Physics Letters B 843 (2023) 138031

Contents lists available at ScienceDirect

Physics Letters B

journal homepage: www.elsevier.com/locate/physletb

A new gravitational action for the trace anomaly

Gregory Gabadadze

Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY, 10003, USA

ARTICLE INFO

Article history:
Received 8 February 2023
Received in revised form 13 June 2023
Accepted 13 June 2023
Available online 15 June 2023
Editor: M. Trodden

ABSTRACT

The question of building a local diff-invariant effective gravitational action for the trace anomaly is reconsidered. General Relativity (GR) combined with the existing action for the trace anomaly is an inconsistent low energy effective field theory. This issue is addressed by extending GR into a certain scalar-tensor theory, which preserves the GR trace anomaly equation, up to higher order corrections. The extension introduces a new mass scale – assumed to be below the Planck scale – that governs four high dimensional terms in a local diff-invariant trace anomaly action. Such terms can be kept, while an infinite number of Planck-suppressed invariants are neglected. The resulting theory maintains two derivative equations of motion. In a certain approximation it reduces to the conformal Gallileon, which could have physical consequences.

1. Introduction and summary

General Relativity (GR), combined with a quantum field theory exhibiting the gravitational trace anomaly [1], and described by its effective action [2,3], is an inconsistent field theory, despite the existence of a local diff-invariant trace anomaly action (see, [4,5], and discussions below). Such a theory is strongly coupled at an arbitrarily low energy scale, as was shown in [6] in a different context. The inconsistency stems from the fact that the conformal part of a metric is not a propagating degree of freedom in classical GR.[1] This issue is fully relevant to GR coupled to the Standard Model (SM) of particle physics.

One way to avoid the inconsistency is to augment GR so that the conformal mode turns into a proper propagating degree of freedom, without spoiling observational consequences of GR, as it is proposed in this work. An alternative to the augmentation of GR would be to cancel the trace anomaly by introducing additional low energy degrees of freedom.

As a reminder, one is concerned with the invariance of a quantum field theory (QFT) – such as the SM – with respect to global scale transformations. At the classical level the scale invariance could be exact, as in the Maxwell theory, or it could be violated explicitly, as in a massive scalar theory, or in GR. The quantum theory violates the scale invariance generically, irrespective of whether the classical theory is or is not scale invariant. This violation appears in the trace of a stress-tensor, as it was first shown for gauge fields [7,8], and subsequently for a gravitational field [1]. The latter will be the focus on this work.

It is useful to build a low energy effective action that would incorporate quantum loops of the matter fields (see, [9] and references therein). The variation of such a quantum effective action would give rise to the equations of motion which capture the trace anomaly [2,3]. These equations could then be solved in various physical settings, notably in astrophysics and cosmology, with a benefit that the solutions would automatically contain quantum effects due to the trace anomaly.

Riegert constructed a local effective action, S_A, which captures the trace anomaly in its equations of motion [2] (Efim Fradkin and Tseytlin [3], built the same action in conformal gravity, practically at the same time [2]). This construction was done for a particular class of constrained fields, and hence was regarded by Riegert as breaking diff-invariance.

E-mail address: gg32@nyu.edu.

[1] A propagating degree of freedom is defined as a mode with proper quadratic time derivative and spatial gradient terms in Minkowski space-time, which is not removable by gauge transformations or by field redefinitions, nor it is restricted by constraints. The scale factor of the Fridmann-Lemaitre-Robertson-Walker (FLRW) metric is a conformal mode which has a ghost-like ("wrong sign") quadratic time derivative term in the GR action. This "wrong-sign" is crucial for FLRW cosmology. However, the scale factor is not a propagating degree of freedom because its spatial dynamics is constrained within GR; had it been a propagating mode, it would have led to unsurmountable instabilities because of its ghost-like quadratic time derivative term. The GR action supports only two propagating modes, the helicity ± 2 states of a massless graviton, often referred as the tensor modes.

https://doi.org/10.1016/j.physletb.2023.138031
0370-2693/

Komargodski and Schwimmer [4] showed that the same functional, S_A, but written in terms of unconstrained fields, gives a diff-invariant trace anomaly action. This finding placed the Riegert action on a solid footing, which it lacked for many years. Moreover, ref. [5], showed that the Riegert action emerges as a local diff-invariant Wess-Zumino term in a coset for a non-linearly realized conformal symmetry, broken by the scale anomaly.

These findings offer an important perspective: the GR action depends on the metric g and its derivatives. The metric g can formally be decomposed as $g = e^{2\sigma}\bar{g}$, and GR can be viewed as a theory non-linearly realizing a Weyl symmetry that shifts σ and conformally transforms $\bar{g}$, keeping g invariant. In GR this is a "spurious" symmetry since the above split of g is arbitrary, and the σ field can be gauged away by the very same Weyl transformations. However, the Riegert action is a local diff-invariant functional of σ and $\bar{g}$ which is not invariant under the "spurious" Weyl transformations [4]. Thus, the local diff-invariant action containing the Einstein-Hilbert and Riegert terms, both written in terms of σ and $\bar{g}$, can be viewed as an action non-linearly realizing the "spurious" Weyl symmetry [5].

The above arguments, however, suggests that something must be wrong: the field σ was "spurious" in GR, but becomes unremovable once GR is supplemented by the Riegert action. Indeed, in the GR action the metric field can absorb the kinetic term of σ, rendering it in the Riegert action infinitely strongly coupled at arbitrarily low energies [6].[2]

One way to resolve this problem is to augment the classical GR action and only then couple it to a quantum field theory. I will show that the following action

$$S_{R-\bar{R}} = M^2 \int d^4x\sqrt{-g}\, R \,-\, \bar{M}^2 \int d^4x\sqrt{-\bar{g}}\, \bar{R}\,, \tag{1}$$

with $\bar{R} \equiv R(\bar{g})$, $M = M_{\rm Pl}/\sqrt{2} >> \bar{M}$, and the two metric tensors connected as

$$g_{\mu\nu} = e^{2\sigma}\, \bar{g}_{\mu\nu}\,, \tag{2}$$

gives a viable low energy theory of gravity with a proper propagating conformal mode, that can be consistently coupled to the Riegert action. The "spurious" Weyl transformation, $\sigma \to \sigma + \beta(x)$, $\bar{g} \to e^{-2\beta(x)}\bar{g}$, with an arbitrary $\beta(x)$, would have made σ gauge-removable in GR, however, this is not a symmetry of the second term in the action (1), and therefore σ can't be gauged away. Furthermore, the action (1) possesses a global symmetry that neither the first nor the second term on the r.h.s. separately has (Section 2).

Combining (1) with the results of [2–4], the total effective action that captures the GR trace anomaly equation reads as follows:

$$S_{eff} = S_{R-\bar{R}} + S_A(\sigma, \bar{g})\,, \tag{3}$$

where S_A has the form [2–4]

$$S_A = a\int d^4x\sqrt{-\bar{g}}\left(\sigma\bar{E} - 4\bar{G}^{\mu\nu}\bar{\nabla}_\mu\sigma\bar{\nabla}_\nu\sigma - 4(\bar{\nabla}^2\sigma)(\bar{\nabla}\sigma)^2 - 2(\bar{\nabla}\sigma)^4\right) + c\int d^4x\sqrt{-\bar{g}}\sigma\bar{W}^2, \tag{4}$$

with the Euler (Gauss-Bonnet) invariant, $\bar{E} = \bar{R}^2_{\mu\nu\alpha\beta} - 4\bar{R}^2_{\mu\nu} + \bar{R}^2$, and the Weyl tensor squared, $\bar{W}^2 = \bar{R}^2_{\mu\nu\alpha\beta} - 2\bar{R}^2_{\mu\nu} + \bar{R}^2/3$. The action (4) emerges as a Wess-Zumino term in a $SO(2,4)/ISO(1,3)$ coset, which can be recast as a boundary term in 5D [5]; this distinguishes (4) from other terms in the effective field theory. Note that the a-terms in S_A belong to the general class of the Horndeski theories giving rise to second order equations of motion [11].

There are an infinite number of additional higher dimensional counter-terms supplementing (3). All these terms will be suppressed by respective powers of the scale $M = M_{\rm Pl}/\sqrt{2}$, or of higher scales, such as $M(M/\bar{M})^2$; they will be neglected. At the same time, S_A retains a finite number of terms which are suppressed by $\bar{M} << M$, or by certain geometric mean scales such as $(\bar{M}M^2)^{1/3}$, all of them lower than M.

It is due to the separation of scales between $\bar{M}$ and M that one can regard (3) as a meaningful low energy action obtained by "integrating out" quantum loops of a QFT. The coefficients a and c depend on numbers and representations of the low energy physical degrees of freedom [1]. The very same degrees of freedom, could also give rise to classical sources for the gravitational field. Hence the action (3) should be supplemented by the classical action for the fields representing those low energy degrees of freedom, but without quantizing those fields further.[3]

The signature used in this work is "mostly plus", $(-,+++)$. The Ricci tensor convention is as follows, $R_{\mu\nu} = \partial_\alpha\Gamma^\alpha_{\mu\nu} - \ldots$; Riegert, following Ref. [9], uses the "mostly minus" signature, and the opposite sign for the curvature tensor. Ref. [4] uses the "mostly minus" signature, the curvature convention opposite to Riegert's, and their τ and g are, respectively, $-\sigma$ and $\bar{g}$ here. The actions in refs. [2,3], [4], and in eq. (4) agree with each other after these different conventions are taken into consideration.

2. The $R - \bar{R}$ theory

Consider the action already quoted in the previous section

$$S_{R-\bar{R}} = M^2 \int d^4x\sqrt{-g}\, R \,-\, \bar{M}^2 \int d^4x\sqrt{-\bar{g}}\, \bar{R}\,, \tag{5}$$

[2] Komargodski and Schwimmer constructed the local diff-invariant Riegert action, and combined it with GR to prove the *a-theorem* (for an earlier work using a non-local Riegert action for the a-theorem, see [10]). They used the metric field for symmetry bookkeeping, but its dynamics was unimportant for the proof itself; hence the metric was frozen, and only the conformal mode (a dilaton) was utilized [4]. The issue discussed here does not affect the Komargodski-Schwimmer proof of the a-theorem, since their construction does not require a dynamical tensor field [4]. See more comments in Section 4.

[3] Not all quantum loops are proportional to the positive powers of $\hbar$ when massive fields are involved, some classical effects emerge from such loops [12]. Hence, one might worry that without considering the quantum loops some classical effects would be lost. However, keeping the respective classical fields in the effective action would enable one to retain those classical effects via nonlinear classical perturbation theory (or via exact classical or numerical solutions).

with the two metric tensors, g and $\bar{g}$, related to one another by (2). At the classical level, the above can be viewed as an action for one metric field – say, the metric g – and for the scalar σ, which gets its proper-sign kinetic term from the $\bar{R}$ term.[4]

Let us consider metric fluctuations above a flat background, $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$, and decompose them in a standard fashion according to the representations of the 3D rotation group (these approximate well fluctuations about an arbitrary nonsingular classical background at length scales much shorter than the characteristic curvature radius of the background):

$$h_{00} = 2n\,,\ \ h_{0j} = v_j + \partial_j u\,,\ \ h_{ij} = t_{ij} + \partial_i w_j + \partial_j w_i + 2\partial_i\partial_j\rho - 2\delta_{ij}\tau\,, \tag{6}$$

where $i, j = 1, 2, 3$, v_j and w_j are transverse three-vectors, t_{ij} is a transverse–traceless tensor, and τ is the conformal mode. Furthermore, one can choose the $u = 0, \rho = 0$ gauge.[5] Then, the scalar part of the action (5) – which decouples from the tensor and vector parts – equals to the space-time integral of the following expression:

$$2M^2\left(-3\dot{\tau}^2 + (\partial_j\tau)^2 - 2n\partial_j^2\tau\right) - 2\bar{M}^2\left(-3(\dot{\tau}+\dot{\sigma})^2 + (\partial_j(\tau+\sigma))^2 - 2(n+\sigma)\partial_j^2(\tau+\sigma)\right). \tag{7}$$

If the terms proportional to $\bar{M}^2$ were absent, the variation of (7) w.r.t. n would have given a constraint, $\partial_j^2\tau = 0$, rendering the conformal mode, τ, non-propagating. This however is no longer the case in (7): its variation w.r.t. n gives another constraint, $\partial_j^2\tau = \epsilon^2\partial_j^2\sigma/(1-\epsilon^2)$, that relates the conformal mode τ to the scalar σ, $\tau = \epsilon^2\sigma/(1-\epsilon^2)$ (here, $\epsilon = \bar{M}/M << 1$, and a zero mode of the Laplacian has been removed by choosing the appropriate spatial boundary conditions.) Substituting the latter into (7), one gets

$$-6M^2\frac{(1-\epsilon^2)}{\epsilon^2}\left(\partial_\mu\tau\right)\left(\partial^\mu\tau\right) = -6\bar{M}^2\frac{1}{(1-\epsilon^2)}\left(\partial_\mu\sigma\right)\left(\partial^\mu\sigma\right), \tag{8}$$

which is the kinetic term for the conformal mode with the proper sign. This is the key feature of the $R - \bar{R}$ theory.

It is convenient to rewrite (5) as follows:

$$S_{R-\bar{R}} = \int d^4x\sqrt{-g}\left(M^2 R(g) - \Phi^2 R(g) - 6(\nabla\Phi)^2\right), \tag{9}$$

where $\Phi \equiv \bar{M}e^{-\sigma}$, and the covariant derivative, ∇, is that of the metric g. Owing to the choice of the sign of the second term on the right hand side of (5), the kinetic term for the scalar Φ has the proper sign in (9); this sign will remain the same after complete diagonalization of the action (9), as shown in Appendix.

The scale $\bar{M}$ enters the action (9) only through the vacuum expectation value (VEV), $\langle\Phi\rangle = \bar{M}$. Since $\bar{M} << M$, this VEV is within the realm of the effective field theory. The last two terms in (9) non-linearly realize a Weyl symmetry which transforms g, and is explicitly broken by the Einstein-Hilbert term.

It is straightforward to check that the action (9) is invariant w.r.t. the following transformations of the fields

$$\Phi \to \Phi + \frac{\omega M(1-(\Phi/M)^2)}{1+\omega\Phi/M},\quad g_{\mu\nu} \to g_{\mu\nu}\frac{(1+\omega\Phi/M)^2}{1-\omega^2}, \tag{10}$$

where $\omega \equiv \tanh(\lambda)$, with λ being an arbitrary constant.[6]

Furthermore, there exists an invariant combination of the metrics g and $\bar{g}$

$$\hat{g}_{\mu\nu} = g_{\mu\nu} - \epsilon^2\,\bar{g}_{\mu\nu} = g_{\mu\nu}\left(1 - \frac{\Phi^2}{M^2}\right), \tag{11}$$

and if a QFT coupled to $\hat{g}$ in a conventional manner, $\mathcal{L}_{\rm SM} = -\frac{1}{4}\sqrt{-\hat{g}}\,\hat{g}^{\mu\nu}\hat{g}^{\alpha\beta}F_{\mu\alpha}F_{\nu\beta} + \cdots$, then it will also be invariant. Since one should require, $\epsilon \equiv (\bar{M}/M) << 1$, the coupling to the QFT is approximated by the coupling to the metric $g_{\mu\nu}$. The resulting classical equations approximate GR, as long as $\bar{M} << M$.[7]

The requirement of the invariance under (10) prohibits certain terms to be added to the action (9). For instance, an additional kinetic term for Φ, the cosmological constant, $\sqrt{-g}$, the quadratic mass term, $\sqrt{-g}\Phi^2$, and the quartic term, $\sqrt{-g}\Phi^4$, don't respect the symmetry (10), and are prohibited to enter the action with arbitrary coefficients. That said, there is a particular combination of the latter three terms which is invariant under (10)

$$\pm\Lambda^4\sqrt{-g}\left(1 - \frac{\Phi^2}{M^2}\right)^2, \tag{12}$$

[4] This would correspond to the Einstein frame. Section 4 will instead regard this action as a functional of $\bar{g}$ and σ, corresponding to the Jordan frame.

[5] The insertion of the $u = 0, \rho = 0$ gauge into the action does not lead to the loss on any constraints which might have emerged through the variation of the action w.r.t. u and ρ. This is so because in the quadratic action u and ρ enter only via a gauge invariant Lagrange multiplier, $n - \dot{u} - \ddot{\rho}$, and therefore, the variation w.r.t. n is what's capturing the most restrictive constraint discussed below.

[6] The symmetry transformations (10) may look mysterious, but their essence is simple: in Appendix it is shown that certain non-linear conformal transformations of g and Φ, bring the action (9) to that of GR coupled to a massless scalar field kinetic term, which is invariant w.r.t. the shift symmetry. The shift transformation, once rewritten in terms of the original variables, gives (10).

[7] One could consider a different approach in which QFT fields would couple to g, instead of $\hat{g}$, thus breaking the symmetry (10). This would result in new terms generated in the effective action due to the QFT loops, notably the mass term for the σ field would be induced. The latter would be proportional to some positive power of the UV scale of the QFT, suppressed by powers of M. Furthermore, σ would couple to the stress-tensor in the linearized approximation, providing a gravity-competing force at distances smaller than inverse mass of σ. One would then need to impose a constraint on $\bar{M}$ to suppress the coupling of σ to the stress-tensor. While this is a logical possibility, it would also lead to the trace anomaly equation being modified as compared with that of GR (such modifications can be kept small by imposing constraints on $\bar{M}$). The present work focuses on a scenario where the symmetry (10) is preserved by the QFT coupling, but can be adopted to the case when the matter couples to g.

with Λ being some arbitrary dimensionful constant. If the above term is included in the action then both the quadratic and quartic terms for Φ will be connected to the cosmological constant. The cosmological constant Λ will be tuned to zero (or be vanishingly small as compared to $\bar{M}$) and this will also nullify the quadratic and quartic terms in (12).

The term in (12) is nothing other than $\pm\Lambda^4\sqrt{-\hat{g}}$. One could also consider the terms, $\sqrt{-\hat{g}}\,\hat{R}$, and $\sqrt{-\hat{g}}(\hat{\nabla}\Phi)^2/(1-\Phi^2/M^2)^2$, to be added to the action (9), with arbitrary coefficients; these terms would individually respect the symmetry (10). However, only one linear combination of the above two terms retains the structure of the action (1) intact, which is necessary to preserve the same trace anomaly equation as one gets in GR (see Section 4). Therefore, in addition to the fine tuning of the cosmological constant, one needs to adopt another fine tuning of the relative coefficients between the above two terms.[8]

The action (1), and the relation (2), are invariant under the following "duality" transformations: $M \leftrightarrow \bar{M}$, $g \leftrightarrow -\bar{g}$, $\sigma \to -\sigma$. The latter leads to $\hat{g} \to \epsilon^{-2}\hat{g}$, rendering the following three terms invariant, $M^4\sqrt{-\hat{g}}$, $M^2\sqrt{-\hat{g}}\,\hat{R}$ and $\sqrt{-\hat{g}}(\hat{\nabla}\Phi)^2/(1-\Phi^2/M^2)^2$; furthermore, coupling of $\hat{g}$ to massless scalars, spinors, and vector fields can be made invariant by inserting appropriate powers of M in front of their kinetic terms. Therefore, the above "duality symmetry" does not help to avoid the need for the fine tuning of the two parameters discussed above.

The equations of motion that follow from (9) read:

$$(M^2-\Phi^2)\left(R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R\right)+\left(\nabla_\mu\nabla_\nu-g_{\mu\nu}\nabla^2\right)\Phi^2=6\left(\nabla_\mu\Phi\nabla_\nu\Phi-\frac{1}{2}g_{\mu\nu}(\nabla\Phi)^2\right),$$
$$-\Phi R+6\nabla^2\Phi=0\,, \tag{13}$$

where one can easily include the matter stress-tensor on the r.h.s. of the first equation. It is straightforward to show that a flat FLRW solution of GR is also a solution of (13), amended by the matter stress-tensor.

3. Toward the quantum effective theory

In general, one could quantize (9) as an effective low energy action (see [13], and references to earlier literature). However, there is no need for a general approach here. Instead, one could regard gravity as a dynamical classical field coupled to a QFT [9]. This is justified at energy scales much below the Planck scale.

Quantum loops of the QFT, including the loops by which one would integrate out possible Planck scale physics, will generate an infinite number of higher-dimensional terms that one needs to include in the effective action. To deal with the loop divergencies one could use, e.g., dimensional regularization.

Due to the symmetry (10) of the action (9) and of its coupling to a QFT, all the higher dimensional terms should also be invariant under (10).[9] Hence, one could identify all the symmetry preserving terms by writing all possible invariants in terms of the metric $\hat{g}$, expressed in terms of g and Φ. Furthermore, the metric $\hat{g}$ approximately equals to g, up to the corrections of the order of $1/M^2$, hence, one could just use g as an approximation. The goal of this section is to understand at what energy scales are those counter-terms significant.

To achieve this goal, let us introduce dimensionful fields

$$H_{\mu\nu}=M\left(g_{\mu\nu}-\eta_{\mu\nu}\right),\quad \Sigma=\bar{M}-\Phi\,, \tag{14}$$

and consider two different limits of the theory (9).

First, consider the limit $M\to\infty$, while $\bar{M}$ is fixed, and H and Σ are finite. It is then straightforward to see that the Lagrangian in (9) reduces to a free field theory of H and Σ

$$-\frac{1}{4}H^{\mu\nu}G_{\mu\nu}(H)-6(\partial\Sigma)^2\,, \tag{15}$$

where $G_{\mu\nu}(H)$ denotes the linearized Einstein tensor. The coupling to the stress-tensor, $T_{\mu\nu}$, is proportional to $H_{\mu\nu}T^{\mu\nu}/M$, up to additional corrections of the order of $\mathcal{O}\left(\bar{M}^2/M^3\right)$, and also vanishes in the limit as long as the stress-tensor is held finite. These considerations show that the loop-generated counter-terms in the full nonlinear theory should vanish in the $M\to\infty$ limit.

Second, consider the limit $M\to\infty$, $\bar{M}\to\infty$, with $\epsilon=(\bar{M}/M)<<1$ being fixed; the fields H and Σ are held finite in this limit, too. The resulting Lagrangian reads

$$-\frac{1-\epsilon^2}{4}H^{\mu\nu}G_{\mu\nu}(H)+2\epsilon\,\Sigma\,R_L(H)-6(\partial\Sigma)^2\,, \tag{16}$$

where $R_L(H)$ denotes the linearized Ricci scalar, and the coupling to the matter stress-tensor is proportional to $H_{\mu\nu}T^{\mu\nu}/M$, up to the additional corrections of the order of $\mathcal{O}\left(\bar{M}^2/M^3\right)$. The expression (16) represents a free field theory of a kinetically mixed tensor and scalar fields, and can easily be diagonalized by a linear conformal transformation of $H_{\mu\nu}$. Therefore, the counter terms of the full nonlinear theory ought to vanish in the second limit, too. The latter condition is more restrictive than the one obtained from the first limit above. It implies that there will not exist counter-terms proportional to $\bar{M}^p/M^q$, with $p\geq q$. In particular, the following counter terms

$$\bar{M}^2\left(\frac{\bar{M}}{M}\right)^n\sqrt{-g}\,\Sigma^2,\quad\left(\frac{\bar{M}}{M}\right)^m\sqrt{-g}\,\Sigma^4, \tag{17}$$

[8] One could relax this tuning somewhat to obtain the trace anomaly equation that would slightly differ from that of GR; this is easy to do but there is no urgency to pursue such extensions.

[9] I will neglect non-perturbative quantum gravity effects due to, e.g., black holes or worm holes, which are expected to violate global symmetries, and in particular to violate (10). Such violations are exponentially suppressed at low energies in the quasi-classical approximation [14]; they should be contrasted to the violation due to the trace anomaly, which is suppressed only by the powers of the scale $\bar{M}$, and will be included in the effective action.

will be absent for arbitrary integers $n \geq 0$ and $m \geq 0$ (i.e., such terms will not be generated with $n \geq 0$ and $m \geq 0$ even if they form combinations that are invariant under (10)).

Let us now summarize the results of the above discussions in terms of the fields of the action (9). One concludes that the counter-terms in (9) will be proportional to

$$\frac{R^3}{M^2}, \; \frac{R\Box R}{M^2}, \; \frac{\Sigma^2 R^2}{M^2}, \; \frac{\Sigma^2 R\Box R}{M^4}, \ldots \tag{18}$$

Each of these counter-terms will have corrections that make them invariant under (10), however, each of these corrections are higher order in $1/M^2$.

All of these counter terms are suppressed by M. Furthermore, due to the VEV, $\langle\Phi\rangle = \bar{M}$, some of the higher dimensional operators will end up depending on $\bar{M}$, too:

$$\frac{\bar{M}^2 R^2}{M^2}, \; \frac{\bar{M}^2 R\Box R}{M^4} \ldots \tag{19}$$

The lowest scale that suppresses the latter operators is $M(M/\bar{M})$, which is significantly higher than M, thanks to $\bar{M} << M$.

All the arguments above apply to conventional local counter-terms encountered in the conventional perturbative series expansion in powers of ∇^2/M^2 [13]. This however says nothing about possible non-local terms that may arise from the loops. Due to their nonlocal nature such terms could remain significant even when $\nabla^2 << M^2$. The trace anomaly introduces precisely such terms, which can then be rewritten as local term but at the expense of introducing a new field, σ. The new terms in the effective action are suppressed by the scale much smaller than M. Hence, it is meaningful to keep the trace-anomaly induced terms, but ignore all the other conventional higher dimensional counter terms suppressed by M and higher scales, as done below.

4. The effective action

The scale anomaly in the trace of a massless QFT stress-tensor reads [1]:

$$T^{\mu}_{\mu} = aE(g) + cW^2(g) + b\Box R(g)\,, \tag{20}$$

where the coefficients a and c depend on the field content of the theory, while b is arbitrary, and for that reason will not be included.[10] The goal is to find an effective action, $\tilde{S}_A(g)$, which incorporates the loop effects of the QFT so that the trace of its metric variation gives (20)

$$T^{\mu}_{\mu} = g^{\alpha\beta}\frac{2}{\sqrt{-g}}\frac{\delta\tilde{S}_A(g)}{\delta g^{\alpha\beta}} = aE(g) + cW^2(g)\,. \tag{21}$$

Riegert [2], argued that $\tilde{S}_A(g)$ cannot be written as a local diff-invariant functional if one uses only one field, g. Yet, the variation of the total action, $S_{GR}(g) + \tilde{S}_A(g)$, should give rise to a local equation for the trace anomaly, with the r.h.s. defined by (21).

To find the action, Riegert introduced a decomposition, $g_{\mu\nu} = e^{2\sigma}\bar{g}_{\mu\nu}$, with the metric $\bar{g}$ restricted to have a fixed determinant [2]. Using this decomposition, Riegert constructed a local action, $S_A(\bar{g},\sigma)$, such that the variation of $S_{GR}(e^{2\sigma}\bar{g}) + S_A(\bar{g},\sigma)$ w.r.t. σ gives the trace anomaly equation. The functional $S_A(\bar{g},\sigma)$ was regarded in [2] as breaking diff-invariance since the determinant of $\bar{g}$ had been fixed [2].

It is however straightforward to restore diff-invariance in $S_A(\bar{g},\sigma)$: one could use exactly the same action, $S_A(\bar{g},\sigma)$, but declare that the restriction on the determinant of $\bar{g}$ has been lifted. Conversely, such a diff-invariant $S_A(\bar{g},\sigma)$ would yield back Riegert's original local but non-invariant $S_A(\bar{g},\sigma)$, after gauge-fixing the determinant of $\bar{g}$.

Logically, Komargodski and Schwimmer [4] reconstructed Riegert's action $S_A(\bar{g},\sigma)$, as a functional of two fields, σ and $\bar{g}$, without assuming any constraint on $\bar{g}$, only requiring that the action reproduce the trace anomaly upon simultaneous Weyl transformations of σ and $\bar{g}$.

Furthermore, the same $S_A(\bar{g},\sigma)$ was obtained as a Wess-Zumino term in a coset construction for non-linearly realized conformal symmetry [5].

The above facts suggest that the local and invariant Riegert action, $S_A(\bar{g},\sigma)$, given in (4), should perhaps play a more fundamental role than it did in Riegert's construction, where it was merely used as an intermediate step toward motivating a non-local action.[11]

To this end, one could add the action (4) to the GR action expressed in terms of σ and $\bar{g}$, $S_{GR}(e^{2\sigma}\bar{g})$, and treat both σ and $\bar{g}$ as dynamical fields without assuming any constraint on $\bar{g}$. This theory, however, describes an infinitely strongly coupled system when restricted to its Minkowski background [6]. It is a strongly coupled theory with an unacceptably low strong coupling scale on any small curvature background. This is so because the theory, $S_{GR}(e^{2\sigma}\bar{g}) + S_A(\bar{g},\sigma)$, does not contain a kinetic term for σ. An apparent kinetic term for σ in $S_{GR}(e^{2\sigma}\bar{g})$ can be removed by a field redefinition of the metric, $\bar{g} = e^{-2\sigma}g$, rendering the σ field endowed with nonlinear interactions in $S_A(g,\sigma)$, but without a quadratic kinetic term. Such a theory is intractable as an effective field theory.[12]

[10] Both, b, and the gauge field trace anomaly [7,8], can easily be included in the effective action [2].

[11] Riegert's subsequent procedure of constructing a non-local but diff-invariant functional introduces a four derivative term in the action [2]. This rout unavoidably leads to a negative energy state, a ghost (or to violation to unitarity). It is also an unnecessary rout – as argued, a local and diff-invariant effective action for the trace anomaly does exist. More general actions with four-derivative terms were explored in [15], with some potentially interesting applications. Regretfully, these actions also have ghosts.

[12] This is not an issue for the work [4] insofar as $\bar{g}$ is not a dynamical field and is frozen to equal to the flat space metric; in the latter case the GR action gives a kinetic term for σ, albeit with a wrong overall sign. This sign can be flipped to the correct one by choosing a “wrong sign” GR term to start with; this does not cause a problem in [4] since the theory does not use the dynamics of tensor perturbations.

To bring some normalcy, one needs to introduce a kinetic term for σ. However, adding in any new term explicitly depending on σ – besides the ones already present in (4) – would spoil the recovery of the correct trace anomaly. One way out is to use the $R - \bar{R}$ theory as a functional of $\bar{g}$ and σ, where the new term, $\bar{R}$, does not depend on σ.

This leads one to explore a new action consisting of the $R - \bar{R}$ theory plus (4):

$$S_{eff}(\bar{g},\sigma) = M^2 \int d^4x\sqrt{-\bar{g}}\, e^{2\sigma}\left(R(\bar{g}) + 6\,(\bar{\nabla}\sigma)^2\right) - \bar{M}^2 \int d^4x\sqrt{-\bar{g}}\, R(\bar{g}) +$$
$$a\int d^4x\sqrt{-\bar{g}}\left(\sigma\bar{E} - 4\bar{G}^{\mu\nu}\bar{\nabla}_\mu\sigma\bar{\nabla}_\nu\sigma - 4(\bar{\nabla}^2\sigma)(\bar{\nabla}\sigma)^2 - 2(\bar{\nabla}\sigma)^4\right) + c\int d^4x\sqrt{-\bar{g}}\,\sigma\,\bar{W}^2\,. \tag{22}$$

This is a local diff-invariant functional of σ and $\bar{g}$. It can be viewed as an effective action for quantized σ and QFT.[13]

The classical matter fields – not shown in (22) – couple to the metric $\hat{g}_{\mu\nu} = g_{\mu\nu}(1 - \bar{M}^2 e^{-2\sigma}/M^2) \simeq g_{\mu\nu} = e^{2\sigma}\bar{g}_{\mu\nu}$. Variation of the action (22) with respect to σ, with a subsequent substitution of $\bar{g} = e^{-2\sigma}g$, gives an equation that does not depend on σ

$$\frac{\delta S_{eff}(\sigma,\bar{g})}{\delta\sigma}|_{\bar{g}_{\mu\nu}=e^{-2\sigma}g_{\mu\nu}} = 0\,, \;\Rightarrow\; 2M^2R = aE(g) + cW^2(g)\,, \tag{23}$$

which is exactly the trace-anomaly equation.

Now that the tensor $\bar{g}$ is an unconstrained dynamical field, one should vary (22) w.r.t. $\bar{g}$, too. This variation will give ten modified Einstein's equations. One can show that the trace of these equations does not coincide with the equation (23). Thus, there are eleven independent equations: one equation of motion of σ, and ten equations for ten components of the symmetric tensor $\bar{g}$. By determining σ and $\bar{g}$ from these equations, one can determine $\hat{g} \simeq g$, which is the metric experienced by the classical matter fields. The modified Einstein equations approximate well the conventional equations as long as $\bar{M} << M$, and as long as the stress-tensor is much smaller than $\bar{M}^4$.

Alternatively, one could rewrite the ten equations for $\bar{g}$ as ten equations for g, which would also depend on σ. The σ equation, on the other hand, can entirely be kept in terms of g (23). Thus, there would still be eleven equations for eleven unknowns, g and σ.

The theory given by (22) is strongly coupled at the scale $\bar{M}$. This can be seen in the flat space expansion of (22) obtained either by taking the limit $M \to \infty$, or by the substitution

$$g_{\mu\nu} = \eta_{\mu\nu}\left(1 - \Phi^2/M^2\right)^{-1}, \quad \bar{g}_{\mu\nu} = e^{-2\sigma}\left(1 - \Phi^2/M^2\right)^{-1}\eta_{\mu\nu}. \tag{24}$$

In either case, the theory reduces to a conformal Galileon of the canonically normalized field, $\pi \equiv \sigma\bar{M}$ [16], with the nonlinear Galileon terms suppressed by the scale $\bar{M}$

$$S_{eff}|_{\bar{g}_{\mu\nu}\simeq e^{-2\pi/\bar{M}}\eta_{\mu\nu}} \simeq \int d^4x\left(-6e^{-2\pi/\bar{M}}(\partial\pi)^2 + 4a\left(-\frac{\partial^2\pi(\partial\pi)^2}{\bar{M}^3} + \frac{(\partial\pi)^2(\partial\pi)^2}{2\bar{M}^4}\right) + \dots\right), \tag{25}$$

where the dots stand for sub-leading terms suppressed by powers of M.

Below the scale $\bar{M}$ the full theory (22) is weakly coupled and describes three degrees of freedom – two helicity states of a massless graviton, and one massless scalar, $\pi \equiv \sigma\bar{M}$. The later becomes strongly coupled at the scale of the order of $\bar{M}$, as seen from the Galileon terms in (25). The Lagrangian (25) makes it clear that the theory without the $\bar{M}^2\sqrt{-\bar{g}}\bar{R}$ term is untenable – taking $\bar{M} \to 0$ leaves the nonlinear terms in (25) infinitely strongly coupled.

The additional scalar, σ, does not couple to the stress tensor in the linearized approximation in Minkowski space, therefore there is no fifth force constraint. It can couple to matter on curved backgrounds. Physical consequences of the action (22) will be studies elsewhere.

Comments added: After this article appeared in the arXiv, Pedro Fernandes communicated that the action which he derived in [17] using different principles, is equivalent to the action given above in (22) (up to some obvious terms). Indeed, if one transforms the action (22) from the Jordan frame to the Einstein frame, and then changes the field σ to $-\phi$, one obtains the respective terms in the action (14) of [17]. Moreover, the “duality” discussed in the present work in Section 2, transforms the action (22) into the action (14) of ref. [17].

The action in ref. [17] was derived based on purely classical principles, by answering the question: what is the most general scalar-tensor theory of gravity with the second order equations of motion that has the scalar equation invariant under the conformal transformations, even if the action is not conformally invariant? Furthermore, the invariance of the scalar equation of motion enabled Fernandes to find exact analytic solutions of the theory [17].

It is encouraging that such a different principle gave rise to the action of [17], which was derived in the present work based on the consistency of the effective field theory of gravity with the quantum trace anomaly. The anomaly perspective of the present work is essential: its outcome is that GR needs to be amended by a scalar degree of freedom – as it is done in the trace anomaly action (22) – in order for GR to be consistently coupled to a QFT with the trace anomaly. If so, then there has to exist a new physical scale, $\bar{M} << M_{\rm Pl}$, governing the interactions of the scalar with itself, and governing – along with $M_{\rm Pl}$ – its interactions with the SM fields. An alternative to the above is to cancel the trace anomaly by introducing other light degrees of freedom.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[13] Note that the Riegert action in the second line in (22) contains the Galileon terms that are not suppressed by M. These can be understood as terms obtained by “integrating in” the σ field to make the otherwise nonlocal anomaly action be local. In such a case, not all the terms containing σ should be suppressed by M. Conversely, integrating out σ would lead to nonlocal terms, all of them suppressed by M.

Data availability

No data was used for the research described in the article.

Acknowledgements

I'd like to thank Massimo Porrati, David Spergel, and Giorgi Tukhashvili for useful discussions, and Pedro Fernandes and Arkady Tseytlin for helpful communications. The work was supported in part by NSF grant PHY-2210349.

Appendix A

A complete diagonalization of (9) can be done by using the following conformal transformation

$$g_{\alpha\beta} = \hat{g}_{\alpha\beta}\left(1 - \frac{\Phi^2}{M^2}\right)^{-1}. \tag{26}$$

It brings the action for the metric $\hat{g}$ to the canonical form

$$S_{R-\bar{R}} = \int d^4x\sqrt{-\hat{g}}\left(M^2\, R(\hat{g}) - \frac{6\,(\hat{\nabla}\Phi)^2}{(1-\Phi^2/M^2)^2}\right). \tag{27}$$

The above conformal transformation is non-singular, and (27) is sensible, as long as

$$\Phi = \bar{M}e^{-\sigma} << M, \quad \Rightarrow \quad \sigma >> -ln(M/\bar{M}). \tag{28}$$

In addition, the effective theory is valid if $|\sigma| << 1$, which is a stronger constraint.

The scalar field action in (27) can be simplified further: it can be reduced to a free field interacting with gravity, thanks to the following non-linear field redefinition

$$U = \left(\frac{1+\Phi/M}{1-\Phi/M}\right)^{1/2}. \tag{29}$$

The resulting action reads:

$$S_{R-\bar{R}} = M^2\int d^4x\sqrt{-\hat{g}}\left(R(\hat{g}) - 6\,(U^{-1}\hat{\nabla}U)^2\right). \tag{30}$$

The latter makes the symmetries of the sigma model more explicit: the Lagrangian is invariant w.r.t. the rescaling

$$U \to e^{\lambda}\, U\,, \tag{31}$$

where λ is an arbitrary constant. This symmetry is the shift symmetry of the scalar field $\ln(U)$ that has only a kinetic term in (30).

Note that $\hat{g}$ and U are related to $\bar{g}$ and σ through nonlinear transformations. However, the path integral $Z(\hat{g})$ does not equal to the path integral $Z(g)$. In other words, the nonlinear conformal transformation from $\bar{g}$ to $\hat{g}$, and the quantization procedure for the scalar and QFT, do not commute with one another. If one were to start with (30) and combine it with the Riegert action for $\hat{g}$, one would get an infinitely strongly coupled theory. Instead, the Riegert action that needs to be added to (30) can be obtained from the Riegert action for $\bar{g}$, by using the respective nonlinear conformal transformation from $\bar{g}$ to $\hat{g}$. The obtained action will be strongly coupled at the scale $\bar{M}$.

References

[1] D.M. Capper, M.J. Duff, Phys. Lett. A 53 (1975) 361;
M.J. Duff, Nucl. Phys. B 125 (1977) 334–348;
S.M. Christensen, M.J. Duff, Phys. Lett. B 76 (1978) 571.
[2] R.J. Riegert, Phys. Lett. B 134 (1984) 56–60.
[3] E.S. Fradkin, A.A. Tseytlin, Phys. Lett. B 134 (1984) 187.
[4] Z. Komargodski, A. Schwimmer, J. High Energy Phys. 12 (2011) 099, arXiv:1107.3987 [hep-th].
[5] G. Gabadadze, G. Tukhashvili, Phys. Rev. D 102 (2) (2020) 024054, arXiv:2005.01729 [hep-th].
[6] J. Bonifacio, K. Hinterbichler, L.A. Johnson, Phys. Rev. D 102 (2) (2020) 024029, arXiv:2004.10716 [hep-th].
[7] R.J. Crewther, Phys. Rev. Lett. 28 (1972) 1421.
[8] M.S. Chanowitz, J.R. Ellis, Phys. Lett. B 40 (1972) 397–400.
[9] N.D. Birrell, P.C.W. Davies, Quantum Fields in Curved Space, Cambridge University Press, 1994.
[10] D. Anselmi, Ann. Phys. 276 (1999) 361–390, arXiv:hep-th/9903059 [hep-th].
[11] G.W. Horndeski, Int. J. Theor. Phys. 10 (1974) 363–384.
[12] B.R. Holstein, J.F. Donoghue, Phys. Rev. Lett. 93 (2004) 201602, arXiv:hep-th/0405239 [hep-th].
[13] J.F. Donoghue, Phys. Rev. D 50 (1994) 3874–3888, arXiv:gr-qc/9405057 [gr-qc], arXiv:gr-qc/9512024 [gr-qc], arXiv:1209.3511 [gr-qc].
[14] R. Kallosh, A.D. Linde, D.A. Linde, L. Susskind, Phys. Rev. D 52 (1995) 912–935, arXiv:hep-th/9502069 [hep-th].
[15] I. Antoniadis, E. Mottola, Phys. Rev. D 45 (1992) 2013–2025;
I. Antoniadis, P.O. Mazur, E. Mottola, Phys. Rev. D 55 (1997) 4770–4784, arXiv:hep-th/9509169 [hep-th];
I. Antoniadis, P.O. Mazur, E. Mottola, Phys. Rev. Lett. 79 (1997) 14–17, arXiv:astro-ph/9611208 [astro-ph];
E. Mottola, J. High Energy Phys. 11 (2022) 037, arXiv:2205.04703 [hep-th].
[16] A. Nicolis, R. Rattazzi, E. Trincherini, Phys. Rev. D 79 (2009) 064036, arXiv:0811.2197 [hep-th].
[17] P.G.S. Fernandes, Phys. Rev. D 103 (10) (2021) 104065, arXiv:2105.04687 [gr-qc].