Quantum gravity constraints from unitarity and analyticity

Brando Bellazzini, Clifford Cheung, and Grant N. Remmen

Quantum Gravity Constraints from Unitarity and Analyticity

Brando Bellazzini,[1] Clifford Cheung,[2] and Grant N. Remmen[2]

[1]*Institut de Physique Théorique, Université Paris Saclay, CEA, CNRS, F-91191 Gif-sur-Yvette, France*
[2]*Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91125, USA*

Abstract

We derive rigorous bounds on corrections to Einstein gravity using unitarity and analyticity of graviton scattering amplitudes. In $D \geq 4$ spacetime dimensions, these consistency conditions mandate positive coefficients for certain quartic curvature operators. We systematically enumerate all such positivity bounds in $D = 4$ and $D = 5$ before extending to $D \geq 6$. Afterwards, we derive positivity bounds for supersymmetric operators and verify that all of our constraints are satisfied by weakly-coupled string theories. Among quadratic curvature operators, we find that the Gauss-Bonnet term in $D \geq 5$ is inconsistent unless new degrees of freedom enter at the natural cutoff scale defined by the effective theory. Our bounds apply to perturbative ultraviolet completions of gravity.

1 Introduction

Low-energy effective theory describes quanta interacting indirectly through kinematically inaccessible states. The dynamics are characterized by an effective action that typically includes all interactions permitted by symmetry with couplings of order unity. However, in certain cases, self-consistency at long distances imposes non-trivial constraints on the coefficients of effective

e-mail: `brando.bellazzini@cea.fr`, `clifford.cheung@caltech.edu`, `gremmen@theory.caltech.edu`

operators. This is famously true in the theory of pions, where the operator coefficients of certain higher-derivative operators are required to be strictly positive to ensure causal particle propagation together with unitarity and analyticity of scattering amplitudes at complex momenta [1–5].

As low-energy criteria, causality, unitarity, and analyticity impose constraints that are independent of the detailed ultraviolet dynamics. Consequently, these consistency conditions offer special utility in the context of quantum gravity, where the ultraviolet completion is not known with certainty. For instance, such bounds have been derived for the effective theory of gravitons and photons [6], where consistency necessitates large charge-to-mass ratios precisely of the form of the weak gravity conjecture [7].

Notably, a proper diagnosis of causality violation in curved spacetime is subtle since particle propagation can be locally superluminal even in healthy theories. For example, it has long been known that photons with certain polarizations travel superluminally in the vicinity of a black hole in the effective theory of photons and gravitons describing our actual universe [8]. Instead, a more global measure of causality, *e.g.*, the existence of closed timelike curves, is necessary to establish a true pathology. On the other hand, unitarity and analyticity offer alternative criteria that are mathematically rigorous and applicable in the flat-space limit.

In this paper, we systematically derive new constraints on curvature corrections in gravity from unitarity and analyticity. The graviton effective theory is described by the action[1]

$$S = \int \mathrm{d}^D x \, \sqrt{-g} \, \sum_{n=1}^{\infty} \mathcal{L}_n, \tag{1}$$

where $\mathcal{L}_n$ are contributions to the action entering at order $2n$ in the derivative expansion and

$$\mathcal{L}_1 = \frac{R}{2\kappa^2} \qquad \text{and} \qquad \mathcal{L}_2 = \lambda(R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2) \tag{2}$$

are the Einstein-Hilbert and Gauss-Bonnet terms. We assume the Gauss-Bonnet form for $\mathcal{L}_2$ throughout since this is the unique ghost-free quadratic curvature invariant [9, 10] in D dimensions. Moreover, we restrict our analysis to effective theories in which $\mathcal{L}_4$ takes the form

$$\mathcal{L}_4 = \sum_{i=1}^{7} c_i \mathcal{O}_i, \tag{3}$$

---

[1]We work in mostly + signature and write $\kappa = \sqrt{8\pi G}$, $R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu}$, and $R^{\mu}{}_{\nu\rho\sigma} = \partial_\rho \Gamma^{\mu}{}_{\nu\sigma} + \cdots$.

expressed in terms of the minimal basis of quartic Riemann operators[2] in Ref. [12],

$$\mathcal{O}_1 = R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma} R^{\alpha\beta\gamma\delta} R_{\alpha\beta\gamma\delta} \qquad \mathcal{O}_2 = R^{\mu\nu\rho\sigma} R_{\mu\nu\rho}{}^{\delta} R^{\alpha\beta\gamma}{}_{\sigma} R_{\alpha\beta\gamma\delta}$$

$$\mathcal{O}_3 = R^{\mu\nu\rho\sigma} R_{\mu\nu}{}^{\alpha\beta} R_{\alpha\beta}{}^{\gamma\delta} R_{\rho\sigma\gamma\delta} \qquad \mathcal{O}_4 = R^{\mu\nu\rho\sigma} R_{\mu\nu}{}^{\alpha\beta} R_{\rho\alpha}{}^{\gamma\delta} R_{\sigma\beta\gamma\delta}$$

$$\mathcal{O}_5 = R^{\mu\nu\rho\sigma} R_{\mu\nu}{}^{\alpha\beta} R_{\rho}{}^{\gamma}{}_{\alpha}{}^{\delta} R_{\sigma\gamma\beta\delta} \qquad \mathcal{O}_6 = R^{\mu\nu\rho\sigma} R_{\mu}{}^{\alpha}{}_{\rho}{}^{\beta} R_{\alpha}{}^{\gamma}{}_{\beta}{}^{\delta} R_{\nu\gamma\sigma\delta} \tag{4}$$

$$\mathcal{O}_7 = R^{\mu\nu\rho\sigma} R_{\mu}{}^{\alpha}{}_{\rho}{}^{\beta} R_{\alpha}{}^{\gamma}{}_{\nu}{}^{\delta} R_{\beta\gamma\sigma\delta}.$$

Note that linear dependences arise among operators as the dimension D of spacetime decreases. At quadratic order, $\mathcal{L}_2$ is unphysical in $D \leq 3$, a total derivative in $D = 4$, and dynamical in $D \geq 5$. Meanwhile, at quartic order, the number of algebraically independent operators $\mathcal{O}_i$ in $D = 4, 5, 6, 7, 8$ is $2, 4, 6, 6, 7$, respectively [12], with one linear combination—the eight-dimensional Euler density—a total derivative in $D = 8$ and hence dynamical only in $D \geq 9$ [13].

Our analysis hinges on the on-shell four-point graviton scattering amplitude, M, whose forward limit is intimately linked to the total cross-section by well-known analyticity arguments [1, 5]. By marginalizing over the external graviton polarizations, we can then systematically derive a rigorous and inclusive set positivity bounds on the coefficients of operators in the graviton effective action. Throughout, we assume a perturbative ultraviolet completion of gravity, so there exists a well-defined expansion in $\hbar$.

We begin with an analysis of quartic curvature corrections, proving that in $D = 4$ the coefficients of the $(R^2)^2$ and $(R\tilde{R})^2$ operators are positive. Our results precisely match those of Ref. [14], which derived bounds from the condition of locally subluminal graviton propagation. We then generalize our arguments to $D = 5$ and $D \geq 6$. Subsequently, we obtain positivity constraints on supersymmetric operators in general D, which in the literature are sometimes denoted $t_8 t_8 R^4$ and $t_8 (R^2)^2$. As a consistency check, we verify that all our constraints are satisfied in the bosonic, type II, and heterotic string theories.

Moving on to quadratic curvature corrections, we argue that unitarity and analyticity exclude theories in which $\lambda \gg 1$ with no new degrees of freedom present at or below the mass scale $\Lambda \sim |\lambda \kappa^2|^{-1/2}$, the natural cutoff associated with the Gauss-Bonnet term and the derivative expansion. Our results precisely accord with those of Maldacena et al. [15], who demonstrated that this class of theories is inconsistent with global causality.

The remainder of this paper is organized as follows. In Sec. 2, we review the arguments of Ref. [1] whereby unitarity and analyticity impose rigorous positivity bounds on operator coefficients in effective theories. Afterwards, in Sec. 3 we apply these methods to establish the

---

[2]Applying leading-order equations of motion to $\mathcal{L}_n$ is equivalent to a field definition modulo new terms generated in $\mathcal{L}_{n+1}$. Repeating this procedure at each order, we can freely impose $R = R_{\mu\nu} = 0$ in a pure gravity theory [11].

positivity of certain coefficients of quartic curvature operators, starting in $D = 4$ and $D = 5$ and then generalizing to $D \geq 6$. We then apply our bounds to supersymmetric theories and string theories. Finally, we study quadratic curvature operators in Sec. 4 and conclude in Sec. 5.

2 Analyticity Argument

In this section we review how operator coefficients in effective field theories are constrained by the analyticity of scattering amplitudes at complex momenta. Our analysis follows that of Ref. [1], which derived positivity bounds on operator coefficients by relating the low-energy limit of forward amplitudes to strictly positive integrals of cross-sections.

Our object of interest is the on-shell amplitude M describing four-point graviton scattering in D dimensions. Here, the choice of the external polarizations is built into the functional form of M, as is the case for helicity amplitudes in $D = 4$. From this viewpoint, helicity is just a quantum number labeling the external states, no different from baryon or lepton number. Sometimes it will be useful to view M as a function of the external particle labels, $M = M(1, 2, 3, 4)$, and other times as a function of the kinematic invariants, $M = M(s, t, u)$, where

$$s = -(k_1 + k_2)^2, \quad t = -(k_1 + k_3)^2, \quad u = -(k_1 + k_4)^2, \tag{5}$$

working in the convention where all momenta are incoming, so $k_1 + k_2 + k_3 + k_4 = 0$.

To derive constraints from analyticity, we will be interested in scattering amplitudes that are simultaneously forward and invariant under crossing in the t-channel. Formally, t-channel crossing symmetry implies invariance under swapping particle labels $1 \leftrightarrow 3$ or $2 \leftrightarrow 4$ while leaving the functional form of M—which encodes the polarization choices—fixed, so

$$M(1, 2, 3, 4) = M(3, 2, 1, 4) = M(1, 4, 3, 2) = M(3, 4, 1, 2). \tag{6}$$

For external gravitons, crossing symmetry is ensured if the exchanged states are indistinguishable. This happens in $D = 4$ if the states have identical helicity and more generally in D dimensions if the states have the same polarization relative to their momenta. Mathematically, M is crossing symmetric provided the momentum and polarization of particle 1 are related by an improper Lorentz transformation to those of particle 3, and likewise for particles 2 and 4. In terms of kinematic invariants, a crossing symmetric amplitude then satisfies

$$M(s, t) = M(-s - t, t), \tag{7}$$

where the momenta are swapped but the polarizations relative to momenta—which in $D = 4$ are the helicities—are untouched.

4

Meanwhile, the forward limit of the amplitude, $M(s, t \to 0)$, corresponds to an identification of particles $1 \leftrightarrow 3$ and $2 \leftrightarrow 4$. This is achieved simultaneously with crossing symmetry if we restrict to following kinematic regime:

$$\begin{matrix} \text{forward and} \\ \text{crossing symmetric} \end{matrix} \implies (k_3, \epsilon_3) \leftrightarrow (-k_1, \epsilon_1) \quad \text{and} \quad (k_4, \epsilon_4) \leftrightarrow (-k_2, \epsilon_2), \tag{8}$$

where ϵ_1 and ϵ_2 are real linear polarizations. We choose a basis of linear polarizations because an amplitude with fixed external circular polarizations cannot be simultaneously crossing symmetric and forward.

The forward and crossing symmetric amplitude, $M(s, t \to 0)$, can then be expanded in a power series in s and t. While analytic singularities in s or t arise, their form is severely restricted by the locality of the underlying theory. As noted earlier, we assume throughout a perturbative ultraviolet completion of gravity, so we are justified in considering only the leading contribution in the $\hbar$ expansion, *i.e.*, tree-level exchange.

At tree level, analytic singularities in kinematic invariants enter at worst as simple poles. Moreover, a t-channel singularity in the forward limit can only arise from non-local terms corresponding to graviton exchange induced by the leading Einstein-Hilbert interactions, so the general form for the forward amplitude is

$$M(s, t \to 0) = \sum_{n=0}^{\infty} f_n s^n + \mathcal{O}(s^2/t). \tag{9}$$

The first term is regular, as it is generated by heavy particle exchange, while the second term is singular because it comes from t-channel graviton exchange scaling as $\sim s^2/t$. The form of Eq. (9) together with the crossing symmetry relation of Eq. (7) implies that

$$M(s, t \to 0) = M(-s, t \to 0) + \mathcal{O}(s), \tag{10}$$

where the first term arises because the limit of a regular function is the function evaluated at the limit of its arguments, while the second term is a residual contribution from applying crossing symmetry to the singular $\mathcal{O}(s^2/t)$ contribution.

The parameters f_n depend on the coefficients of operators in the effective action of the low-energy theory. To determine analyticity constraints, we consider the order n residue of $M(s, t \to 0)$ in the complex s plane, yielding

$$f_n = \frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{\mathrm{d}s}{s^{n+1}} \left[M(s, t \to 0) + \mathcal{O}(s^2/t) \right], \tag{11}$$

where $\mathcal{C}$ denotes a small contour encircling the origin.

As previously noted [1,6], the $\mathcal{O}(s^2/t)$ singular contribution is formally infinite in the strictly forward limit and therefore a major obstacle to deriving bounds from analyticity. Nevertheless, for $n \neq 2$ this term is eliminated by the residue theorem. While forward singularities of order s^n/t can arise from loop-level graviton exchange diagrams, we are working at leading order in the $\hbar$ expansion so these contributions are formally subdominant. On the other hand, $n = 2$ is more subtle, but we will show that in certain parameter regimes the $\mathcal{O}(s^2/t)$ term can be subdominant to the rest of the amplitude, allowing for a bound to be placed. In any case, we leave a detailed discussion of these issues for later sections and for now simply drop the $\mathcal{O}(s^2/t)$ contribution. Terms subleading in the forward limit of the Einstein-Hilbert amplitude must by power counting go as $\mathcal{O}(s)$ and will thus be eliminated in the contour integral for all $n \geq 2$.

By Cauchy's theorem, we can blow up $\mathcal{C}$ into a new contour $\mathcal{C}'$ that runs just above and below the real s axis, plus a circular boundary contour at infinity, yielding

$$f_n = \frac{1}{2\pi i} \left(\int_{-\infty}^{-s_0} + \int_{s_0}^{\infty} \right) \frac{\mathrm{d}s}{s^{n+1}} \mathrm{Disc}[M(s, t \to 0)] + \text{boundary integral}, \tag{12}$$

where s_0 is any scale above zero and below the first massive threshold in the ultraviolet completion. We note that $\mathrm{Disc}[M(s, t \to 0)] = M(s + i\epsilon, t \to 0) - M(s - i\epsilon, t \to 0)$. By the Schwarz reflection principle, $M(s^*, t \to 0) = M(s, t \to 0)^*$, so

$$\mathrm{Disc}[M(s, t \to 0)] = 2i\mathrm{Im}[M(s, t \to 0)]. \tag{13}$$

The crossing symmetry relation in Eq. (10) then implies that

$$\mathrm{Disc}[M(-s, t \to 0)] = -\mathrm{Disc}[M(s, t \to 0)], \tag{14}$$

dropping the $\mathcal{O}(s)$ term that arose from the $\mathcal{O}(s^2/t)$ singularity.

Throughout this paper we assume that $|M(s)|$ is less divergent than $|s|^n$ at large s so that the boundary term in Eq. (12) can be dropped.[3] This is a physically reasonable assumption applicable to any ultraviolet completion in which the large s behavior of the amplitude at fixed finite physical $t \ll s$ grows more slowly in s than the Einstein-Hilbert contribution, which scales as s^2/t. A theory that fails this criterion would actually have worse ultraviolet behavior than Einstein gravity. Operationally, this translates into the assumption that $|M(s)|$ grows more slowly than $|s|^2$ at large s. For example, this can be verified explicitly in the Regge behavior of string theory amplitudes, which scale as $s^{\alpha(t)}/t$ where $\alpha(t) < 2$ for $t < 0$ [1].

---

[3]Strictly speaking, positivity bounds only require that the boundary term be non-negative, which is sometimes true given specific assumptions about the ultraviolet [16]. We do not consider this possibility here.

Combining Eq. (12) with Eqs. (13) and (14) yields

$$f_n = \frac{(-1)^n + 1}{\pi} \int_{s_0}^{\infty} \frac{\mathrm{d}s}{s^{n+1}} \mathrm{Im}[M(s, t \to 0)]. \tag{15}$$

For n odd, this result trivially implies $f_n = 0$ as required by crossing symmetry of M, while for n even, it imposes a positivity condition. In particular, we use the optical theorem to write $\mathrm{Im}[M(s, t \to 0)] = s\sigma(s)$, where $\sigma(s)$ is the total cross-section.[4] Crucially, in an interacting theory with new heavy states, $\sigma(s)$ is strictly positive, so

$$f_n = \frac{2}{\pi} \int_{s_0}^{\infty} \mathrm{d}s \frac{\sigma(s)}{s^n} > 0, \tag{16}$$

thus establishing positivity of f_n for even n.

Here f_n corresponds to the s^n contribution to the low-energy amplitude, which is proportional to the operator coefficients of $\mathcal{L}_n$. By power counting, we know that the low-energy amplitude can be expanded in powers of Mandelstam variables, so

$$M = \sum_{n=1}^{\infty} M_n, \tag{17}$$

where the leading contribution arises from the Einstein-Hilbert action, which in the forward limit gives an amplitude

$$M_1(s, t \to 0) = -\epsilon_{1\mu\nu} \epsilon_1^{\mu\nu} \epsilon_{2\rho\sigma} \epsilon_2^{\rho\sigma} \frac{\kappa^2 s^2}{t} + \mathcal{O}(s), \tag{18}$$

where the $\mathcal{O}(s)$ terms are regular in the forward limit. The remaining contributions M_n are generated by $\mathcal{L}_n$. In the subsequent sections, we derive precise analyticity bounds for the quartic and quadratic curvature corrections, $\mathcal{L}_4$ and $\mathcal{L}_2$.

3 Bounds on Quartic Curvature Corrections

In this section we derive bounds on $\mathcal{L}_4$, which encodes quartic curvature corrections to Einstein gravity. The leading contributions from $\mathcal{L}_4$ are quartic graviton vertices, which contribute to graviton scattering amplitudes via contact interactions. Since these corrections are free

---

[4]While the total cross-section diverges in the presence of a t-channel singularity, $\mathrm{Im}(s, t \to 0)$ and by extension $\sigma(s) = \mathrm{Im}(s, t \to 0)/s$ are really proxies for the finite sum over all residues of heavy states in the complex s plane. By factorization, each contribution is positive since $M(hh \to hh) \sim -M(hh \to X)M(X \to hh)/(s - m^2 + i\epsilon)$ on the s-channel resonance of a massive state X of mass m.

from kinematic singularities, their forward limit is regular. Thus, to obtain a forward, crossing symmetric amplitude, we simply set $t = 0$, $\epsilon_3 = \epsilon_1$, and $\epsilon_4 = \epsilon_2$, as derived in Eq. (8).

After a lengthy calculation, we compute the quartic corrections to the graviton scattering amplitude in the forward limit to be

$$
\begin{aligned}
M_4(s, t \to 0) = \frac{\kappa^4 s^4}{2} \Big[& (2c_6 + c_7)(\epsilon_{1\mu\nu}\epsilon_1{}^{\mu\nu}\epsilon_{2\rho\sigma}\epsilon_2{}^{\rho\sigma}) + (32c_1 + 4c_2 + 2c_6)(\epsilon_{1\mu\nu}\epsilon_2{}^{\mu\nu})^2 \\
& + (4c_2 + 16c_3 + 2c_6)(\epsilon_1{}^{\mu\nu}\epsilon_{2\nu\rho}\epsilon_1{}^{\rho\sigma}\epsilon_{2\sigma\mu}) + (4c_2 + 8c_4 + 2c_7)(\epsilon_1{}^{\mu\nu}\epsilon_{1\nu\rho}\epsilon_2{}^{\rho\sigma}\epsilon_{2\sigma\mu}) \Big].
\end{aligned}
\tag{19}
$$

Eq. (16) bounds f_4, corresponding to the coefficient of the s^4 contribution to the amplitude, to be positive. To determine the constraint on the coefficients of $\mathcal{L}_4$, we should marginalize over all possible values of the independent polarizations, ϵ_1 and ϵ_2.

To determine the full set of bounds, it will be convenient to map the question of positivity to a linear algebra problem. To do so, we work in the center-of-mass frame, where the polarization tensors, $\epsilon_{1\mu}{}^{\nu}$ and $\epsilon_{2\mu}{}^{\nu}$ are real, symmetric $(D-2)$-by-$(D-2)$ matrices satisfying the usual tracelessness and normalization conditions,

$$
\text{Tr}(\epsilon_1) = \text{Tr}(\epsilon_2) = 0 \qquad \text{and} \qquad \text{Tr}(\epsilon_1 \cdot \epsilon_1) = \text{Tr}(\epsilon_2 \cdot \epsilon_2) = 1.
\tag{20}
$$

Furthermore, we can define Hermitian matrices $H_+ = \{\epsilon_1, \epsilon_2\}/2$ and $H_- = i[\epsilon_1, \epsilon_2]/2$ encoding the polarization information, which enter the amplitude in terms of the invariants

$$
x = \text{Tr}(H_+)\text{Tr}(H_+), \qquad y = \text{Tr}(H_+ \cdot H_+), \qquad z = \text{Tr}(H_- \cdot H_-).
\tag{21}
$$

We can then express the analyticity bound as

$$
(2c_6 + c_7) + (32c_1 + 4c_2 + 2c_6)x + (8c_2 + 16c_3 + 8c_4 + 2c_6 + 2c_7)y + (-16c_3 + 8c_4 - 2c_6 + 2c_7)z > 0, \tag{22}
$$

for all (x, y, z) spanned by the graviton polarizations ϵ_1 and ϵ_2. What is the allowed space of (x, y, z)? An obvious set of necessary conditions are

$$
0 \le x, y, z \le 1 \qquad \text{and} \qquad y + z \le 1,
\tag{23}
$$

from familiar linear algebra inequalities. In general D, finding the space spanned by the allowed (x, y, z) is a highly nontrivial problem in matrix inequalities.

In the next subsections, we will study various physically well-motivated scenarios, including general theories in $D = 4$ and supersymmetric theories in arbitrary D.

3.1 Theories in $D = 4$

The number of linearly independent curvature invariants monotonically increases with the dimension of spacetime. In $D = 4$, there are only two independent quartic curvature invariants. Hence, $\mathcal{L}_4$ in Eq. (4) collapses to

$$\mathcal{L}_4 = c_1 \mathcal{O}_1 + \tilde{c}_1 \tilde{\mathcal{O}}_1, \tag{24}$$

where $\mathcal{O}_1$ is defined as in Eq. (4) but $\tilde{\mathcal{O}}_1$ is unique to $D = 4$,

$$\mathcal{O}_1 = R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma} R^{\alpha\beta\gamma\delta} R_{\alpha\beta\gamma\delta} \qquad \text{and} \qquad \tilde{\mathcal{O}}_1 = R^{\mu\nu\rho\sigma} \tilde{R}_{\mu\nu\rho\sigma} R^{\alpha\beta\gamma\delta} \tilde{R}_{\alpha\beta\gamma\delta}, \tag{25}$$

where $\tilde{R}_{\mu\nu\rho\sigma} = R_{\mu\nu}{}^{\alpha\beta} \epsilon_{\alpha\beta\rho\sigma}/2$ is the dual Riemann tensor. The operator $\tilde{\mathcal{O}}_1$ can be written as a linear combination of any two the operators in Eq. (4) modulo contributions proportional to R and $R_{\mu\nu}$, which can be eliminated by the equations of motion. For example,

$$\tilde{\mathcal{O}}_1 = 4\mathcal{O}_2 - 4\mathcal{O}_3 = -4\mathcal{O}_2 + 8\mathcal{O}_4 = \cdots, \tag{26}$$

corresponding to a choice of operator coefficients, $(c_1, 4\tilde{c}_1, -4\tilde{c}_1, 0, 0, 0, 0)$, $(c_1, -4\tilde{c}_1, 0, 8\tilde{c}_1, 0, 0, 0)$, etc. The ellipses in Eq. (26) denote equivalent representations in terms of other operators, which are not unique due to the linear dependence in $D = 4$ of all but two of the operators in Eq. (4).

In $D = 4$, the invariants (x, y, z) are constructed from real, symmetric, traceless 2-by-2 matrices, which we can parameterize by

$$\begin{aligned} \epsilon_1 &= \vec{\epsilon}_1 \cdot \vec{\sigma}/\sqrt{2} \\ \epsilon_2 &= \vec{\epsilon}_2 \cdot \vec{\sigma}/\sqrt{2}, \end{aligned} \tag{27}$$

where $\vec{\epsilon}_1$ and $\vec{\epsilon}_2$ are real unit polarization vectors and $\vec{\sigma}$ are the Pauli matrices. Since ϵ_1 and ϵ_2 are real and symmetric, they only have components in σ_1 and σ_3, since σ_2 is imaginary and anti-symmetric. From standard matrix identities, we see that $\{\epsilon_1, \epsilon_2\} = \vec{\epsilon}_1 \cdot \vec{\epsilon}_2$ and $[\epsilon_1, \epsilon_2] = i(\vec{\epsilon}_1 \times \vec{\epsilon}_2) \cdot \vec{\sigma}$. Defining θ to be the angle between $\vec{\epsilon}_1$ and $\vec{\epsilon}_2$, we obtain

$$(x, y, z) = \cos^2 \theta \ (1, \tfrac{1}{2}, 0) + \sin^2 \theta \ (0, 0, \tfrac{1}{2}), \tag{28}$$

which defines an interval whose endpoints are $(1, \tfrac{1}{2}, 0)$ and $(0, 0, \tfrac{1}{2})$. Inserting these (x, y, z) values, along with the coefficient choice given by Eqs. (24) and (26), the bound (22) takes the suggestive form

$$c_1 \cos^2 \theta + \tilde{c}_1 \sin^2 \theta > 0, \tag{29}$$

which obviously implies positivity of both coefficients separately,

$$c_1 > 0 \quad \text{and} \quad \tilde{c}_1 > 0, \tag{30}$$

which correspond to parallel or perpendicular polarization vectors, respectively. Our results exactly coincide with those derived from requiring subluminal graviton propagation [14].

9

3.2 Theories in $D = 5$

In $D = 5$, there are four linearly independent quartic curvature invariants. For the sake of generality we use the basis of Eq. (4) with the linear dependences among operators assumed. For this analysis, we ascertain the physically allowed region for the invariants (x, y, z), which in $D = 5$ are constructed from real, symmetric, traceless 3-by-3 matrices. This requirement constrains (x, y, z) to lie in the plane $1 + 2x - 6y - 2z = 0$. Specifically, (x, y, z) are restricted to a planar triangular region,

$$(x, y, z) = \sum_{i=1}^{3} \tau_i v_i, \tag{31}$$

defined by three vectors

$$v_1 = (0, 0, \tfrac{1}{2}), \qquad v_2 = (1, \tfrac{1}{2}, 0), \qquad \text{and} \qquad v_3 = (0, \tfrac{1}{6}, 0) \tag{32}$$

for the real parameters $\tau_1, \tau_2, \tau_3 \geq 0$ such that $\tau_1 + \tau_2 + \tau_3 = 1$. The vertices (32) of this triangle can be reached by choices of physical polarizations. In particular,

$$
v_1 : \quad \epsilon_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \epsilon_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},
$$

$$
v_2 : \quad \epsilon_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \epsilon_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{33}
$$

$$
v_3 : \quad \epsilon_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \epsilon_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.
$$

Plugging in Eqs. (31) and (32) back into Eq. (22), we obtain

$$
\begin{bmatrix} \tau_1(-8c_3 + 4c_4 + c_6 + 2c_7) \\ + \quad \tau_2(32c_1 + 8c_2 + 8c_3 + 4c_4 + 5c_6 + 2c_7) \\ + \quad \tfrac{1}{3}\tau_3(4c_2 + 8c_3 + 4c_4 + 7c_6 + 4c_7) \end{bmatrix} > 0, \tag{34}
$$

where we have repackaged the terms independent of (x, y, z) in Eq. (22) into the coefficients τ_1, τ_2, τ_3 by re-expressing 1 as $\tau_1 + \tau_2 + \tau_3$. Thus, the necessary and sufficient set of bounds on quartic curvature corrections in $D = 5$ are

$$-8c_3 + 4c_4 + c_6 + 2c_7 > 0$$

$$32c_1 + 8c_2 + 8c_3 + 4c_4 + 5c_6 + 2c_7 > 0 \tag{35}$$

$$4c_2 + 8c_3 + 4c_4 + 7c_6 + 4c_7 > 0,$$

coming from analyticity of the four-point graviton scattering amplitude.

10

3.3 Theories in $D \geq 6$

Consider finally the general case of $D \geq 6$. It is a non-trivial linear algebra problem to determine the parameter space of (x, y, z) corresponding to physical polarization configurations. Each physically allowed point (x, y, z) yields a positivity bound via Eq. (22). The set of all positive linear combinations of such bounds is given by plugging into Eq. (22) the set of all points in the convex hull S spanning physically allowed values of (x, y, z). Fully characterizing all such (x, y, z) is beyond the scope of the present work. However, we can derive a general collection of necessary conditions from a subset of extremal vertices on the boundary of S. The details of the calculation are given in App. A, but the vertices are

$$
\begin{aligned}
v_1 &= \left(0, 0, \tfrac{1}{2}\right) \\
v_2 &= \left(1, 1 - \tfrac{3}{D-2} + \tfrac{1}{D-3}, 0\right) \\
v_3 &= \left(0, \tfrac{D-4}{2(D-2)}, 0\right) \\
v_4 &= \left(1, \tfrac{1}{D-2}\left[1 + \tfrac{4(D \bmod 2)}{(D-1)(D-3)}\right], 0\right) \\
v_5 &= \left(0, 0, 0\right).
\end{aligned}
\tag{36}
$$

These vectors can be realized by physical polarization choices. The bounds associated with the (x, y, z) values in Eq. (36) are necessary for analyticity of four-point scattering amplitudes and moreover are a subset of the minimal basis of sufficient bounds. Numerical evaluation, via the explicit computation of x, y, and z for pseudorandom, traceless, unit-norm matrix pairs of various dimensions, shows that the convex hull defined by the vertices in Eq. (36) is in fact equal to the full hull S for even D but is slightly smaller than S in odd D. Note that the vectors (36) are a generalization of those we saw in earlier sections, so v_1, v_2, and v_3 coincide with the vectors from $D = 5$. Moreover, each corner corresponds to a certain extreme configuration of polarizations. For example, v_1 corresponds anticommuting polarizations as in Eq. (32), while v_2, v_3, v_4, and v_5 correspond to commuting polarizations. For the latter, the polarizations are mutually diagonalizable and can without loss of generality be represented as traceless diagonal matrices. See App. A for details.

Plugging the vectors in Eq. (36) into the bound in Eq. (4), we obtain the positivity bounds

$$
-8c_3 + 4c_4 + c_6 + 2c_7 > 0
$$
$$
2\left(1 - \tfrac{3}{D-2} + \tfrac{1}{D-3}\right)(4c_2 + 8c_3 + 4c_4 + c_6 + c_7) + 32c_1 + 4c_2 + 4c_6 + c_7 > 0
$$
$$
\left(\tfrac{D-4}{D-2}\right)(4c_2 + 8c_3 + 4c_4 + c_6 + c_7) + 2c_6 + c_7 > 0 \tag{37}
$$
$$
\left(\tfrac{2}{D-2}\right)\left[1 + \tfrac{4(D \bmod 2)}{(D-1)(D-3)}\right](4c_2 + 8c_3 + 4c_4 + c_6 + c_7) + 32c_1 + 4c_2 + 4c_6 + c_7 > 0
$$
$$
2c_6 + c_7 > 0,
$$

which are a stringent set of requirements on quartic curvature corrections to general relativity in $D \geq 6$, necessary to guarantee analyticity of scattering amplitudes.

3.4 Supersymmetric Theories

We now consider supersymmetric quartic curvature corrections. Conveniently, Ref. [17] derived a basis for independent off-shell supersymmetric quartic curvature invariants,

$$\mathcal{L}_4 = A\mathcal{O}_A + B\mathcal{O}_B + C\mathcal{O}_C, \tag{38}$$

where $\mathcal{O}_A$, $\mathcal{O}_B$, and $\mathcal{O}_C$ are proportional to more familiar looking forms denoted in the literature [17, 18] by $t_8 t_8 R^4$, $t_8(R^2)^2$, and $\epsilon_{10}\epsilon_{10}R^4$, respectively. In terms of the basis defined in Eq. (4), these supersymmetric operators are

$$\mathcal{O}_A = \mathcal{O}_1 - 16\mathcal{O}_2 + 2\mathcal{O}_3 - 32\mathcal{O}_5 + 16\mathcal{O}_6 + 32\mathcal{O}_7$$

$$\mathcal{O}_B = -\mathcal{O}_1 + 8\mathcal{O}_2 - 2\mathcal{O}_3 + 4\mathcal{O}_4 \tag{39}$$

$$\mathcal{O}_C = \mathcal{O}_1 - 16\mathcal{O}_2 + 2\mathcal{O}_3 + 16\mathcal{O}_4 - 32\mathcal{O}_5 + 16\mathcal{O}_6 - 32\mathcal{O}_7,$$

corresponding to the following choice of operator coefficients:

$$c_1 = A - B + C \qquad c_2 = -16A + 8B - 16C \qquad c_3 = 2A - 2B + 2C$$

$$c_4 = 4B + 16C \qquad c_5 = -32A - 32C \qquad c_6 = 16A + 16C \tag{40}$$

$$c_7 = 32A - 32C.$$

Plugging this choice into Eq. (22), we obtain

$$A + B(y + z) > 0. \tag{41}$$

Note that C drops out of the calculation completely, since at quartic order in graviton perturbations it is a total derivative in all dimensions [13]. Recall from App. A that while simple matrix identities imply that $y + z \leq 1$, no point in the hull S actually saturates this bound. For example, in $D = 4$, Eq. (28) implies that $y + z = \frac{1}{2}$, so $A + \frac{1}{2}B > 0$. In $D = 5$, Eq. (32) implies that $\frac{1}{6} \leq y + z \leq \frac{1}{2}$, so $A + \frac{1}{6}B > 0$ and $A + \frac{1}{2}B > 0$. Finally, in $D \geq 6$, inputting the vertices in Eq. (36) into Eq. (41) yields the complete set of positivity bounds for quartic curvature operators in supergravity theories. In summary, we find

$$A + \tfrac{1}{6}B > 0 \qquad (D = 5)$$

$$A > 0 \qquad (D \geq 6) \tag{42}$$

$$A + \left(1 - \tfrac{3}{D-2} + \tfrac{1}{D-3}\right)B > 0, \qquad (\text{any } D)$$

noting that in $D = 4$ and $D = 5$, the final bound reduces to $A + \frac{1}{2}B > 0$.

3.5 String Theories

As a consistency check, we now apply our bounds to string theory, which is arguably the leading candidate for the ultraviolet completion of gravity. Conveniently, quartic curvature corrections have been been dutifully computed at tree-level in the existing literature for the bosonic [19, 20], type II [21–23], and heterotic string [22, 23]. The type I string is dual to the heterotic string and has the same low-energy effective action [18, 24], so we need not consider it as a separate case. The resulting effective theory is described by

$$\mathcal{L}_4 = A\mathcal{O}_A + B\mathcal{O}_B + C\mathcal{O}_C + \Delta\mathcal{O}_\Delta, \tag{43}$$

where $\mathcal{O}_A$, $\mathcal{O}_B$, and $\mathcal{O}_C$ are the supersymmetric operators from the previous section and $\mathcal{O}_\Delta$ is a non-supersymmetric operator defined as

$$\mathcal{O}_\Delta = -\mathcal{O}_1 + 10\mathcal{O}_2 + \mathcal{O}_4. \tag{44}$$

In various string theories, the operator coefficients are

	A	B	C	Δ
bosonic	$\zeta(3)$	0	$-\zeta(3)$	16
type II	$\zeta(3)$	0	$-\zeta(3)$	0
heterotic	$\zeta(3)$	1	$-\zeta(3)$	0

$$\tag{45}$$

where each entry is normalized by a factor of $\alpha'^3/1024\kappa^2$.

As expected, since the type II and heterotic string theories are supersymmetric, their coefficients in Eq. (45) satisfy the bound for supersymmetric theories in Eq. (42). Since the bosonic string is non-supersymmetric, the bound is more complicated. In particular, plugging the corresponding operator coefficients into Eq. (22), we obtain

$$\zeta(3) + 2(x + 11y + z) > 0, \tag{46}$$

which is indeed positive, as $x, y, z \geq 0$. Thus, we have verified that quartic curvature corrections in bosonic, type II, and heterotic string theory are consistent with unitarity and analyticity.

4 Bounds on Quadratic Curvature Corrections

Next, we consider analyticity constraints on $\mathcal{L}_2$, which characterizes quadratic curvature corrections in the graviton effective theory. As shown in Ref. [9], the Gauss-Bonnet term

$$\mathcal{L}_2 = \lambda(R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2) \tag{47}$$

13

does not introduce ghost modes in any dimension D, so in this basis the graviton propagator is unmodified. To avoid ghost pathologies, we only consider curvature invariants of this form. For $D = 4$, the Gauss-Bonnet term is furthermore a total derivative and thus does not affect local dynamics. As recently shown [25], however, the Gauss-Bonnet term is critical for computing and interpreting the leading ultraviolet divergences of pure gravity.

Expanding to leading order in the Gauss-Bonnet coefficient λ, we compute the quadratic curvature correction to the graviton scattering amplitude in the forward limit,

$$M_2(s, t \to 0) = 4\lambda\kappa^4 s^2 \Big[\epsilon_1^{\mu\nu} \epsilon_{3\mu\nu} \epsilon_2^{\rho\sigma} \epsilon_{4\rho\sigma} + \epsilon_1^{\mu\nu} \epsilon_{3\nu\rho} \epsilon_2^{\rho\sigma} \epsilon_{4\sigma\mu} + \epsilon_1^{\mu\nu} \epsilon_{3\nu\rho} \epsilon_4^{\rho\sigma} \epsilon_{2\sigma\mu}$$
$$+ \frac{2}{t} \Big(k_2^\mu k_4^\nu \epsilon_{2\nu}{}^\rho \epsilon_{4\rho\mu} \epsilon_{1\alpha\beta} \epsilon_3^{\alpha\beta} + k_1^\mu k_3^\nu \epsilon_{1\nu}{}^\rho \epsilon_{3\rho\mu} \epsilon_{2\alpha\beta} \epsilon_4^{\alpha\beta} \Big) \Big], \tag{48}$$

where we have expanded formally in t-dependence arising from propagator denominators, but we have yet to evaluate the numerators in the forward limit.

The first line of Eq. (48) is manifestly regular in the forward limit $t = 0$, so for these terms we can simply set $\epsilon_3 = \epsilon_1$ and $\epsilon_4 = \epsilon_2$. On the other hand, the second line of Eq. (48) is naively singular since $1/t$ diverges as $t \to 0$. However, this singularity is canceled by the numerator factor, which vanishes in the forward limit as $\epsilon_3 \to \epsilon_1$ and $\epsilon_4 \to \epsilon_2$. It will be convenient to rewrite this expression in terms of the momentum transfer,

$$q = k_1 + k_3 = -(k_2 + k_4), \tag{49}$$

where $t = -q^2$. For real kinematics, q is spacelike and vanishes in the forward limit. Note that $q^\mu q^\nu / q^2$ is simply a projection operator in the direction of the spacelike exchanged momentum.

We note that k_3 is simply a real spatial rotation of $-k_1$, and likewise for k_4 and k_2. By symmetry, this then implies that $\epsilon_1^{\mu\nu} k_{3\nu} = \epsilon_3^{\mu\nu} k_{1\mu} = \epsilon_1^{\mu\nu} q_\mu$ and $\epsilon_2^{\mu\nu} k_{4\nu} = \epsilon_4^{\mu\nu} k_{2\mu} = -\epsilon_2^{\mu\nu} q_\mu$ at leading order in q. Rewriting Eq. (48) in terms of q, we then have

$$M_2(s, t \to 0) = 4\lambda\kappa^4 s^2 \Big[\epsilon_{1\mu\nu} \epsilon_1^{\mu\nu} \epsilon_{2\rho\sigma} \epsilon_2^{\rho\sigma} + 2\epsilon_1^{\mu\nu} \epsilon_{1\nu\rho} \epsilon_2^{\rho\sigma} \epsilon_{2\sigma\mu}$$
$$- \frac{2q^\mu q^\nu}{q^2} \Big(\epsilon_{2\mu}{}^\rho \epsilon_{2\rho\nu} \epsilon_{1\alpha\beta} \epsilon_1^{\alpha\beta} + \epsilon_{1\mu}{}^\rho \epsilon_{1\rho\nu} \epsilon_{2\alpha\beta} \epsilon_2^{\alpha\beta} \Big) \Big], \tag{50}$$

which is regular because the projection operator $q^\mu q^\nu / q^2$ is finite in the forward limit. To obtain a bound on λ, we consider all possible choices for the external momenta and polarizations and impose positivity bounds on the forward amplitude in Eq. (50).

As expected, quadratic curvature corrections to graviton scattering scale as $M_2 \sim \lambda\kappa^4 s^2$, so to extract an analyticity bound we should apply Eq. (16) for a second-order residue, corresponding

14

to $n = 2$. Unfortunately, this choice also extracts the t-channel singular contribution from leading-order graviton exchange, $M_1 \sim -\kappa^2 s^2/t$. In the forward limit, this contribution is formally infinite. Of course, in any physical experiment there is an infrared scale μ that regulates these contributions from long distance physics. This would arise, *e.g.*, from a finite detector resolution or beam width [26]. As is common practice for infrared divergences in scattering amplitudes, we introduce a mass regulator, sending $t \to t - \mu^2$ in the denominator. This approach was also used in Ref. [1] to make sense of a theory of interacting massless scalars with trilinear couplings. Note that as in gauge theory, the mass μ^2 is a formal regulator that leaves the number degrees of freedom untouched—so the vDVZ discontinuity [27, 28], which arises for a physical graviton mass included via a Fierz-Pauli Lagrangian term, does not apply here.

While μ^2 tames the formal infrared divergences, for $\lambda \lesssim 1$ the forward amplitude will be dominated by finite but large contributions from Einstein-Hilbert interactions because $|M_1| \gg |M_2|$ in this regime.[5] However, by explicit calculation, we can see from Eq. (18) that $M_1 \sim +\kappa^2 s^2/\mu^2$, which is positive. So while positivity is satisfied, we learn nothing beyond what is already borne out from scattering via the leading Einstein-Hilbert term.

To place a bound on the coefficient λ, we must then restrict to a parameter regime where $|M_1| \lesssim |M_2|$, so the contributions from graviton exchange are subdominant to those from the Gauss-Bonnet term. This implies that $1/\mu^2 \lesssim |\lambda\kappa^2|$. Together with the requirement that $|s| \gg \mu^2$, necessary to treat μ as a regulator, this forces us to consider the regime

$$\sqrt{|s|} \gg \mu \gtrsim \Lambda, \tag{51}$$

where $\Lambda \sim |\lambda\kappa^2|^{-1/2}$ is the scale of the would-be natural cutoff associated with the derivative expansion. We assume throughout that $\Lambda \ll \kappa^{\frac{2}{2-D}}$ so that it is below the Planck scale in D dimensions.

Of course, Eq. (51) points to a naively pathological region of the effective field theory, given the reasonable expectation of new degrees of freedom of mass m where $m \sim \Lambda$. Moreover, Eq. (51) indicates that the infrared regulator μ must be larger than some other energy scale Λ. Nevertheless, one can *a priori* envision an ultraviolet completion in which $m \gg \Lambda$, so new degrees of freedom enter at a parametrically higher scale. In that case, Λ is not the scale of any physical states in the theory and is merely the combination of parameters that appears in the higher-dimension operator. Indeed, at the level of the scattering amplitude, there are no discontinuities that appear around Λ to signal new degrees of freedom.

---

[5]Note that taking t strictly to zero is not required to derive a positivity bound [29] and positivity holds for any non-negative t below μ^2 [30]. However, we will not need this more general result for our purposes.

Thus, μ remains smaller than any physical mass scale in the theory and indeed can be treated consistently as an infrared regulator. In the absence of new states at Λ, the Gauss-Bonnet term acts effectively as a primordial contact operator over a wide range of scales. Precisely such a scenario was considered in Ref. [15], where it was found that such a low-energy effective theory is acausal without new states at or below $m \lesssim \Lambda$. Other authors [31] have likewise argued that a pure Gauss-Bonnet theory is inconsistent with black hole thermodynamics. We will likewise find a pathology in this theory coming from unitarity and analyticity.

To apply constraints from unitarity and analyticity, we must first ensure that the low-energy theory is sensible enough that we can even speak of a long-distance scattering amplitude. Indeed, Eq. (51) is plainly strange since $|s| \gg \Lambda^2$ violates the derivative expansion. This was required in order for the Gauss-Bonnet interactions to dominate over the Einstein-Hilbert action, as was also assumed in Ref. [15]. Naively, one would expect a gross departure from perturbative unitarity, *e.g.*, probability amplitudes greater than one as well as a breakdown of the loop expansion. Nevertheless, there is a wide range of scales where neither sickness actually arises. This hinges on the fact that the theory depends on Λ as well as κ, the gravitational coupling constant.

In particular, note that amplitudes can still be perturbatively small in the regime specified by Eq. (51). For example, $M_2 \sim \kappa^2 s^2 / \Lambda^2$ is still sensible provided κ is sufficiently small, corresponding to the weak gravity limit. We can make this more precise by considering the leading effect of the Gauss-Bonnet term, which is a cubic vertex of the schematic form $\lambda \kappa^3 \partial^4 h^3$. Inserting this vertex into low-energy amplitudes, we find that the theory remains under perturbative control provided $\lambda \kappa^3$ ($\sim \kappa/\Lambda^2$) times the appropriate powers of energy is sufficiently small. In D dimensions this implies that

$$|s| \ll \left(\frac{\Lambda^2}{\kappa} \right)^{\frac{4}{2+D}} \tag{52}$$

to safely reside within the regime of perturbativity.[6] Moreover, Eq. (52) also ensures a perturbative loop expansion, since radiative corrections always introduce additional insertions of the Gauss-Bonnet interactions.

For our purposes, we assume a weak gravity limit defined by Eq. (52), so the low-energy theory is perturbatively unitary. When then apply the method of Sec. 2, where the contour around the origin in the complex s plane is widened so as to satisfy Eq. (51), ensuring that s is large compared to the scale of the infrared cutoff and that the Gauss-Bonnet term dominates the amplitude. Note also that the initial contour encircles a region below the heavy particle

---

[6]A similar statement applies to pions, which have quartic vertices of the form $\partial^4 \pi^4 / \Lambda^2 v^2$ where v is the breaking scale and Λ controls the derivative expansion. The theory maintains perturbative control provided $s \ll \Lambda v$.

threshold, $m \gg \Lambda$.

To see how a pathology arises, it will be convenient to define coordinates transverse to the incoming momenta, $(x_1, x_2, \ldots x_{D-2})$. Without loss of generality, we take the forward limit such that the infinitesimal momentum transfer lies in the x_1 direction, which we henceforth refer to as the "direction of approach." In turn, $q^\mu q^\nu / q^2$ is a projection operator onto this direction. Next, we define a particular subset of polarizations in the transverse plane, defined by rank-two diagonal matrices of the form

$$d^{(i,j)} = \frac{1}{\sqrt{2}} \operatorname{diag}(0, \ldots, 0, \overbrace{1}^{x_i}, 0, \ldots, 0, \overbrace{-1}^{x_j}, 0, \ldots), \tag{53}$$

with zero entries except in the x_i and x_j directions. As the only preferred direction is x_1, labeling the direction from which we approach the forward limit, the relevant physical polarizations are $d^{(1,2)}$, $d^{(2,3)}$, and $d^{(3,1)}$. The forward limit of the quadratic correction to the graviton scattering amplitude in Eq. (50) for various polarization combinations is

$$M(s, t \to 0) = 2\lambda \kappa^4 s^2 \times \begin{cases} 0, & \epsilon_1 = \epsilon_2 = d^{(1,2)} \\ 4, & \epsilon_1 = \epsilon_2 = d^{(2,3)} \\ -1, & \epsilon_1 = d^{(1,2)} \text{ and } \epsilon_2 = d^{(1,3)}. \end{cases} \tag{54}$$

In the first case, $\epsilon_1 = \epsilon_2 = d^{(1,2)}$, corresponding to polarizations that have support in the direction of approach. In the case of $D = 4$, this is required because the transverse space only has two dimensions. As expected, the amplitude vanishes in this regime because the Gauss-Bonnet term is a total derivative in $D = 4$. Meanwhile, the second case, $\epsilon_1 = \epsilon_2 = d^{(2,3)}$ occurs when both polarizations are orthogonal to the direction of approach. Of course, this requires dimensions $D \geq 5$. Finally, in the last case, $\epsilon_1 = d^{(1,2)}$ and $\epsilon_2 = d^{(1,3)}$, the polarizations occupy different planes but share support in the direction of approach, which is only possible in $D \geq 5$.

The upshot of Eq. (54) is that in $D \geq 5$, different polarization configurations can yield opposite signs for the corrections to the forward scattering amplitude. As a result, this excludes both signs of λ and thus forbids it entirely. Of course, we made the assumption of Ref. [15] that the Gauss-Bonnet term is an effectively primordial contact operator insofar as new degrees of freedom enter only at a scale far above the naive cutoff. Hence, the positivity violation in Eq. (54) simply implies that this assumption is false. We conclude that a primordial Gauss-Bonnet term is forbidden and new degrees of freedom are required at or below the cutoff Λ.

5 Conclusions

In this paper, we have derived rigorous bounds on the coefficients of quartic and quadratic curvature corrections in the low-energy effective theory of gravitons. Our results hinge on very general principles: quantum mechanical unitarity and analyticity of scattering amplitudes. Consequently, these constraints apply to any consistent perturbative ultraviolet completion of gravity. For the quartic curvature operators defined in Eqs. (3), (4), (24), and (25), we derived the positivity bounds in Eq. (30) in $D = 4$, Eq. (35) in $D = 5$, and Eq. (37) in arbitrary $D \geq 6$. We also presented constraints on supergravity theories and checked that all of our results are consistent with known calculations in weakly-coupled string theories. For the quadratic curvature correction in Eq. (47), we showed that both signs of its coefficient λ are inconsistent unless new degrees of freedom enter at the natural cutoff $\Lambda \sim |\lambda\kappa^2|^{-1/2}$ specified by the effective theory. In short, a primordial Gauss-Bonnet term is forbidden.

Many possibilities remain for future work. While four-point graviton scattering cannot probe curvature operators beyond quartic order, little is known of higher-point amplitudes. Such amplitudes are functions of many more kinematic invariants and should thus enforce commensurately more positivity constraints. Another issue meriting further study is that of cubic curvature operators. Here, positivity bounds encounter technical challenges due to the vanishing of the associated amplitude in the forward limit [30, 32]. As noted in Ref. [32], this problem is closely related to the a-theorem in $D = 6$.

Distinguishing low-energy effective theories that are consistent with ultraviolet completion from those that are not presents a significant challenge. Systematizing this procedure is important for delineating the space of possible physical laws, but has also become important for model-building in more phenomenological contexts [33] and in inflation [34–38]. In this paper, the low-energy tools of analyticity and unitarity enabled us to find solutions to this problem in gravitational theories, allowing us to constrain higher-curvature corrections to gravity in our own universe—applying our quartic curvature results to $D = 4$—and further discover bounds applicable in any consistent theory.

A   Bounding Invariants in General Dimension

We have shown that the graviton scattering amplitude can be expressed in terms of invariant products of graviton polarizations ϵ_1 and ϵ_2, which are real, symmetric, traceless $(D-2)$-by-$(D-2)$ matrices with unit normalization. To recapitulate from Sec. 3, given the Hermitian matrices

$$H_+ = \{\epsilon_1, \epsilon_2\}/2 \qquad \text{and} \qquad H_- = i[\epsilon_1, \epsilon_2]/2, \tag{A.1}$$

we can define the invariants

$$x = \text{Tr}(H_+)\text{Tr}(H_+), \qquad y = \text{Tr}(H_+ \cdot H_+), \qquad z = \text{Tr}(H_- \cdot H_-). \tag{A.2}$$

The space of physical polarizations ϵ_1 and ϵ_2 then maps onto a physical region in (x, y, z), which through Eq. (22) implies positivity constraints on operator coefficients in the effective theory.

What are the bounds on (x, y, z)? We first note that since H_+ and H_- are Hermitian, their squares are positive semidefinite, so $x, y, z \geq 0$. Moreover, a straightforward application of the Cauchy-Schwarz inequality implies $x = \text{Tr}(\epsilon_1 \cdot \epsilon_2)\text{Tr}(\epsilon_1 \cdot \epsilon_2) \leq \text{Tr}(\epsilon_1 \cdot \epsilon_1)\text{Tr}(\epsilon_2 \cdot \epsilon_2) = 1$, with equality if and only if $\epsilon_1 = \pm\epsilon_2$, and similarly $y + z = \text{Tr}(\epsilon_1 \cdot \epsilon_1 \cdot \epsilon_2 \cdot \epsilon_2) \leq \text{Tr}(\epsilon_1 \cdot \epsilon_1)\text{Tr}(\epsilon_2 \cdot \epsilon_2) = 1$. There are, however, many additional constraints on (x, y, z), which we now discuss.

Crucially, a weighted average of any number of positivity bounds yields another valid positivity bound. This implies that a space of necessary conditions can be derived by constructing a convex hull S in (x, y, z) that contains the physically allowed region. Without loss of generality, S is

$$S = (x, y, z) = \left\{ \sum_{i=1}^{n} \tau_i v_i \;\middle|\; \tau_i \geq 0 \text{ and } \sum_{i=1}^{n} \tau_i = 1 \right\}, \tag{A.3}$$

where v_i denote extremal points. In this Appendix, we will construct the subset of the v_i that are on the edges of the unit cube in (x, y, z); let the convex hull described by these vertices be $\tilde{S}$. In even dimension, numerical evaluation suggests that $S = \tilde{S}$, while in odd $D > 6$, it is possible for points to lie slightly outside $\tilde{S}$.

Let us first consider the case where ϵ_1 and ϵ_2 are anticommuting, so $x = y = 0$ and we wish to maximize z. Going to a basis in which ϵ_1 is diagonal, we find that antisymmetry of $\epsilon_1 \cdot \epsilon_2$ implies that for each i, j,

$$(\epsilon_{1ii} + \epsilon_{1jj})\epsilon_{2ij} = 0 \tag{A.4}$$

and

$$z = -\text{Tr}(\epsilon_1 \cdot \epsilon_2 \cdot \epsilon_1 \cdot \epsilon_2) = \sum_{i,j} \epsilon_{1ii}^2 \epsilon_{2ij}^2. \tag{A.5}$$

Since Eq. (A.4) implies $\epsilon_{1ii}\epsilon_{2ii} = 0$ for each i, the normalization condition $\sum_{i,j} \epsilon_{2ij}^2 = 1$ implies by Eq. (A.5) that nonzero diagonal terms in ϵ_2 can only decrease z. We therefore take ϵ_2 to have vanishing diagonal. Similarly, since $\sum_i \epsilon_{1ii}^2$ is fixed to unity, we should require that, for each i for which $\epsilon_{1ii} \neq 0$, there exists j such that $\epsilon_{2ij} \neq 0$; letting $\epsilon_{1i_0i_0}$ be nonvanishing for some i_0 even if $\epsilon_{2i_0j} = 0$ for all j would decrease z by Eq. (A.5). Writing $\sum_j \epsilon_{2ij}^2 = \rho_i$, where $\sum_i \rho_i = 1$ by the normalization constraint, we can then consider z to be a weighted average over the ϵ_{1ii}^2. Thus, z is maximized when we weight the average most in favor of the i for which ϵ_{1ii}^2 is maximal. Suppose there are N such i, which we can without loss of generality take to be 1 through N, for which ϵ_{1ii}^2 takes its maximal value, i.e., $\epsilon_{1i^*i^*}^2 = \max_i \epsilon_{1ii}^2 \equiv \varepsilon^2$ for all $i^* \in \{1, ..., N\}$. Then z is maximized when we have $\rho_i = 1/N$ for $i \in \{1, ..., N\}$ and $\rho_i = 0$ otherwise, for which we obtain $z = \varepsilon^2$. Finally, it remains to determine the maximal possible value of ε^2. Since ϵ_1 is of unit norm, its maximal value is attained when we load as much of the normalization into as few of the ϵ_{1ii}^2 as possible. By tracelessness of ϵ_1, at least two of the ϵ_{1ii} must be nonzero. Thus, ε^2 takes its maximum value of $1/2$ when $\epsilon_1 \propto \sigma_3$ in some 2-by-2 block, up to permutation of coordinate labels. That is, a choice of polarizations that maximizes z for $x = y = 0$ is

$$\epsilon_1 = \tfrac{1}{\sqrt{2}}\sigma_3 \oplus 0_{D-4} \qquad \text{and} \qquad \epsilon_2 = \tfrac{1}{\sqrt{2}}\sigma_1 \oplus 0_{D-4}, \tag{A.6}$$

which yields the point

$$v_1 = \left(0, 0, \tfrac{1}{2}\right). \tag{A.7}$$

Let us henceforth consider the case where $z = 0$ and explore in x, y. This means that the (real, symmetric) matrices $\epsilon_{1,2}$ commute and so are simultaneously diagonalizable. Taking $x = 1$, we can ask how large y can be, which will give a vertex of S. Since $\epsilon_1 = \pm\epsilon_2$ for $x = 1$, we have $y = \text{Tr}(\epsilon_1 \cdot \epsilon_1 \cdot \epsilon_1 \cdot \epsilon_1)$. That is, y has positive first and second derivatives in each of the $|\epsilon_{1ii}|$ values; y is therefore maximized when one of the $|\epsilon_{1ii}|$ is as large as possible and the others are equal and small. (If the smaller numbers in the list were unequal, we could always make y larger by shifting some weight back to the element in the list with the largest absolute value.)

20

That is, a choice of polarizations maximizing y for $x = 1$ and $z = 0$ is

$$\epsilon_1 = \epsilon_2 = \frac{1}{\sqrt{(D-2)(D-3)}}\mathrm{diag}(1, 1, \ldots, -(D-3)), \tag{A.8}$$

which corresponds to the vertex

$$v_2 = \left(1, 1 - \frac{3}{D-2} + \frac{1}{D-3}, 0\right). \tag{A.9}$$

Next, still taking $z = 0$, we consider a different extreme, setting $x = 0$ and maximizing y. Again simultaneously diagonalizing ϵ_1 and ϵ_2, we have $y = \sum_i \epsilon_{1ii}^2 \epsilon_{2ii}^2$. Analogously with the case of v_1, we can write $\rho_i = \epsilon_{1ii}^2$ and consider y to be a weighted average over the ϵ_{2ii}^2. Let $\max_i \epsilon_{2ii}^2 \equiv \varepsilon^2$ and, without loss of generality, suppose that $\epsilon_{2ii}^2 = \varepsilon^2$ for $i \in \{1, ..., N\}$ for some N. Then y is maximized if we take $\rho_i = 1/N$ for $i \in \{1, ..., N\}$ and $\rho_i = 0$ otherwise, in which case $y = \varepsilon^2$. Now, by the unit normalization of ϵ_2, ε^2 is maximized when as much of the normalization as possible is loaded into as few terms as possible, *i.e.*, N is minimized. Since ϵ_1 is traceless, at least two of the ρ_i are nonzero, so $N \geq 2$. The maximum value of ε^2 thus occurs when $N = 2$, which fixes $\epsilon_1 \propto \sigma_3 \oplus 0_{D-4}$. We now must maximize the common absolute value of the first two entries in ϵ_{2ii}, subject to the constraints that $\sum_i \epsilon_{2ii} = 0$, $\sum_i \epsilon_{2ii}^2 = 1$, and, since $x = 0$, $\sum_i \epsilon_{1ii}\epsilon_{2ii} = 0$. This last constraint implies that the first two entries in ϵ_{2ii} have the same sign. Thus, y is maximized for $x = z = 0$ for the choice of polarizations

$$\epsilon_1 = \frac{1}{\sqrt{2}}\sigma_3 \oplus 0_{D-4} \qquad \text{and} \qquad \epsilon_2 = \sqrt{\frac{2}{(D-2)(D-4)}}\mathrm{diag}\left(\frac{D-4}{2}, \frac{D-4}{2}, -1, \ldots, -1\right), \tag{A.10}$$

for which we find the vertex

$$v_3 = \left(0, \frac{D-4}{2(D-2)}, 0\right). \tag{A.11}$$

Note that the polarization configuration in Eq. (A.10), and hence the vertex in Eq. (A.11), requires $D \geq 5$.

On the other hand, we can minimize y/x for $z = 0$. Using symmetry and reality to diagonalize $H_+ = \mathrm{diag}\,\vec{h}_+$, we have $x = |\vec{h}_+ \cdot \vec{n}|^2$, where $\vec{n} = (1, 1, \ldots, 1)$, so the Cauchy-Schwarz inequality implies

$$x \leq |\vec{h}_+|^2 |\vec{n}|^2 = (D-2)y. \tag{A.12}$$

Eq. (A.12) is saturated when $H_+ \propto 1_{D-2}$. If D is even, this choice is possible with

$$\epsilon_1 = \epsilon_2 = \frac{1}{\sqrt{D-2}}\mathrm{diag}\left(1, -1, 1, -1, \ldots\right), \tag{A.13}$$

which yields

$$(x, y, z) = \left(1, \frac{1}{D-2}, 0\right). \tag{A.14}$$

21

Let us now consider the odd-dimensional case where $z = 0$ and $x = 1$. Simultaneously diagonalizing ϵ_1 and ϵ_2, we have $y = \sum_i \epsilon_{1ii}^4$. Again, y has positive first and second derivatives in $|\epsilon_{1ii}|$, so it is minimized when the $|\epsilon_{1ii}|$ are all equal. In odd dimension, this is not possible while retaining tracelessness, so the best one can do, making the $|\epsilon_{1ii}|$ as equal as possible, is the choice

$$\epsilon_1 = \epsilon_2 = \sqrt{\tfrac{D-3}{(D-1)(D-2)}}\,\mathrm{diag}\Big(\underbrace{1, \ldots, 1}_{\tfrac{D-1}{2}}, -\tfrac{D-1}{D-3}, ..., -\tfrac{D-1}{D-3}\Big), \tag{A.15}$$

which results in the vertex

$$(x, y, z) = \Big(1, \tfrac{2}{D-1} + \tfrac{2}{D-3} - \tfrac{3}{D-2}, 0\Big). \tag{A.16}$$

We can combine Eqs. (A.14) and (A.16) to write the vertex of S as

$$v_4 = \Big(1, \tfrac{1}{D-2}\Big[1 + \tfrac{4(D \bmod 2)}{(D-1)(D-3)}\Big], 0\Big). \tag{A.17}$$

We note that for both $D = 4$ and $D = 5$, v_2 and v_4 are the same point. Moreover, v_3 generalizes the third vertex applicable in $D = 5$, while the polarization choice for v_3 does not apply in $D = 4$. In $D \geq 6$, there is one remaining linearly independent vertex, which can be obtained by choosing $\epsilon_1 \cdot \epsilon_2 = 0_{D-2}$, e.g.,

$$\epsilon_1 = \tfrac{1}{\sqrt{2}}\sigma_3 \oplus 0_{D-4} \qquad \text{and} \qquad \epsilon_2 = \tfrac{1}{\sqrt{2}}0_{D-4} \oplus \sigma_3, \tag{A.18}$$

which results in the point

$$v_5 = (0, 0, 0). \tag{A.19}$$

Together, Eqs. (A.7), (A.9), (A.11), (A.17), and (A.19) are the vertices of $\tilde{S}$ given in Eq. (36). They correspond via Eq. (22) to a set of linearly independent bounds (37) that must be satisfied in any gravity theory in $D \geq 6$.

References

[1] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, "Causality, analyticity and an IR obstruction to UV completion," *JHEP* **0610** (2006) 014, arXiv:hep-th/0602178 [hep-th].

[2] B. Bellazzini, L. Martucci, and R. Torre, "Symmetries, Sum Rules and Constraints on Effective Field Theories," *JHEP* **1409** (2014) 100, arXiv:1405.2960 [hep-th].

[3] J. Distler, B. Grinstein, R. A. Porto, and I. Z. Rothstein, "Falsifying Models of New Physics via WW Scattering," *Phys.Rev.Lett.* **98** (2007) 041601, `arXiv:hep-ph/0604255 [hep-ph]`.

[4] A. V. Manohar and V. Mateu, "Dispersion Relation Bounds for $\pi\pi$ Scattering," *Phys. Rev.* **D77** (2008) 094019, `arXiv:0801.3222 [hep-ph]`.

[5] T. N. Pham and T. N. Truong, "Evaluation of the Derivative Quartic Terms of the Meson Chiral Lagrangian From Forward Dispersion Relation," *Phys. Rev.* **D31** (1985) 3027.

[6] C. Cheung and G. N. Remmen, "Infrared Consistency and the Weak Gravity Conjecture," *JHEP* **1412** (2014) 087, `arXiv:1407.7865 [hep-th]`.

[7] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, "The string landscape, black holes and gravity as the weakest force," *JHEP* **06** (2007) 060, `arXiv:hep-th/0601001 [hep-th]`.

[8] I. T. Drummond and S. J. Hathrell, "QED Vacuum Polarization in a Background Gravitational Field and Its Effect on the Velocity of Photons," *Phys. Rev.* **D22** (1980) 343.

[9] B. Zwiebach, "Curvature squared terms and string theories," *Phys. Lett.* **B156** (1985) 315.

[10] B. Zumino, "Gravity theories in more than four dimensions," *Phys. Rept.* **137** (1986) 109.

[11] H. Georgi, "On-shell effective field theory," *Nucl. Phys.* **B361** (1991) 339–350.

[12] S. Fulling, R. C. King, B. Wybourne, and C. Cummins, "Normal forms for tensor polynomials. 1: The Riemann tensor," *Class.Quant.Grav.* **9** (1992) 1151–1197.

[13] S. Deser and D. Seminara, "Tree amplitudes and two-loop counterterms in $D = 11$ supergravity," *Phys. Rev.* **D62** (2000) 084010, `arXiv:hep-th/0002241 [hep-th]`.

[14] A. Gruzinov and M. Kleban, "Causality Constrains Higher Curvature Corrections to Gravity," *Class. Quant. Grav.* **24** (2007) 3521–3524, `arXiv:hep-th/0612015 [hep-th]`.

[15] X. O. Camanho, J. D. Edelstein, J. Maldacena, and A. Zhiboedov, "Causality Constraints on Corrections to the Graviton Three-Point Coupling," `arXiv:1407.5597 [hep-th]`.

[16] I. Low, R. Rattazzi, and A. Vichi, "Theoretical Constraints on the Higgs Effective Couplings," *JHEP* **04** (2010) 126, `arXiv:0907.5413 [hep-ph]`.

[17] M. de Roo, H. Suelmann, and A. Wiedemann, "Supersymmetric R^4-actions in ten dimensions," *Phys. Lett.* **B280** (1992) 39–46.

[18] A. A. Tseytlin, "Heterotic – type I superstring duality and low-energy effective actions," *Nucl. Phys.* **B467** (1996) 383–398, `arXiv:hep-th/9512081 [hep-th]`.

[19] I. Jack, D. R. T. Jones, and N. Mohammedi, "The four-loop metric β-function for the bosonic σ-model," *Phys. Lett.* **B220** (1989) 171.

[20] I. Jack, D. R. T. Jones, and N. Mohammedi, "A four-loop calculation of the metric β-function for the bosonic σ-model and the string effective action," *Nucl. Phys.* **B322** (1989) 431.

[21] D. J. Gross and E. Witten, "Superstring modifications of Einstein's equations," *Nucl. Phys.* **B277** (1986) 1.

[22] D. J. Gross and J. H. Sloan, "The quartic effective action for the heterotic string," *Nucl. Phys.* **B291** (1987) 41–89.

[23] Y. Kikuchi, C. Marzban, and Y. J. Ng, "Heterotic string modifications of Einstein's and Yang-Mills' actions," *Phys. Lett.* **B176** (1986) 57.

[24] K. Becker, M. Becker, and J. H. Schwarz, *String Theory and M-Theory: A Modern Introduction.* Cambridge University Press, 2007.

[25] Z. Bern, C. Cheung, H.-H. Chi, S. Davies, L. Dixon, and J. Nohle, "Evanescent Effects Can Alter Ultraviolet Divergences in Quantum Gravity without Physical Consequences," `arXiv:1507.06118 [hep-th]`.

[26] K. Melnikov and V. G. Serbo, "Processes with the t-channel singularity in the physical region: finite beam sizes make cross-sections finite," *Nucl. Phys.* **B483** (1997) 67–82, `arXiv:hep-ph/9601290 [hep-ph]`.

[27] H. van Dam and M. J. G. Veltman, "Massive and mass-less Yang-Mills and gravitational fields," *Nucl. Phys.* **B22** (1970) 397–411.

[28] V. I. Zakharov, "Linearized gravitation theory and the graviton mass," *JETP Lett.* **12** (1970) 312.

[29] A. Martin, "Extension of the axiomatic analyticity domain of scattering amplitudes by unitarity. 1.," *Nuovo Cim.* **A42** (1965) 930–953.

[30] A. Nicolis, R. Rattazzi, and E. Trincherini, "Energy's and amplitudes' positivity," *JHEP* **05** (2010) 095, `arXiv:0912.4258 [hep-th]`. [Erratum: JHEP11,128(2011)].

[31] S. Sarkar and A. C. Wall, "Second law violations in Lovelock gravity for black hole mergers," *Phys. Rev.* **D83** (2011) 124048, `arXiv:1011.4988 [gr-qc]`.

[32] H. Elvang, D. Z. Freedman, L.-Y. Hung, M. Kiermaier, R. C. Myers, and S. Theisen, "On renormalization group flows and the *a*-theorem in 6d," *JHEP* **10** (2012) 011, `arXiv:1205.3994 [hep-th]`.

[33] C. Cheung and G. N. Remmen, "Naturalness and the Weak Gravity Conjecture," *Phys. Rev. Lett.* **113** (2014) 051601, `arXiv:1402.2287 [hep-ph]`.

[34] D. Baumann, D. Green, H. Lee, and R. A. Porto, "Signs of Analyticity in Single-Field Inflation," `arXiv:1502.07304 [hep-th]`.

[35] D. Croon, V. Sanz, and J. Setford, "Goldstone Inflation," `arXiv:1503.08097 [hep-ph]`.

[36] B. Heidenreich, M. Reece, and T. Rudelius, "Weak Gravity Strongly Constrains Large-Field Axion Inflation," `arXiv:1506.03447 [hep-th]`.

[37] A. de la Fuente, P. Saraswat, and R. Sundrum, "Natural Inflation and Quantum Gravity," *Phys. Rev. Lett.* **114** no. 15, (2015) 151303, `arXiv:1412.3457 [hep-th]`.

[38] J. Brown, W. Cottrell, G. Shiu, and P. Soler, "Fencing in the Swampland: Quantum Gravity Constraints on Large Field Inflation," `arXiv:1503.04783 [hep-th]`.