)\right]$$
$$= \dot u + \left(\frac{\dot a}{a} + \frac{\dot N}{N} + 2\,\frac{\dot R}{R} + 2\,\frac{N'}{a} + 2\,\frac{R'}{R}\,\frac{N}{a}\right) u\,.$$

Using Eqn (66) to eliminate $\dot u$, we arrive at

$$\nabla_\mu l^\mu = 2\left(\frac{\dot R}{R}\,u + \frac{R'}{R}\,\frac{N}{a}\,u\right) = 2 l^\mu \partial_\mu R\,. \qquad (68)$$

Therefore, the trapped surface is indeed such that $R(t, r(t))$ decreases along the outgoing null geodesic.

As an example, for a contracting spatially flat Universe, we have $a = a(t)$, $R = a(t) r$, and the right-hand side of Eqn (68) is negative for $r > -1/\dot a$; a sphere of the radius $R = ar > |H|^{-1}$ is a trapped surface. By time reversal, a sphere of a radius $R > |H|^{-1}$ in the expanding Universe is an anti-trapped surface.

Appendix B

We briefly discuss why a contracting universe becomes strongly inhomogeneous and anisotropic if the dominant matter obeys $p < \rho$, and why, on the contrary, it becomes more isotropic in the course of contraction in the opposite case. We consider a simplified version of the anisotropic Universe described by the homogeneous anisotropic metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t) \sum_{a=1}^{3} \mathrm{e}^{2\beta_a(t)}\, e^{(a)}_i e^{(a)}_j\, \mathrm{d}x^i\, \mathrm{d}x^j\,,$$

where $e^{(a)}_i$ are three linearly independent vectors that are constant in time. We assume for simplicity that these vectors are orthogonal to each other (the dynamics is much more complicated in the general situation, but this turns out to be largely irrelevant from our standpoint; see the comment below). The function $a(t)$ is chosen such that

$$\sum_a \beta_a = 0\,; \qquad (69)$$

in other words, $\det g_{ij} = a^6$. The Einstein equations give

$$\left(\frac{\dot a}{a}\right)^2 = \frac{1}{6}\sum_a \dot\beta_a^2 + \frac{8\pi}{3}\,G\rho\,, \qquad (70a)$$

$$\ddot\beta_a + 3\,\frac{\dot a}{a}\,\dot\beta_a = 0\,. \qquad (70b)$$

Equation (70b) gives

$$\dot\beta_a = \frac{d_a}{a^3}\,, \qquad (71)$$

and in view of (69), the constants d_a obey $\sum_a d_a = 0$. Equation (70a) then becomes

$$\left(\frac{\dot a}{a}\right)^2 = \frac{1}{6a^6}\sum_a d_a^2 + \frac{8\pi}{3}\,G\rho\,. \qquad (72)$$

This equation shows that the overall contraction rate (the rate at which $\det g_{ij}$ decreases) is determined at small a by the anisotropy rather than matter, if ρ increases more slowly than a^{-6}. For metric (69), the covariant energy conservation condition still gives Eqn (5) with $H = \dot a/a$, and hence the above property holds for $p < \rho$. Therefore, we can set $\rho = 0$ late at the collapsing stage, and systems of equations (71) and (72) have the Kasner solution:

$$a(t) = |t|^{1/3}\,, \qquad \beta_a = d_a \ln|t|\,,$$
$$\sum_a d_a = 0\,, \qquad \sum_a d_a^2 = \frac{2}{3}\,.$$

Hence, the anisotropy increases as the Universe collapses. In general, when the vectors $e^{(a)}_i$ are not orthogonal to each other, this regime continues for a finite time, and then the values of the parameters d_a change in a rather abrupt manner [124]. The vectors $e^{(a)}_i$ also change. This change occurs infinitely many times in the limit $t \to 0$, which corresponds to chaotic anisotropic collapse.

These results show that the Universe is very anisotropic before the bounce. In fact, the processes we described occur independently in Hubble-size regions and are very different in each of them because of their chaotic properties, and therefore the Universe also becomes strongly inhomogeneous. This picture remains valid after the bounce, at least in the classical theory framework. A strong inhomogeneity of the Universe after the bounce is inconsistent with the smallness of the primordial cosmological perturbations, and hence the entire bounce scenario is up in the air.

To solve this problem, we invoke matter with a super-stiff equation of state $p = w\rho$, $w > 1$. Its energy density behaves as

$\rho \propto a^{-3(1+w)}$, and hence increases faster than a^{-6}. The second term in the right-hand side of Eqn (72) dominates, and the scale factor decreases as $a(t) \propto |t|^{\alpha}$ with $\alpha < 1/3$ [see Eqn (49)]. It then follows from Eqn (71) that the parameters β_a tend to constants as $t \to 0$. If the Universe is nearly homogeneous at the early stages of collapse, and anisotropy is not strong, then the Universe becomes more and more homogeneous in the process of contraction (see the details in Ref. [120]).

Appendix C

We show in this appendix that the results in Section 4.4 are in agreement with the general results in Ref. [13].

Definition [13]. *Let the metric have form* (62). *The R-region is a region where normal vectors* $R_\mu = \partial_\mu R$ *to the hypersurfaces* $R = \text{const}$ *are spacelike,* $g^{\mu\nu} R_\mu R_\nu < 0$. *Since* $g^{\mu\nu} R_\mu R_\nu = N^{-2}\dot{R}^2 - a^{-2}R'^2 < 0$, *there is no place in an R-region where* $R' = 0$, *and therefore the sign of* R' *is the same in the entire R-region. An R-region where* $R' > 0$ *is called an* R_+*-region, while an R-region where* $R' < 0$ *is called an* R_-*-region. The T-region is a region where normal vectors* R_μ *to the hypersurfaces* $R = \text{const}$ *are timelike,* $g^{\mu\nu} R_\mu R_\nu > 0$. *There,* $\dot{R}$ *is nonzero everywhere. Hence, the sign of* $\dot{R}$ *is the same everywhere. A T-region where* $\dot{R} > 0$ *is called a* T_+*-region, while a T-region where* $\dot{R} < 0$ *is called a* T_-*-region.* T_+*- and* T_-*-regions are regions of expansion and contraction, respectively.*

We consider the model in Section 4.4. In the above nomenclature, the whole space is initially an R_+-region. At the time t_1, a T_+-region appears. One of its boundaries moves toward smaller r, and another moves toward larger r. One of the results in Ref. [13] is that for $\rho + p < 0$ ($\beta < 0$ in the nomenclature of Ref. [13]), the boundary between the inner R_+-region and the T_+-region is necessarily space-like. We verify that our geometry is consistent with this result.

In our case, $N = 1$, $a \approx 1$ [see (61)]), and $R = a(r,t)r$. The boundary between the left R_+-region and the T_+-region is determined by $\dot{a}r = a$, i.e.,

$$r - H^{-1} = 0\,.$$

The normal to this hypersurface is the vector

$$\left(\frac{\dot{H}}{H^2}\,,1\,,0\,,0\right),$$

which is timelike because

$$\frac{\dot{H}}{H^2} \sim \big(H(t_{*,\text{in}} - t)\big)^{-1} \sim \frac{M_{\text{Pl}}^2 Y_0^2 (t_{*,\text{in}} - t)^2}{M^4} \gg 1\,.$$

Hence, the hypersurface separating the R_+- and T_+-regions is spacelike, in agreement with the general result in Ref. [13].

The outer boundary of the T_+-region can in principle be either spacelike or timelike [13]. For the same reason as above, it is actually spacelike in our case.

References

1. Flanagan E E, Wald R M *Phys. Rev. D* **54** 6233 (1996); gr-qc/9602052
2. Penrose R *Phys. Rev. Lett.* **14** 57 (1965)
3. Starobinskii A A *Sov. Astron. Lett.* **4** 82 (1978); *Pis'ma Astron. Zh.* **4** 155 (1978)
4. Berezin V A, Kuzmin V A, Tkachev I I, in *Quantum Gravity. Proc. of the Third Seminar on Quantum Gravity, October 23–25, 1984, Moscow, USSR* (Eds M A Markov, V A Berezin, V P Frolov) (Singapore: World Scientific, 1985) p. 605
5. Berezin V A, Kuz'min V A, Tkachev I I *JETP Lett.* **41** 547 (1985); *Pis'ma Zh. Eksp. Teor. Fiz.* **41** 446 (1985)
6. Farhi E, Guth A H *Phys. Lett. B* **183** 149 (1987)
7. Starobinskii A A *JETP Lett.* **30** 682 (1979); *Pis'ma Zh. Eksp. Teor. Fiz.* **30** 719 (1979)
8. Starobinsky A A *Phys. Lett. B* **91** 99 (1980)
9. Guth A H *Phys. Rev. D* **23** 347 (1981)
10. Linde A D *Phys. Lett. B* **108** 389 (1982)
11. Albrecht A, Steinhardt P J *Phys. Rev. Lett.* **48** 1220 (1982)
12. Linde A D *Phys. Lett. B* **129** 177 (1983)
13. Berezin V A, Kuz'min V A, Tkachev I I *Sov. Phys. JETP* **66** 654 (1987); *Zh. Eksp. Teor. Fiz.* **93** 1159 (1987)
14. Berezin V A, Kuzmin V A, Tkachev I I *Phys. Lett. B* **207** 397 (1988)
15. Farhi E, Guth A H, Guven J *Nucl. Phys. B* **339** 417 (1990)
16. Fischler W, Morgan D, Polchinski J *Phys. Rev. D* **41** 2638 (1990); *Phys. Rev. D* **42** 4042 (1990)
17. Linde A *Nucl. Phys. B* **372** 421 (1992); hep-th/9110037
18. Garriga J, Vilenkin A *Phys. Rev. D* **57** 2230 (1998); astro-ph/9707292
19. Garriga J et al. *Int. J. Theor. Phys.* **39** 1887 (2000); astro-ph/9909143
20. Dutta S, Vachaspati T *Phys. Rev. D* **71** 083507 (2005); astro-ph/0501396
21. Aguirre A, Johnson M C *Phys. Rev. D* **73** 123529 (2006); gr-qc/0512034
22. Lee W et al. *Phys. Rev. D* **74** 123520 (2006); hep-th/0604064
23. Piao Y-S *Nucl. Phys. B* **803** 194 (2008); arXiv:0712.4184
24. Frolov V P, Markov M A, Mukhanov V F *Phys. Lett. B* **216** 272 (1989); *Phys. Rev. D* **41** 383 (1990)
25. Guendelman E I *Int. J. Mod. Phys. D* **19** 1357 (2010); arXiv:1003.3975
26. Lukash V N, Mikheeva E V, Strokov V N *Phys. Usp.* **55** 204 (2012); *Usp. Fiz. Nauk* **182** 216 (2012); arXiv:1110.6820
27. Lukash V N, Strokov V N *Int. J. Mod. Phys. A* **28** 1350007 (2013); arXiv:1301.5544
28. Mukhanov V, Brandenberger R *Phys. Rev. Lett.* **68** 1969 (1992)
29. Brandenberger R, Mukhanov V, Sornborger A *Phys. Rev. D* **48** 1629 (1993); gr-qc/9303001
30. Trodden M, Mukhanov V F, Brandenberger R H *Phys. Lett. B* **316** 483 (1993); hep-th/9305111
31. Lee B-H et al. *Phys. Rev. D* **77** 063502 (2008); arXiv:0710.4599
32. Guendelman E I, Sakai N *Phys. Rev. D* **77** 125002 (2008); *Phys. Rev. D* **80** 049901(E) (2009), Erratum; arXiv:0803.0268
33. Yeom D, arXiv:0912.0068
34. Hwang D, Yeom D *Class. Quantum Grav.* **28** 155003 (2011); arXiv:1010.3834
35. Morris M S, Thorne K S *Am. J. Phys.* **56** 395 (1988)
36. Morris M S, Thorne K S, Yurtsever U *Phys. Rev. Lett.* **61** 1446 (1988)
37. Visser M *Lorentzian Wormholes: From Einstein to Hawking* (Woodbury, N.Y.: American Institute of Physics, 1995)
38. Hochberg D, Visser M *Phys. Rev. D* **58** 044021 (1998); gr-qc/9802046
39. Novikov I D, Kardashev N S, Shatskii A A *Phys. Usp.* **50** 965 (2007); *Usp. Fiz. Nauk* **177** 1017 (2007)
40. Shatskii A A, Novikov I D, Kardashev N S *Phys. Usp.* **51** 457 (2008); *Usp. Fiz. Nauk* **178** 481 (2008); arXiv:0810.0468
41. Rubakov V A *Theor. Math. Phys.* **149** 1651 (2006); *Teor. Mat. Fiz.* **149** 409 (2006); hep-th/0604153
42. Libanov M et al. *JCAP* (08) 010 (2007); arXiv:0704.1848
43. Horndeski G W *Int. J. Theor. Phys.* **10** 363 (1974)
44. Fairlie D B, Govaerts J, Morozov A *Nucl. Phys. B* **373** 214 (1992); hep-th/9110022
45. Fairlie D B, Govaerts J *Phys. Lett. B* **281** 49 (1992); hep-th/9202056
46. Fairlie D B, Govaerts J *J. Math. Phys.* **33** 3543 (1992); hep-th/9204074
47. Dvali G, Gabadadze G, Porrati M *Phys. Lett. B* **485** 208 (2000); hep-th/0005016
48. Luty M A, Porrati M, Rattazzi R *JHEP* (09) 029 (2003); hep-th/0303116

49. Nicolis A, Rattazzi R *JHEP* (06) 059 (2004); hep-th/0404159
50. Nicolis A, Rattazzi R, Trincherini E *Phys. Rev. D* **79** 064036 (2009); arXiv:0811.2197
51. Deffayet C, Esposito-Farèse G, Vikman A *Phys. Rev. D* **79** 084003 (2009); arXiv:0901.1314
52. de Rham C, Tolley A J *JCAP* (05) 015 (2010); arXiv:1003.5917
53. Goon G L, Hinterbichler K, Trodden M *Phys. Rev. Lett.* **106** 231102 (2011); arXiv:1103.6029
54. Goon G, Hinterbichler K, Trodden M *JCAP* (07) 017 (2011); arXiv:1103.5745
55. Kamada K et al. *Phys. Rev. D* **83** 083515 (2011); arXiv:1012.4238
56. Kobayashi T, Yamaguchi M, Yokoyama J *Prog. Theor. Phys.* **126** 511 (2011); arXiv:1105.5723
57. Babichev E, Deffayet C, Ziour R *Int. J. Mod. Phys. D* **18** 2147 (2009); arXiv:0905.2943
58. Deffayet C et al. *JCAP* (10) 026 (2010); arXiv:1008.0048
59. Kobayashi T, Yamaguchi M, Yokoyama J *Phys. Rev. Lett.* **105** 231302 (2010); arXiv:1008.0603
60. Pujolàs O, Sawicki I, Vikman A *JHEP* (11) 156 (2011); arXiv: 1103.5360
61. Charmousis C et al. *Phys. Rev. Lett.* **108** 051101 (2012); arXiv: 1106.2000
62. Charmousis C et al. *Phys. Rev. D* **85** 104040 (2012); arXiv:1112.4866
63. Deffayet C et al. *Phys. Rev. D* **84** 064039 (2011); arXiv:1103.3260
64. Padilla A, Sivanesan V *JHEP* (04) 032 (2013); arXiv:1210.4026
65. Sivanesan V, arXiv:1307.8081
66. Creminelli P, Nicolis A, Trincherini E *JCAP* (11) 021 (2010); arXiv:1007.0027
67. Creminelli P et al. *JHEP* (02) 006 (2013); arXiv:1209.3768
68. Hinterbichler K et al. *JCAP* (12) 030 (2012); arXiv:1209.5742
69. Hinterbichler K et al. *Phys. Rev. Lett.* **110** 241303 (2013); arXiv: 1212.3607
70. Rubakov V A *Phys. Rev. D* **88** 044015 (2013); arXiv:1305.2614
71. Elder B, Joyce A, Khoury J *Phys. Rev. D* **89** 044027 (2014); arXiv:1311.5889
72. Adams A et al. *JHEP* (10) 014 (2006); hep-th/0602178
73. de Rham C, Fasiello M, Tolley A J, arXiv:1308.2702
74. Arkani-Hamed N et al. *JHEP* (05) 074 (2004); hep-th/0312099
75. Sergienko A V, Rubakov V A *Theor. Math. Phys.* **173** 1709 (2012); *Teor. Mat. Fiz.* **173** 441 (2012); arXiv:0803.3163
76. Libanov M V, Rubakov V A, Sazhina O S, Sazhin M V *JETP* **108** 226 (2009); *Zh. Eksp. Teor. Fiz.* **135** 253 (2009); *Phys. Rev. D* **79** 083521 (2009); arXiv:0812.1459
77. Cline J M, Jeon S, Moore G D *Phys. Rev. D* **70** 043543 (2004); hep-ph/0311312
78. Buniy R V, Hsu S D H, Murray B M *Phys. Rev. D* **74** 063518 (2006); hep-th/0606091
79. Senatore L *Phys. Rev. D* **71** 043512 (2005); astro-ph/0406187
80. Creminelli P et al. *JHEP* (12) 080 (2006); hep-th/0606090
81. Buchbinder E I, Khoury J, Ovrut B A *Phys. Rev. D* **76** 123503 (2007); hep-th/0702154
82. Creminelli P, Senatore L *JCAP* (11) 010 (2007); hep-th/0702165
83. Khoury J et al. *Phys. Rev. D* **64** 123522 (2001); hep-th/0103239
84. Khoury J et al. *Phys. Rev. D* **65** 086007 (2002); hep-th/0108187
85. Padilla A, Saffin P M, Zhou S-Y *JHEP* (12) 031 (2010); arXiv: 1007.5424
86. Padilla A, Saffin P M, Zhou S-Y *JHEP* (01) 099 (2011); arXiv: 1008.3312
87. Padilla A, Saffin P M, Zhou S-Y *Phys. Rev. D* **83** 045009 (2011); arXiv:1008.0745
88. Hinterbichler K, Trodden M, Wesley D *Phys. Rev. D* **82** 124018 (2010); arXiv:1008.1305
89. Nishi S et al., arXiv:1401.1045
90. Levasseur L P, Brandenberger R, Davis A-C *Phys. Rev. D* **84** 103512 (2011); arXiv:1105.5649
91. Linde A, Mukhanov V *Phys. Rev. D* **56** R535 (1997); astro-ph/9610219
92. Enqvist K, Sloth M S *Nucl. Phys. B* **626** 395 (2002); hep-ph/0109214
93. Lyth D H, Wands D *Phys. Lett. B* **524** 5 (2002); hep-ph/0110002
94. Moroi T, Takahashi T *Phys. Lett. B* **522** 215 (2001); *Phys. Lett. B* **539** 303 (2002), Erratum; hep-ph/0110096
95. Dimopoulos K et al. *JHEP* (07) 053 (2003); hep-ph/0304050
96. Dvali G, Gruzinov A, Zaldarriaga M *Phys. Rev. D* **69** 023505 (2004); astro-ph/0303591
97. Kofman L, astro-ph/0303614
98. Dvali G, Gruzinov A, Zaldarriaga M *Phys. Rev. D* **69** 083505 (2004); astro-ph/0305548
99. Hinshaw G et al. (WMAP Collab.) *Astrophys. J. Suppl.* **208** 19 (2013); arXiv:1212.5226
100. Ade P A R et al. (Planck Collab.), arXiv:1303.5076
101. Rubakov V A, Osipov M A *JETP Lett.* **93** 52 (2011); *Pis'ma Zh. Eksp. Teor. Fiz.* **93** 56 (2011); Osipov M, Rubakov V, arXiv: 1007.3417
102. Rubakov V A *JCAP* (09) 030 (2009); arXiv:0906.3693
103. Hinterbichler K, Khoury J *JCAP* (04) 023 (2012); arXiv:1106.1428
104. Hinterbichler K, Joyce A, Khoury J *JCAP* (06) 043 (2012); arXiv:1202.6056
105. Libanov M, Rubakov V *JCAP* (11) 045 (2010); arXiv:1007.4949
106. Libanov M, Mironov S, Rubakov V *Prog. Theor. Phys. Suppl.* **190** 120 (2011); arXiv:1012.5737
107. Libanov M, Mironov S, Rubakov V *Phys. Rev. D* **84** 083502 (2011); arXiv:1105.6230
108. Libanov M, Ramazanov S, Rubakov V *JCAP* (06) 010 (2011); arXiv:1102.1390
109. Mironov S A, Ramazanov S R, Rubakov V A, arXiv:1312.7808
110. Qiu T et al. *JCAP* (10) 036 (2011); arXiv:1108.0593
111. Easson D A, Sawicki I, Vikman A *JCAP* (11) 021 (2011); arXiv:1109.1047
112. Cai Y-F, Easson D A, Brandenberger R *JCAP* (08) 020 (2012); arXiv:1206.2382
113. Osipov M, Rubakov V *JCAP* (11) 031 (2013); arXiv:1303.1221
114. Koehn M, Lehners J-L, Ovrut B A, arXiv:1310.7577
115. Lifshitz E M, Khalatnikov I M *Sov. Phys. Usp.* **6** 495 (1964); *Usp. Fiz. Nauk* **80** 391 (1963)
116. Lifshitz E M, Khalatnikov I M *Adv. Phys.* **12** 185 (1963)
117. Belinskii V A, Lifshitz E M, Khalatnikov I M *Sov. Phys. Usp.* **13** 745 (1971); *Usp. Fiz. Nauk* **102** 463 (1970)
118. Belinskii V A, Khalatnikov I M, Lifshitz E M *Adv. Phys.* **19** 525 (1970)
119. Belinskii V A, Lifshitz E M, Khalatnikov I M *Sov. Phys. JETP* **35** 838 (1972); *Zh. Eksp. Teor. Fiz.* **62** 1606 (1972)
120. Erickson J K et al. *Phys. Rev. D* **69** 063514 (2004); hep-th/0312009
121. Easson D A, Sawicki I, Vikman A *JCAP* (07) 014 (2013); arXiv:1304.3903
122. Dubovsky S L, Sibiryakov S M *Phys. Lett. B* **638** 509 (2006); hep-th/0603158
123. Hawking S W *Phys. Rev. D* **46** 603 (1992)
124. Landau L D, Lifshitz E M *The Classical Theory of Fields* (Oxford: Pergamon Press, 1975); Translated from Russian: *Teoriya Polya* (Moscow: Nauka, 1988)