Causality constraints on corrections to Einstein gravity

Simon Caron-Huot,[a] **Yue-Zhou Li,**[a] **Julio Parra-Martinez**[b] **and David Simmons-Duffin**[b]

[a]*Department of Physics, McGill University,*
3600 Rue University, Montréal, H3A 2T8, QC Canada

[b]*Walter Burke Institute for Theoretical Physics, Caltech,*
Pasadena, California 91125, U.S.A.

E-mail: schuot@physics.mcgill.ca, liyuezhou@physics.mcgill.ca,
jparram@caltech.edu, dsd@caltech.edu

ABSTRACT: We study constraints from causality and unitarity on $2 \to 2$ graviton scattering in four-dimensional weakly-coupled effective field theories. Together, causality and unitarity imply dispersion relations that connect low-energy observables to high-energy data. Using such dispersion relations, we derive two-sided bounds on gravitational Wilson coefficients in terms of the mass M of new higher-spin states. Our bounds imply that gravitational interactions must shut off uniformly in the limit $G \to 0$, and prove the scaling with M expected from dimensional analysis (up to an infrared logarithm). We speculate that causality, together with the non-observation of gravitationally-coupled higher spin states at colliders, severely restricts modifications to Einstein gravity that could be probed by experiments in the near future.

Contents

1 Introduction

Einstein's theory of general relativity (GR) has been extraordinarily successful since its inception over a century ago. Nevertheless, modifications to GR are often discussed in relation to various puzzles, such as the nature of dark energy and dark matter, see [1–3] for reviews. An important class of modifications add higher-derivative terms to the equations of motion, which lead to physical effects which grow at short distances. Such corrections, handily classified in [4], arise naturally in string theory, and presumably in any UV-complete theory of quantum gravity.

In this paper, we consider $2 \to 2$ scattering of gravitons and ask a simple question: assuming that graviton scattering respects causality at *all* energies, by how much can the

low-energy amplitude differ from the predictions of general relativity? A well-known result by [5, 6] shows any Lorentz-invariant theory of massless spin-2 particles must reproduce general relativity at large distances. Our goal will be to bound the corrections to this limit, assuming that relativistic causality (as we understand it) holds.

Our answer will depend on the spectrum of the theory. Let us describe our setup and assumptions. At low energies, we assume there exists a massless graviton, together with a finite number of fields of spin ≤ 2, which can be described by an effective field theory (EFT). Schematically, the low-energy effective action encodes modifications to Einstein gravity:

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{g} \left(R + g_{R^{(3)}} \mathrm{Riem}^3 + g_{R^{(4)}} \mathrm{Riem}^4 + \dots \right) + S_{\mathrm{matter}}, \qquad (1.1)$$

where Riem^n denotes (possibly non-unique) contractions of products of n Riemann tensors.

An important idea is that the sizes of the Wilson coefficients $g_{R^{(3)}}, g_{R^{(4)}}, \dots$ are constrained by causality, that is, the notion that signals cannot travel faster than light. For example in [7] it was observed by Camanho-Edelstein-Maldacena-Zhiboedov (CEMZ) that in the presence of $g_{R^{(3)}}$, the two polarization modes of the graviton would move at different velocities in certain backgrounds, and inevitably one of them moves faster than "light". By considering a setup with large enough black holes and mirrors, this effect could lead to closed timelike curves, with ensuing grandparent-type paradoxes; the conclusion is that a classical theory with $g_{R^{(3)}} \neq 0$ is inconsistent.[1] Ref. [7] further pointed out that paradoxes can be avoided at the quantum level if the graviton couples to higher-spin states. Denoting the mass of the lightest higher-spin state (spin 4 or higher) by M, this led to a parametric bound: $|g_{R^{(3)}}| \lesssim \frac{1}{M^4}$.

Our main goal in this paper will be to quantitatively bound higher-derivative corrections in (1.1) in terms of the mass M of higher-spin states. This mass M provides a UV-cutoff scale for the low-energy EFT in (1.1). We will assume a large hierarchy between the Plank scale and this cutoff

$$M^2 \ll M_{\mathrm{pl}}^2, \qquad (1.2)$$

or, equivalently, $GM^2 \ll 1$, so that gravity is weakly-interacting below the cutoff. Our methods will test causality of graviton scattering with arbitrary center-of-mass energies, although physically the most important region for us will be near the cutoff M.

The notion of causality is subtle in gravitational EFTs because there is no globally well-defined lightcone in nontrivial backgrounds. As discussed recently in [10, 11], one should contrast "asymptotic causality," which exploits a fixed causal structure at large distances, with "infrared causality" which compares local time delays between species [11]. We will use asymptotic causality, but crucially, imposed at all energy scales and not only within the EFT regime. This leads to sharp mathematical statements involving crossing symmetry, analyticity, and Regge boundedness of scattering amplitudes [12]. Many works have examined how these conditions constrain EFTs and their UV completions, see e.g. [13–37] and references therein. In particular, these conditions give rise to dispersion relations that relate high and low energies. In some cases, dispersion relations can be interpreted as

---

[1] A related effect is that the classical initial-value problem is generically ill-posed when higher-derivative terms dominate [8, 9].

expressing the commutativity of coincident shockwaves [38]. Initially, the use of dispersion relations in gravitational EFTs was hindered by divergences related to the forward limit of the low-energy graviton amplitude. These technical issues were recently partially overcome in [15] by studying dispersion relations in impact parameter space; hence our renewed interest in this problem.

There is a natural motivation to study scattering events which have center-of-mass energies above the "cutoff" M. In any scenario where a higher-derivative correction to GR might be observed in an astrophysical or other large-distance process, the suppression scale would have to be very low, $M \ll 1 \text{ TeV} = (2 \times 10^{-19}\text{m})^{-1}$: energies above such a cutoff are routinely probed at colliders. However, collider experiments have not yet reported any higher-spin particles of the type suggested above. Is it at all possible to modify GR in a way that simultaneously: (1) satisfies collider constraints, (2) is relevant at large scales, and (3) respects causality as we understand it?

In this paper we take a modest step toward answering this question, by quantitatively relating higher-derivative corrections to the mass M of higher-spin states, using methods from [15]. Our main results will be that dimensionless ratios, of the schematic form $g_{R^{(3)}} M^4$ or $g_{R^{(4)}} M^6$, are bounded by order-unity constants, times an infrared logarithmic divergence $\log(M/m_{\text{IR}})$. Alternatively, given a measurement of such couplings, we constrain the mass M of new states and their couplings to gravitons. The task of bounding their couplings to Standard Model fields, as well as and potential collider signals, is left to future work.

The operational definition of "gravity" in this paper is a force which grows linearly with energy at high speeds, corresponding in particle physics language to exchange of a spin-2 particle. We stress that static long-range forces, which could come from direct interactions between matter and light spin-0 or spin-1 particles (also sometimes called fifth forces), are unconstrained by our arguments.

The infrared logarithmic divergence in our bounds is related to the divergence of the eikonal phase in four dimensions. In the context of scalar scattering in AdS/CFT, this logarithmic divergence gets regulated by the AdS curvature scale, yielding rigorous, finite bounds proportional to $\log M R_{\text{AdS}}$ [39]. We expect the same mechanism to apply to graviton scattering as well. Thus, our bounds can be interpreted as finite bounds on gravitational Wilson coefficients in AdS$_4$. The key feature of our bounds is the absence of power-law infrared divergences; eventually, we hope that infrared logarithms can be removed by studying suitable IR finite observables. In our view, an incredibly conservative assumption would be to replace m_{IR} with the Hubble scale, which for $M \sim 1$ TeV, multiplies our bounds by a factor of only $\log M/m_{\text{IR}} \sim 100$. Note that this would still strongly rule out a modification of GR that simultaneously satisfies (1), (2), and (3) above.

This paper is organized as follows. In section 2, we state our assumed axioms encoding causality of graviton scattering, review dispersive sum rules, and some of their known implications, notably positivity bounds from forward limits. In section 3, we review the impact parameter approach of [15] and provide example positive functionals which prove upper bounds on gravitational EFT coefficients. We explain why the bounds are only weakly affected by possible light matter fields. We then describe our numerical strategy to systematically search for optimal bounds, and how CEMZ-like bounds are automatically included. In section 4, we report the bounds obtained with this method and comment on

their relations with known theories. In a more conjectural part 4.5 we speculate about the possibility of modifications associated with a low scale M. We summarize in section 5. In appendix A, we review partial waves for graviton scattering amplitudes. In appendix B we present graviton amplitudes from light exchanges of spin-0 and spin-2 particles. Finally, we record our numerical set-ups in appendix C.

2 Review: dispersive sum rules

In this section, we describe our key physical assumptions and tools: effective field theory at low energies (section 2.1), unitarity (section 2.2), causality (section 2.3), and Kramers-Kronig-type dispersion relations and some of their consequences (section 2.4). Most of this material is standard, except perhaps for our discussion of Regge boundedness in section 2.3, using compact-support wavefunctions.

Let us first briefly motivate our assumptions, before we state them technically.

By unitarity, we mean that initial and final states of scattering processes are elements of positive-definite Hilbert spaces, whose norms are conserved by time evolution. This represents the idea that the probabilities of all possible events are positive and add up to one. The reason we assume this is obvious: we wouldn't know how to interpret negative probabilities.

By causality, we refer to the notion that "signals can't travel faster than light". We assume causality because of its tremendous explanatory power and past successes: by forbidding instantaneous action at a distance, it qualitatively explains why electromagnetic and gravitational waves *must* exist, why forces in nature are mediated by particles, why antiparticles exist, how their interactions are quantitatively related [40], and so much more. Abandoning causality without a good replacement principle seems akin to opening Pandora's box.[2]

There is an interesting interplay between unitarity and causality, as displayed by quantum fields with wrong-sign kinetic terms, sometimes called "ghosts". In one quantization, positive-frequency modes propagate forward in time but have negative norms. In an alternative quantization choice, norms are fine but positive-frequency modes propagate backward in time. Lee and Wick famously proposed that the ensuing acausality could be made unobservably small if one treats backward-moving modes as resonances which decay to normal forward-moving modes [41], as is indeed seen at low orders in perturbation theory [42, 43]. A problem is that, as soon as interactions with normal matter are included, negative-frequency modes make the vacuum unstable against resonant particle production. This instability can be nicely discussed in connection with a classical theorem by Ostrogradsky [44]; it seems incompatible with a long-lived Universe [45].[3]

---

[2]If one were not worried about instantaneous action at a distance, any many-body Hamiltonian such as

$$H = \sum_i \frac{\vec{p_i}^2}{2m_i} + \sum_{i<j} V_{ij}$$

with potential $V_{ij} = \frac{-Gm_i m_j}{|\vec{x}_i - \vec{x}_j|}$, or others, would trivially define a "quantum theory of gravity".

[3]As was pointed out with much deference to the original authors, the Lee-Wick prescription is ambiguous and "an additional prescription would be needed to completely define the theory" [46]. In our view, the Lee-Wick idea fails to address the vacuum stability issue for the simple reason that the timescales in the relevant vacuum diagram are shorter than the decay time of Lee-Wick quanta.

In short, we assume unitarity and causality because *we see no alternatives*. It is possible that Nature does not conform to these principles as we understand them, but resulting bounds can be viewed as tests of these principles.

2.1 Helicity amplitudes and low-energy EFT

Four-dimensional gravitons possess two helicity states. A complete set of independent amplitudes for graviton-graviton scattering is

$$\mathcal{M}(1^+2^-3^-4^+) = \langle 23 \rangle^4 [14]^4 \, f(s,u) \,, \tag{2.1}$$

$$\mathcal{M}(1^+2^+3^+4^-) = ([12][13]\langle 14 \rangle)^4 \, g(s,u) \,, \tag{2.2}$$

$$\mathcal{M}(1^+2^+3^+4^+) = \frac{[12]^2[34]^2}{\langle 12 \rangle^2 \langle 34 \rangle^2} \, h(s,u) \,, \tag{2.3}$$

and permutations thereof. At tree-level in Einstein's theory, only the maximal-helicity-violating (MHV) amplitude $f(s,u)$ is nonvanishing, and a large fraction of our results will be derived by studying only this amplitude. Here, we use spinor-helicity variables (see [47]) and we have introduced the Mandelstam invariants

$$s = -(p_1 + p_2)^2 \,, \qquad t = -(p_2 + p_3)^2 \,, \qquad u = -(p_1 + p_3)^2 \,, \tag{2.4}$$

with $s + t + u = 0$. The functions $f(s,u)$, $g(s,u)$, and $h(s,u)$ are analytic in the upper-half plane with $\operatorname{Im} s > 0$, and crossing symmetric:

$$f(s,u) = f(u,s) \,, \qquad g(s,u) = g(t,s) = g(u,t) \,, \qquad h(s,u) = h(t,s) = h(u,t) \,. \tag{2.5}$$

Other helicity amplitudes may be obtained by complex conjugation and Schwarz reflection, for example the $---+$ amplitude is $\overline{g}(s,u) = (g(s^*,u^*))^*$.

At low energies, we assume a spectrum comprising massless gravitons together with possible light particles of spin ≤ 2. These can be described by an effective field theory (EFT) of the generic form (1.1), with higher derivative terms encoding modifications to Einstein gravity generated by physics above the EFT cutoff M. We assume $M \ll M_{\rm pl}$ and thus neglect loops within the EFT. States with spin greater than two are genuinely gravitational, and assumed to have mass above the cutoff, $m > M$.

By contrast, particles of spin two or less and mass $m_\ell \ll M$ can be interpreted as additional states in the Standard Model of particle physics and its extensions, or states arising from Kaluza-Klein reduction of massless gravity in higher-dimensions. Angular momentum conservation forbids the decay of a state of half-integer or odd spin to a pair of gravitons, so only matter fields of spins 0 and 2 can affect graviton scattering at tree level. We will refer to both as "matter", even though this nomenclature is slightly unconventional for spin 2 fields.

The best way to enumerate EFT couplings is to list how they modify graviton scattering amplitudes. On-shell three-particle vertices are determined by Lorentz invariance up to overall parameters

$$\mathcal{M}(1^+, 2^+, 3^-) = \sqrt{8\pi G} \frac{[12]^6}{[13]^2[23]^2} \,, \qquad \mathcal{M}(1^+, 2^+, 3^+) = \frac{\widehat{g}_3}{2} \sqrt{8\pi G} ([12][13][23])^2 \,. \tag{2.6}$$
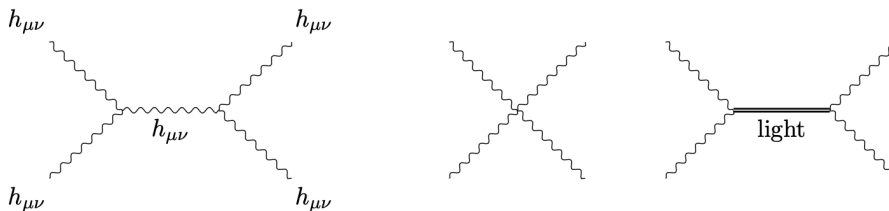
Figure 1. 2-to-2 scattering amplitudes of gravitons within the low-energy effective theory. We include at tree-level both the graviton exchange and (higher-derivative) contact diagrams, as well as exchanges of possible light spin-0 and spin-2 particles. Other spins are forbidden by angular momentum conservation.

The tree-level four-particle amplitudes (2.3) may then be written in terms of exchange diagrams, plus a sum of contact interactions, which are simply polynomials with the symmetry (2.5):

$$f_{\text{low}}(s,u) = \frac{8\pi G}{stu} + \frac{2\pi G su}{t}|\widehat{g}_3|^2 + g_4 + g_5 t + g_6\, t^2 - g_6'\, su + \ldots + f_{\text{matter}}(s,u)$$
$$+ O(\text{loops})\,, \tag{2.7a}$$

$$g_{\text{low}}(s,u) = \frac{4\pi G}{stu}\widehat{g}_3 + \frac{1}{2}\widehat{g}_6'' + \ldots + g_{\text{matter}}(s,u) + O(\text{loops})\,, \tag{2.7b}$$

$$h_{\text{low}}(s,u) = 40\pi G\widehat{g}_3 stu + \frac{1}{2}\widehat{g}_4(s^2 + t^2 + u^2)^2 + 2\widehat{g}_5 stu(s^2 + t^2 + u^2)$$
$$+ \widehat{g}_6(s^2 + t^2 + u^2)^3 + \widehat{g}_6' s^2 t^2 u^2 + \ldots + h_{\text{matter}}(s,u) + O(\text{loops})\,, \tag{2.7c}$$

where hatted couplings are complex (the real and imaginary part representing parity-even and parity-odd couplings, respectively). The subscript "low" emphasizes that this expansion is used only for $|s| < M^2$. The signs on the first line have been chosen so that our couplings relate simply to those in [48].[4] The matter contributions $f_{\text{matter}}(s,u)$, $g_{\text{matter}}(s,u)$, and $h_{\text{matter}}(s,u)$ are recorded in appendix B.

It is straightforward to write down Lagrangians that give rise to the above amplitudes. Before doing so, it is important to note that Lagrangian densities are only defined modulo field redefinitions (which change contact interactions by equation of motions) and total derivatives. In particular, any higher-derivative term involving the Ricci tensor $R_{\mu\nu}$ or scalar R is removable, so only powers of the Riemann curvature $R_{\mu\nu\sigma\rho}$ must be kept.[5] Furthermore, numerous identities relate various contractions of Riemann tensors and derivatives. This is the reason why we do not include R^2: R^2-terms can be recast into the Gauss-Bonnet term, which is topological in $d = 4$. In contrast, the amplitudes (2.7) are unambiguous.

---

[4]The conversion is simply:

$$\{g_4,\ g_5,\ g_6,\ g_6'\}^{\text{here}} = \{a_0,\ a_1,\ a_{2,0},\ a_{2,1}\}^{\text{there}}.$$

In our notation the subscript always denotes half the number of derivatives in the contact interaction.

[5]It is well-known for example that $f(R)$ gravity is equivalent to standard Einstein gravity minimally coupled to a scalar field with a specific potential. From our perspective, $f(R)$ gravity thus does not constitute a higher-derivative correction to Einstein's gravity. Instead, it is a specific choice of matter sector.

With this being said, it is straightforward to list a minimal set of irreducible higher-dimension operators and map them to the amplitudes (2.7) by computing the resulting tree-level amplitudes. For example, the parity-even sector of cubic gravity contains 10 different operators, but field redefinitions and various identities leave us with only one independent operators [49]. Up to dimension eight, our effective action is

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \Big[R - \frac{1}{3!} \left(\alpha_3 R^{(3)} + \tilde{\alpha}_3 \tilde{R}^{(3)} \right) $$
$$+ \frac{1}{4} \left(\alpha_4 (R^{(2)})^2 + \alpha_4' (\tilde{R}^{(2)})^2 + 2\tilde{\alpha}_4 R^{(2)} \tilde{R}^{(2)} \right) + \dots \Big] + S_{\text{matter}} \,, \tag{2.8}$$

where we defined

$$R^{(2)} = R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} \,, \qquad \tilde{R}^{(2)} = R_{\mu\nu\rho\sigma} \tilde{R}^{\mu\nu\rho\sigma} \,, \qquad \tilde{R}_{\mu\nu\rho\sigma} \equiv \tfrac{1}{2} \epsilon_{\mu\nu}{}^{\alpha\beta} R_{\alpha\beta\rho\sigma} \,,$$
$$R^{(3)} = R_{\mu\nu}{}^{\rho\sigma} R_{\rho\sigma}{}^{\alpha\beta} R_{\alpha\beta}{}^{\mu\nu} \,, \qquad \tilde{R}^{(3)} = R_{\mu\nu}{}^{\rho\sigma} R_{\rho\sigma}{}^{\alpha\beta} \tilde{R}_{\alpha\beta}{}^{\mu\nu} \,. \tag{2.9}$$

It is then straightforward to expand $g_{\mu\nu} = \eta_{\mu\nu} + \sqrt{32\pi G} h_{\mu\nu}$ and apply the standard Feynman techniques to evaluate scattering amplitudes and compare with eqs. (2.7):

$$\hat{g}_3 = \alpha_3 + i\tilde{\alpha}_3, \qquad g_4 = 8\pi G(\alpha_4 + \alpha_4') \,, \qquad \hat{g}_4 = 8\pi G(\alpha_4 - \alpha_4' + i\tilde{\alpha}_4) \,. \tag{2.10}$$

Note that we absorbed a factor of $8\pi G$ in three-point couplings but not in four-point couplings.

2.2 High energies: partial waves and unitarity

We will assume that graviton scattering remains sensible even at center-of-mass energies that exceed the EFT cutoff M (where the parametrization (2.7) no longer applies). Our minimal assumptions are that the amplitude remains causal (that is, analytic) and unitary, and that the spectrum is relativistic so that it can be organized in terms of mass, m^2, and spin, J.

In other words, the amplitude admits a partial wave expansion of the form [50]

$$\mathcal{M}(1^{h_1} 2^{h_2} 3^{h_3} 4^{h_4}) = 16\pi \sum_J (2J+1) a_J^{\{h\}}(s) d_{h_{12}, h_{34}}^J \left(1 + \tfrac{2t}{s} \right) \,. \tag{2.11}$$

Here, $d_{\alpha,\beta}^J(x)$ are the well-known Wigner-D functions, which are explicitly written in appendix A, and $h_{ij} = h_i - h_j$. The partial wave coefficients, $a_J^{\{h\}}(s)$, encode all dynamical information. Without boost invariance, partial waves would be more complicated, but study of causality constraints has been initiated in [51, 52].

Unitarity of the S-matrix $S = 1 + i\mathcal{M}$ imposes crucial positivity properties on the "absorptive part" of the amplitude, through the familiar relation: $i(\mathcal{M}^\dagger - \mathcal{M}) = \mathcal{M}^\dagger \mathcal{M}$. The matrix structure will be important and is in contrast to the scalar case studied in [13–19]. In terms of partial waves,

$$i \left[\left(a_J^{-h_4, -h_3, -h_2, -h_1}(s) \right)^* - a_J^{h_1, h_2, h_3, h_4}(s) \right] = \sum_X \left(a_J^{-h_3, -h_4 \to X}(s) \right)^* a_J^{h_1, h_2 \to X}(s), \quad (2.12)$$

where X runs over intermediate states. The right-hand-side is a positive semi-definite matrix. To capture its positive properties we adopt an abbreviated notation from [53] and omit the X sum, writing the right-hand-side simply as $2(c^{-h_4,-h_3})^* c^{h_1 h_2}$. Specializing to the MHV amplitude, we have

$$\text{Im}\, a_J^{+-+-}(s) = |c_{J,s}^{+-}|^2, \qquad \text{Im}\, a_J^{++--}(s) = |c_{J,s}^{++}|^2, \tag{2.13}$$

where $c_{J,s}^{+-}$ is real. In particular, the quantities in (2.13) are positive.[6] Note this would not be so for the permutation $a_J^{+--+} = (-1)^J a_J^{+-+-}$. Partial waves admit two-sided bounds, which follow from applying the same argument to the unitary matrix $-S = 1 + i(2i - \mathcal{M})$. In particular, we have[7]

$$0 \le \text{Im}\, a_J^{+-+-}(s) \le 1, \quad \text{and} \quad 0 \le \text{Im}\, a_J^{++--}(s) \le 2. \tag{2.14}$$

Positivity of the spectral density is key for establishing bounds on the low-energy EFT Wilson coefficients [13–18].

Explicitly, the MHV amplitude $f(s,u)$ has distinct discontinuities in the s- and t- channels:

$$s = m^2 > 0: \qquad \text{Im}\, f(m^2, -p^2) = \frac{16\pi}{m^8} \sum_{J \ge 4} (2J+1)|c_{J,m^2}^{+-}|^2\, \tilde{d}_{4,4}^J \left(1 - \frac{2p^2}{m^2}\right), \tag{2.15a}$$

$$t = m^2 > 0: \qquad \text{Im}\, f(-p^2, p^2 - m^2) = \frac{16\pi}{m^8} \sum_{\substack{J \ge 0 \\ \text{even}}} (2J+1)|c_{J,m^2}^{++}|^2\, \tilde{d}_{0,0}^J \left(1 - \frac{2p^2}{m^2}\right), \tag{2.15b}$$

where $\tilde{d}_{\alpha,\beta}^J$ are Wigner-D functions with stripped helicity factors, see appendix A for more details. The overall m^{-8} originates from the prefactor in (2.3).

For other helicity configurations, we have similar relations, except that the "imaginary part" gets replaced by the discontinuity $\widetilde{\text{Im}}\, a \equiv [a(s + i0) - a(s - i0)]/(2i)$, and the right-hand-sides are now complex numbers:

$$\widetilde{\text{Im}}\, a_J^{+++-}(s) = c_{J,s}^{++} c_{J,s}^{+-}, \qquad \widetilde{\text{Im}}\, a_J^{++++}(s) = (c_{J,s}^{++})^2. \tag{2.16}$$

The corresponding partial wave expansions are

$$\widetilde{\text{Im}}\, g(m^2, -p^2)\Big|_{s=m^2} = \frac{16\pi}{m^{12}} \sum_{\substack{J \ge 4 \\ \text{even}}} (2J+1) c_{J,s}^{++} c_{J,s}^{+-}\, \tilde{d}_{4,0}^J \left(1 - \frac{2p^2}{m^2}\right), \tag{2.17}$$

$$\widetilde{\text{Im}}\, h(m^2, -p^2)\Big|_{s=m^2} = 16\pi \sum_{\substack{J \ge 0 \\ \text{even}}} (2J+1)(c_{J,s}^{++})^2\, \tilde{d}_{0,0}^J \left(1 - \frac{2p^2}{m^2}\right). \tag{2.18}$$

---

[6]Even though $c_{J,s}^{+-}$ is real, we nonetheless write the absolute value sign $|c_{J,s}^{+-}|^2$ throughout to emphasize that its square is positive.

[7]In more detail: in the even spin sector, there are three incoming states for the helicities $|h_1 h_2\rangle$: namely $|++\rangle$, $|--\rangle$, and $\frac{1}{\sqrt{2}}(|+-\rangle + |-+\rangle)$. The corresponding diagonal elements of the S-matrix are $1 + ia^{--++}$, $1 + ia^{++--}$, and $1 + 2ia^{+-+-}$. In the odd-spin sector, the S-matrix is a 1×1 matrix with element $1 + 2ia^{+-+-}$. By unitarity, each of these diagonal elements must have real part in $[-1, 1]$, which leads to (2.14).

2.3 Regge boundedness and all that

Low and high energies are related by Kramers-Kronig-type dispersion relations. It will be crucial that we can predict beforehand which dispersion relations converge.

Typically one assumes a Froissart-Martin-like bound at fixed momentum transfer and large complex energies:

$$\lim_{|s| \to \infty} \mathcal{M}/s^2 \to 0 \qquad \text{at fixed} \qquad t < 0 \qquad (\textit{not} \text{ what we'll assume}). \tag{2.19}$$

For example, in tree-level string theory, $\mathcal{M} \sim s^{2+\alpha' t} < s^2$. However the validity of this bound is not generally established in an abstract theory of quantum gravity. Martin's original proof of the Froissart-Martin bound in axiomatic field theory [54] does not apply to gravity, due to the absence of a mass gap. For holographic theories it has been argued that the behavior (2.19) holds for physical kinematics as a consequence of the chaos bound [55, 56].

The constraint we will use is in fact weaker than (2.19): we will only assume that it holds after integrating in t against suitable wavefunctions. We also assume analyticity in the upper-half s-plane, at least for s large, as well as sub-exponential growth for $|s| \to \infty$, as described more precisely below. This is equivalent to assuming that the unsmeared amplitude satisfies dispersion relations with a finite number of subtractions, which is a weak form of UV locality.

The difficulty with (2.19) is a physical one and not merely technical: to bound amplitudes at large complex energies, one must generally combine analyticity with some boundedness property on the real axis, as we do shortly. The difficulty is that analyticity holds at fixed momentum, while boundedness holds at fixed impact-parameter; these two spaces are related by a Fourier transform which is not easy to control. Namely, it is not straightforward to estimate large-impact-parameter contributions in the absence of a mass gap or of an explicit model of the dynamics. Thankfully, large-impact-parameter physics however seems immaterial for bounding EFT couplings at the scale M. The intuition, stressed in [14, 16], is that EFT parameters at the scale M satisfy sum rules saturated by impact parameters $b \sim M^{-1}$.

Let us explain how we sidestep (2.19) by adapting a recent method from [39], which showed that the conclusions from flat space sum rules apply to quantum gravity in AdS (defined as a CFT with large but finite central charges and single-trace gap). The method is simple: we integrate scattering amplitudes against wavepackets that have finite support in momentum space and decay rapidly at large impact parameters b. Formally, for a wavefunction $\Psi(p)$, we define the smeared amplitude:

$$\mathcal{M}_\Psi(\tilde{s}) = \int_0^M dp \, \Psi(p) \mathcal{M}\left(\tilde{s} + \tfrac{1}{2}p^2, -p^2\right). \tag{2.20}$$

It is apparent that for $|\tilde{s}| > \frac{1}{2}M^2$, all amplitudes on the right-hand-side are in the physical region where the partial wave expansion (2.11) applies. (The offset of s by $\frac{1}{2}p^2$ is not essential but ensures that $s \leftrightarrow u$ crossing symmetry is simply reflection of $\tilde{s}$.) Furthermore, thanks to compactness of the integral, $\mathcal{M}_\Psi(\tilde{s})$ inherits the analyticity properties of the original

amplitude: our fundamental assumption is that a crossing path exists which connects the two points $\tilde{s} = \pm\frac{1}{2}M^2$, and that the amplitude is analytic outside of that arc.

Fast decay in b requires $\Psi(p)$ to be smooth and to vanish rapidly enough at the endpoints; the precise condition is detailed below (see (3.16)). The upshot is that if the decay sets in at some $b > b_*$, then the spin sum in (2.11) is effectively limited to $J \leq \sqrt{s}b_*$. Since individual a_J are bounded (see (2.14)), one trivially gets the bound

$$|\mathcal{M}_\Psi(s)| \leq s \times \text{constant} \qquad (|s| > \tfrac{1}{2}M^2, \text{ real}). \tag{2.21}$$

We thus have an analytic function which is bounded on the real axis. Unless this function grows exponentially at complex energies (which would imply blatant time advances when Fourier transformed to the time domain, a behavior which was not seen in theories of quantum gravity in AdS realized by unitary CFTs [39]), it must be bounded in all complex directions by a version of the maximum principle called Phragmén-Lindelöf principle (see [55]):

$$|\mathcal{M}_\Psi(s)| \leq |s| \times \text{constant} \qquad (\text{complex } s \text{ outside } s \sim M^2 \text{ arc}). \tag{2.22}$$

The results presented in this paper rely only on the above properties of smeared amplitudes $\mathcal{M}_\Psi$, and not on (2.19). It is intriguing that this reasoning produces an exponent in (2.22) that is formally stronger than that in (2.19): $|\mathcal{M}_\Psi/s|$ bounded versus $|\mathcal{M}_\Psi/s^2| \to 0$. The extra power will not be exploited in the present paper, but it would be interesting to see if it has any implications. We will only use that $\lim_{|s|\to\infty} |\mathcal{M}_\Psi(s)|/|s|^2 = 0$, which is easily implied by (2.22) and is strictly weaker than (2.19).[8]

To be fully explicit, since smearing is technically involved, in this paper we will only use smeared amplitudes when strictly required by the $t \to 0$ singularities of amplitudes. For higher-subtracted dispersion relations, that do not suffer from a graviton pole, we will effectively use an assumption similar to (2.19), namely $|\mathcal{M}/s^4| \to 0$ for $t < 0$, as detailed in section 3.4. This is similar to the assumption made in ref. [48].[9]

Now that our mathematical assumptions (analyticity and boundedness) have been stated precisely, we would like to explain why we believe these are conservative and physically relevant notions directly related to causality and unitarity. Analyticity is related to the notion that "signals cannot move faster than light" in the context of signals that are waves. For example for vacuum two-point functions in quantum field theory, well-known arguments show that the vanishing of spacelike commutators requires the existence of antiparticles. This happens because waves components of different frequencies can only interfere destructively in the spacelike region if they are related to each other in a very specific way, namely by an analytic continuation in the upper-half energy plane, which relates positive and negative frequencies. The same interpretation applies to Kramers-Kronig dispersion relations.

---

[8]After the first arXiv version of this article appeared, the paper [57] analyzed the high-energy behavior of smeared amplitudes and independently confirmed (2.22) for a suitable class of smearing functions in $d \geq 5$ where there are no infrared divergences.

[9]Note that the spin-2 dispersion relations that we use are nonperturbative statements that need not be satisfied order-by-order, for example by the individual one-loop diagrams studied in [48].

We interpret S-matrix dispersion relations in a similar way: for (asymptotic) measurements at space-like separated points A and B to commute (in the presence of other particles), the amplitude for a particle moving from A to B must be related, by analytic continuation, to that for an antiparticle moving the other way. In this way, causality is entwined with analyticity and crossing symmetry [58]. While we find this picture intuitive and compelling, we should note that analyticity is nontrivial to prove mathematically even in quantum field theory (for recent discussions see [59, 60]). In quantum gravity, of course, no axioms are established and rigorous proofs are impossible. That our assumptions are compatible with gravity is nonetheless supported by the recent work [39], which showed that well-established CFT axioms imply that graviton scattering in AdS space satisfies dispersion relations.

The implications of (2.22) depend on the helicity of scattered particles. Recalling that for physical kinematics $\langle ij \rangle = \pm[ij]^*$, and

$$|\langle 12 \rangle| = |\langle 34 \rangle| = \sqrt{|s|}\,, \qquad |\langle 23 \rangle| = |\langle 14 \rangle| = \sqrt{|t|}\,, \qquad |\langle 13 \rangle| = |\langle 24 \rangle| = \sqrt{|u|}\,, \quad (2.23)$$

we find that for the component amplitudes f, g and h from (2.3), the condition (2.22) yields:

$$\lim_{|s|\to\infty} f(s, -p^2) \leq Cs^{-3}\,, \qquad \lim_{|s|\to\infty} f(s, p^2 - s) \leq Cs\,,$$
$$\lim_{|s|\to\infty} g(s, -p^2) \leq Cs^{-3}\,, \qquad \lim_{|s|\to\infty} h(s, -p^2) \leq Cs\,, \quad \text{(after smearing in } p). \quad (2.24)$$

The first line gives respectively the fixed-u and fixed-t Regge limits of the MHV amplitude. Notice that certain limits enjoy improved behavior $\sim s^{-3}$: when amplitudes are normalized so that contact interactions are polynomial (see eq. (2.7)), they vanish in some high-energy limits. This phenomenon is known as superconvergence and is the main reason why we will find stronger constraints on graviton contact interactions than for scalars.[10]

Superconvergence is also related to the observation of [62] that a very limited number of graviton contact interactions obey (or more precisely, saturate) the classical bound (2.19). Although it is simpler to prove, the bound (2.22) is stronger and is not satisfied by any individual graviton contact interaction.

2.4 Dispersive sum rules

We are now ready to write dispersive sum rules for the amplitudes $f(s,u)$, $g(s,u)$, and $h(s,u)$. We begin with the MHV amplitude f. From the behavior (2.24), we get two types of constraints: from fixed-u and fixed-t. For fixed-u we can separate f into combinations

---

[10]In general, superconvergence occurs in scattering of particles of spins J_1 and J_2 whenever $J_1 + J_2 - 1 > J_0$, where J_0 is the Regge intercept of the theory, see e.g. [38, 61]. The bound (2.22) amounts to $J_0 \leq 1$ but all we ultimately use in this paper is $J_0 < 2$.

that are even/odd under $s \leftrightarrow t$ and obtain the following basis of sum rules, for integer k:

$$B_k^{(1)}(p^2) = \oint_\infty \frac{ds}{4\pi i} \frac{2s - p^2}{[s(s-p^2)]^{\frac{k-2}{2}}} \left[f(s, -p^2) + f(p^2 - s, -p^2) \right] = 0 \qquad (k \geq 2 \text{ even}),$$
$$(2.25\text{a})$$

$$B_k^{(1)}(p^2) = \oint_\infty \frac{ds}{4\pi i} \frac{1}{[s(s-p^2)]^{\frac{k-3}{2}}} \left[-f(s, -p^2) + f(p^2 - s, -p^2) \right] = 0 \qquad (k \geq 3 \text{ odd}),$$
$$(2.25\text{b})$$

where the integrals are along a large circle at infinity. We additionally have three fixed-t dispersion relations, which also integrate to zero for $k \geq 2$ even:

$$\left\{ B_k^{(2)}, B_k^{(3)}, B_k^{(4)} \right\}(p^2) = \oint_\infty \frac{ds}{4\pi i} (2s - p^2) \left\{ \frac{f(s, p^2 - s)}{[s(s-p^2)]^{\frac{k+2}{2}}}, \frac{g(s, p^2 - s)}{[s(s-p^2)]^{\frac{k-2}{2}}}, \frac{h(s, p^2 - s)}{[s(s-p^2)]^{\frac{k+2}{2}}} \right\}.$$
$$(2.26)$$

To avoid confusion between different channels, we always write the fixed momentum transfer as p. These sum rules become strictly valid after the p-dependence is integrated against appropriate wavepackets as in eq. (2.20). In the limit $|s| \to \infty$ the subtractions only give an inverse power of s, so the convergence of the smeared sum rules in eqs. (2.25)–(2.26) directly follows from that of the amplitudes in eq. (2.24).

The subscript k indicates the Regge spin of a sum rule. This concept is closely related, but distinct, from the "number of subtractions" or power of $1/s$ inserted to improve high-energy convergence. For example, $B_2^{(1)}$ has fewer subtractions than $B_2^{(2)}$ (and is even "anti-subtracted" since it has no denominator!), yet they possess the same convergence properties. The nomenclature is motivated by the fact that exchange of a single t-channel particle of spin J yields an amplitude that grows like $\mathcal{M} \sim s^J$: we say that a sum rule has spin k if it converges on exchanges with $J < k$ (and marginally diverges on spin k).

Regge spin is more important than subtraction-counting because the Regge growth (2.22) translates into the simple convergence criterion $k > 1$. This is the same criterion as convergence of the Froissart-Gribov formula which extracts partial waves of spin $J > 1$, or of the analogous Lorentzian inversion formula [63–65] which extracts CFT data for spin $J > 1$.

Sum rules are obtained by deforming the contour towards the real axis but avoiding the low-energy region: the contour in figure 2 relates low-energy data at the scale M and heavy data above M:

$$-B_k^{(i)}(p^2) \Big|_{\text{low}} = B_k^{(i)}(p^2) \Big|_{\text{high}}$$
$$(2.27)$$

$$\oint_{s=M^2}^{u=M^2} \frac{ds}{4\pi i} (\cdots) = \int_{M^2}^\infty \frac{ds}{\pi} \text{Im} (\cdots).$$
$$(2.28)$$

Note that the s and u channel cuts contribute identically due to symmetry of eqs. (2.25), so we included only the right cut. (The contour on the left is in reality the union of upper and lower half-circles, separated by the branch cut of the amplitude.)
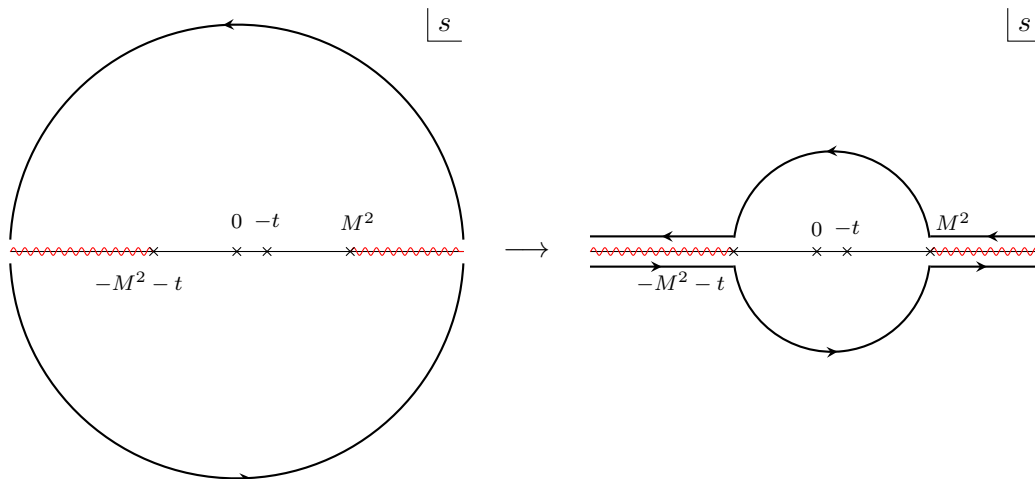
Figure 2. Contour deformation which gives rise to sum rules eq. (2.27). The final contour relates low-energy EFT data along the arcs to heavy discontinuities along the branch cuts.

Let us focus on the first sum rule for simplicity. At tree-level we find only two residues, from $s = 0$ and $u = 0$, which contribute the same amount:

$$-B_k^{(1)}(p^2)\Big|_{\text{low}} = \underset{s=0}{\text{Res}} \left(\frac{2s - p^2}{[s(s - p^2)]^{\frac{k-2}{2}}} \left[f(s, -p^2) + f(p^2 - s, -p^2)\right] \right) \quad \text{(tree-level)}. \quad (2.29)$$

Substituting in the low-energy amplitude (2.7a), only the exchange graphs contribute for $k = 2, 3$:

$$-B_2^{(1)}(p^2)\Big|_{\text{low}} = \frac{16\pi G}{p^2} + 2\pi G|\widehat{g}_3|^2 p^6 + O(\text{matter and loops}), \quad (2.30a)$$

$$-B_3^{(1)}(p^2)\Big|_{\text{low}} = -2\pi G|\widehat{g}_3|^2 p^4 + O(\text{matter and loops}). \quad (2.30b)$$

The absence of contact term contributions is a hallmark of superconvergent sum rules. Examples that probe contact interactions include:

$$-B_4^{(1)}(p^2)\Big|_{\text{low}} = 2g_4 + (4\pi G|\widehat{g}_3|^2 + g_5)p^2 + (g_6 + g_6')p^4 + \dots, \quad (2.31a)$$

$$-B_5^{(1)}(p^2)\Big|_{\text{low}} = g_5 + (g_6 - g_6')p^2 + \dots, \quad (2.31b)$$

$$-B_2^{(2)}(p^2)\Big|_{\text{low}} = 2\pi G|\widehat{g}_3|^2 \frac{1}{p^2} + g_6' + \dots. \quad (2.31c)$$

A salient feature is that the same couplings appear in multiple sum rules: this reflects crossing symmetry. Another feature is the appearance of the cubic self-coupling in $B_4^{(1)}$: this is due to the rapid growth with t of the t-channel exchange diagram with derivatives. This rapid energy growth at zero impact parameter will turn out to be a powerful mechanism to bound $\widehat{g}_3$, as was proposed in section 7 of [62]; this mechanism is distinct from the spin-2 growth at large impact parameter that was exploited by CEMZ [7].

At high energies $s \geq M^2$, the amplitudes are beyond our knowledge. We can nevertheless evaluate the contribution to the dispersive sum rule, by inserting the partial wave

decomposition and using eq. (2.15) for even k:

$$B_k^{(1)}(p^2)\Big|_{\text{high}} = \left\langle (2m^2 - p^2) \frac{|c_{J,m^2}^{++}|^2 \, \tilde{d}_{0,0}^J(x) + |c_{J,m^2}^{+-}|^2 \, \tilde{d}_{4,4}^J(x)}{m^{k+4}(m^2 - p^2)^{\frac{k-2}{2}}} \right\rangle \qquad (k \geq 2 \text{ even}), \quad (2.32)$$

where $x = 1 - \frac{2p^2}{m^2}$ and for later convenience we define the heavy densities $\mathcal{C}$ and (dimensionless) averages[11]

$$B_k(p^2)\Big|_{\text{high}} = \left\langle \mathcal{C}_{k,-p^2}[m^2, J] \right\rangle = 16 \sum_J (2J+1) \int_{M^2}^{\infty} \frac{dm^2}{m^2} \mathcal{C}_{k,-p^2}[m^2, J]. \qquad (2.33)$$

The remaining four sum rules are:

$$B_k^{(1)}(p^2)\Big|_{\text{high}} = \left\langle \frac{|c_{J,m^2}^{++}|^2 \, \tilde{d}_{0,0}^J(x) - |c_{J,m^2}^{+-}|^2 \, \tilde{d}_{4,4}^J(x)}{m^{k+3}(m^2 - p^2)^{\frac{k-3}{2}}} \right\rangle \qquad (k \geq 3 \text{ odd}), \qquad (2.34a)$$

$$\left\{ B_k^{(2)}, B_k^{(3)}, B_k^{(4)} \right\}(p^2)\Big|_{\text{high}} =$$

$$\left\langle (2m^2 - p^2) \left\{ \frac{(-1)^J |c_{J,m^2}^{+-}|^2 \, \tilde{d}_{4,-4}^J(x)}{m^{k+8}(m^2 - p^2)^{\frac{k+2}{2}}}, \frac{c_{J,m^2}^{++} c_{J,m^2}^{+-} \, \tilde{d}_{4,0}^J(x)}{m^{k+8}(m^2 - p^2)^{\frac{k-2}{2}}}, \frac{(c_{J,m^2}^{++})^2 \, \tilde{d}_{0,0}^J(x)}{m^k(m^2 - p^2)^{\frac{k+2}{2}}} \right\} \right\rangle \qquad (k \geq 2 \text{ even}).$$

$$(2.34b)$$

In summary, the constraints of analyticity and unitarity used in this paper are embodied in the relation (2.27), which connects EFT couplings (eqs. (2.30)–(2.31)) to heavy averages like (2.32) that involve a positive measure.

2.5 Review of simple bounds from forward limits

Let us explore the above sum rules. Evidently, the spin-2 sum rules (2.30) diverge in the forward limit $p \to 0$. This is the famous "graviton pole" problem, whose resolution using smeared sum rules will be described in the next section. However, higher-spin sum rules have smooth limits which have been extensively studied.

For the spin-4 sum rule, the forward limit of (2.31a) and (2.32) yields simply [66, 67]:

$$g_4 = \left\langle \frac{|c_{J,m^2}^{++}|^2 + |c_{J,m^2}^{+-}|^2}{m^8} \right\rangle \geq 0. \qquad (2.35)$$

Similarly, taking the forward limit of the same-helicity amplitude ($B^{(4)}$ sum rule) we get

$$\widehat{g}_4 = \left\langle \frac{(c_{J,m^2}^{++})^2}{m^8} \right\rangle, \qquad (2.36)$$

where we recall that the complex couplings $\widehat{g}_4$ and c_{J,m^2}^{++} combine parity even and odd parts. Comparison yields the obvious bound

$$|\widehat{g}_4| \leq g_4. \qquad (2.37)$$

---

[11]To avoid unnecessary clutter, we set undefined coefficients to zero, *ie.* $c^{+-} = 0$ when $J < 4$ and $c^{++} = 0$ when J is odd.

In particular this implies the two positivity constraints [66, 67]: $\alpha_4, \alpha_4' \geq 0$, in the Lagrangian (2.8) and (2.10).

As stressed before, forward limits can be dangerous. Should we trust these bounds?

From the present perspective, one can argue that the answer is: yes, as long as the size of g_4 is much larger than the loop effects which cause the danger at loop level. Let us estimate this in detail. The one-loop amplitude diverges in the forward limit $p \to 0$ like $\mathcal{M} \sim i\pi \frac{G^2 s^3}{p^2} \log p^2$. This divergence can be avoided simply by replacing the limit with evaluation at a small scale $p \sim p_* \ll M$. (This will spoil positivity at large $b = \frac{2J}{m}$, but if we assume that this region is controlled by known low-energy-computable eikonal physics, positivity should not be crucial there.) The bound (2.35) becomes schematically (up to logarithms)

$$ g_4 \geq -\frac{G^2}{M^2 p_*^2} \qquad \text{including loops} \,. \tag{2.38} $$

On the other hand, as discussed in the next section, in the presence of spin-4 particles at the scale M the expected size is $g_4 \sim \frac{G}{M^6} \gg \frac{G^2}{M^4}$. By choosing appropriately p_*, it is easy to make the loop corrections negligible in comparison:

$$ \frac{G^2}{M^2 p_*^2} \ll \frac{G}{M^6} \quad \Rightarrow \quad \frac{M^2}{M_{\text{pl}}} \ll p_* \ll M \,. \tag{2.39} $$

This can be satisfied since we assume that the EFT cutoff is parametrically below the Planck scale. A similar argument was described in [67].

By the same token we can trust the Taylor expansion around the forward limit of other sum rules with spin $k \geq 3$. For $k = 5$ we find

$$ g_5 = \left\langle \frac{|c_{J,m^2}^{++}|^2 - |c_{J,m^2}^{+-}|^2}{m^{10}} \right\rangle \,, \tag{2.40} $$

from which the two-sided bound $-g_4 \leq g_5 M^2 \leq g_4$ readily follows (again up to loop corrections). From the forward limit of $B_6^{(1)}$, we have

$$ g_6 = \left\langle \frac{|c_{J,m^2}^{++}|^2 + |c_{J,m^2}^{+-}|^2}{m^{12}} \right\rangle \quad \Rightarrow \quad 0 \leq g_6 \leq \frac{g_4}{M^4} \,. \tag{2.41} $$

We get additional sum rules of the same scaling dimensions by considering also forward limit derivatives of $k = 4$ and $k = 5$ sum rules:

$$ g_6' = \left\langle \frac{1}{m^{12}} \left(\mathcal{J}|c_{J,m^2}^{++}|^2 - (\mathcal{J} - 22)|c_{J,m^2}^{+-}|^2 \right) \right\rangle \,, \tag{2.42a} $$

$$ 0 = \left\langle \frac{1}{m^{12}} \left(\mathcal{J}(\mathcal{J} - 6)|c_{J,m^2}^{++}|^2 + (\mathcal{J}^2 - 50\mathcal{J} + 596)|c_{J,m^2}^{+-}|^2 \right) \right\rangle \,, \tag{2.42b} $$

where $\mathcal{J} = J(J+1)$. The novelty is the presence of a "null constraint", that is, a sum rule with vanishing low-energy contribution. By taking linear combinations of (2.41) and (2.42)

that maintain the positivity of r.h.s., these reproduce the known bounds derived in [48]:[12]

$$g_6 \geq 0\,, \qquad -\frac{90}{11} \leq \frac{g_6'}{g_6} \leq 6\,. \tag{2.43}$$

We will see in section 4.4 that the above bounds are not optimal; we will get tighter ones by using more null constraints. In particular we will show that indeed $g_6' \geq 0$ (up to $1/M_{\rm pl}^2$ corrections from loops, like other bounds in this paper).

A generic observation is that all dimensionless ratios of the form $g_k M^{2(k-4)}/g_4$ satisfy two-sided bounds, consistent with dimensional analysis scaling. We do not have a formal proof for all couplings, but we expect this to hold in analogy with the scalar theory case studied in [15]. The novel feature for gravitons is that we will be able to upper-bound g_4 itself by a multiple of G/M^6, as we now discuss.

3 Bounds that relate gravity and higher derivatives

3.1 Impact parameter functionals

As we reviewed in the previous section, the simplest positivity bound is $g_4 \geq 0$, which is trivially established by evaluating $B_4^{(1)}$ at $p \to 0$. However, this says little about the size of g_4 since Newton's constant does not appear. On the other hand, one may expect that all graviton interactions shut down if $G = 0$, a concrete example being the CEMZ bound [7]. To probe Newton's constant with dispersion relations, we must inevitably use the spin-2 sum rules B_2 and deal with the $1/p^2$ pole. This precludes doing Taylor expansions around the forward limit.

Our strategy will follow [15]: we tame the graviton pole by considering processes with finite impact parameter $b \sim 1/M$. In spacetime dimensions $D > 4$ this removes all divergences. In $D = 4$, as we are considering in this paper, we will be left with infrared logarithms, still a major improvement over power-law divergences in our opinion. Being mindful of the compact-support property (see section 2.3), we consider functionals of the generic form:

$$\mathcal{F} = \int_0^M dp \sum_i \psi_i(p) B_i(p^2)\,, \tag{3.1}$$

where $B_i(p^2)$ generally denotes dispersive sum rules and $\psi_i(p)$ is compact functions in p. The reader might worry that the integral in eq. (3.1) is done all the way up to the cutoff, pushing the convergence of the EFT expansion to the limit. For the time being we will focus on super-convergent sum rules, for which this is not an issue. In section 3.4 we will review how this problem can be avoided by considering *improved* sum rules [14, 15], which only receive contributions from a finite number of EFT coefficients.

---

[12]If we denote the three sum rules in the order they appear as A, B, C, the relevant linear combinations are respectively: $6A - B + C \geq 0$ and $90A + 11B + \frac{1}{2}C \geq 0$.

In order to get a bound, the idea is to look for functionals whose action on any state above the cutoff is nonnegative:

$$\text{if} \quad \mathcal{F}[J, m^2] = \int_0^M dp \sum_i \psi_i(p) \mathcal{C}_{i,-p^2}[J, m^2] \geq 0, \quad \forall\, m \geq M, J \in \mathbb{Z},$$

$$\text{then} \quad -\int_0^M dp \sum_i \psi_i(p) B_i(p^2)\Big|_{\text{low}} \geq 0, \tag{3.2}$$

where $\mathcal{C}_{i,-p^2}[J, m^2]$ are the heavy densities defined in eq. (2.33). In words, any nonnegative functional yields an inequality on low-energy observables, thanks to the relation between low and high energies (2.27).

In $D = 4$ there will be a tension between finiteness and positivity: finiteness on the graviton pole requires $\psi_i(p)$ to vanish faster than p at the origin, which is impossible for the Fourier transform of a positive function. In practice, we begin by finding functionals which are rigorously positive at all impact parameters but logarithmically diverge on the graviton pole. We then regulate by adding an infrared cutoff $m_{\text{IR}} \ll M$, and accept that this causes negativity at large impact parameters.

Before explaining how we produce positive functionals, we first detail how we ascertain positivity.

3.2 Example positive functionals involving gravity

We now present explicit bounds that combine two ingredients:

- The spin-2,3 sum rules $B_2^{(1)}(p), B_3^{(1)}(p)$, smeared against suitable wavepackets $\psi(p)$

- The forward limit of spin-4 sum rules and its first derivative: $B_4^{(1)}(0)$, $\partial_{p^2} B_4^{(1)}(0)$.

These use only $B^{(1)}$ (2.25): fixed-u dispersion relation for the MHV amplitude.

In principle the spin-4 sum rules should also be smeared to make them rigorously valid, but as discussed in section 2.3 this is a technical modification, which we will ignore here.

Using these ingredients and the method of section 2.4, we have constructed the following two functionals:

$$\mathcal{F}_{g3} = \int_{m_{\text{IR}}}^M dp \, (1-p)^3 \big(p(65 + 155p + 47p^2) B_2^{(1)} + 2(81 - 1927p + 4601p^2) B_3^{(1)} \big)$$
$$- 5\partial_{p^2}(B_4^{(1)} + B_4^{(2)})|_{p=0}, \tag{3.3a}$$

$$\mathcal{F}_{g4} = \int_{m_{\text{IR}}}^M dp(1-p)^3 \big(p(99 + 282p + 49p^2) B_2^{(1)} + (21 - 474p + 1216p^2) B_3^{(1)} \big)$$
$$- \big(7B_4^{(1)} + \partial_{p^2} B_4^{(1)} + \partial_{p^2} B_4^{(2)} \big)|_{p=0}, \tag{3.3b}$$

in units where $M = 1$. Notice the lower cutoff m_{IR} makes the functionals infrared-safe.

We claim that:

- Without the cutoff m_{IR}, the functionals are positive for all states with $m > M$

- With the cutoff m_{IR}, they only become negative at some large $b \sim m_{\text{IR}}^{-2/3}$

- They imply the respective bounds:

$$|\widehat{g}_3|^2 M^8 \leq 37.8 \log(M/m_{\mathrm{IR}}) - 45.4 - \mathcal{F}_{g_3}\big|_{\mathrm{matter}}, \tag{3.4a}$$

$$g_4 M^6 \leq 8\pi G \left(14.1 \log(M/m_{\mathrm{IR}}) - 15.5 - 0.04|\widehat{g}_3|^2\right) - \mathcal{F}_{g_4}\big|_{\mathrm{matter}}, \tag{3.4b}$$

where the matter contributions (coming from possible light scalars or Kaluza-Klein modes and detailed in eq. (3.8) below) is sign definite: $\mathcal{F}\big|_{\mathrm{matter}} \geq 0$.

The bound (3.4a) is a sharp version of the CEMZ constraint [7]: $\widehat{g}_3 \lesssim \frac{1}{M^4}$, up to logarithms. This shows that a cubic coupling of size $\frac{1}{M^4} R^3$ cannot be turned on without having a heavy state at the mass M or lighter. The bound (3.4b) is similar for quartic couplings.

The functionals (3.3) are not optimal: their main virtue is to be explicit enough to be analyzed in full detail in this section. They establish our main conceptual result: higher dimensional couplings can be bounded in terms of Einstein gravity. Optimal bounds are presented in the next section, see eq. (4.4).

How do we ascertain that a functional is positive on all heavy states? A simple strategy is to plot the action $\mathcal{F}[m^2, J]$ as a function of m for various discrete spins $J = 0, 2, 4, 5, 6, \ldots$. Note that the heavy contribution (2.32) is the sum of two positive unknowns ($|c_{J,m^2}^{++}|^2$ and $|c_{J,m^2}^{+-}|^2$) and we must ascertain that the coefficient of each is positive: we will refer to those as $\mathcal{F}^{++}$ and $\mathcal{F}^{+-}$ below. Thus two plots must be made for each (even) value of the spin, and one plot for each odd spin.

Since we cannot plot infinitely many spins, it is fruitful to exploit regularity of the functionals. We find that when plotted as a function of "impact parameter" $b = 2J/m$, and m, the curves vary smoothly with spin and display simple asymptotic trends. This allows us to draw contour plots where positivity is easy to ascertain, and potentially dangerous regions can be easily identified for finer sampling.

We display $\mathcal{F}_{g_3}$ in figure 3. Unless noted otherwise, all plots in this section are in units where $M = 1$. To make the plot, we considered data up to $J_{\max} = 300$, more specifically, $J = 0, 2, \ldots, 300$ for $\mathcal{F}_{g_3}^{++}$ and $J = 4, 5, \ldots, 300$ for $\mathcal{F}_{g_3}^{+-}$, accounting for spin selection rules. For each spin, we sample the interval $m \in [1, 16]$, with a larger density of points closer to the origin (for example 80 points between 1 and 1.2 and 700 between 1.2 and 16). At the highest value $m = 16$ we have safely reached the $m \to \infty$ limit, further discussed below, and the value $b = 40$ is well past any interesting structure. Therefore figure 3 establishes positivity of $\mathcal{F}_{g_3}$. Similar plots are displayed for $\mathcal{F}_{g_4}$ in figure 4, where we used the same sampling.

In addition to sampling a finite range, we find it useful to verify positivity in a $m \to \infty$ scaling limit. The limit is nontrivial if keeping impact parameter $b = 2J/m$ is fixed, and is dominate by the B_2 component of the sum rules. It is essentially the Fourier transform of its coefficient [15]:

$$\lim_{m \to \infty} \mathcal{F}[m^2, \tfrac{bm}{2}] = \frac{2}{m^4} \int_0^1 dp\, \psi(p) \times \left(\left(|c_{J,m^2}^{++}|^2 + |c_{J,m^2}^{+-}|^2\right) J_0(pb) + \ldots\right), \tag{3.5}$$

where $\ldots$ represents higher orders in $1/m$. With $m_{\mathrm{IR}} = 0$, positivity is easy to ascertain at all b, as shown for $\mathcal{F}_{g_3}$ in figure 5. (In this limit, $\mathcal{F}^{++}$ and $\mathcal{F}^{+-}$ coincide, so there is only one curve.) However, with $m_{\mathrm{IR}} = 0$ the action on low-energy gravity is infinite, and

(a) action of $\mathcal{F}_{g_3}$ on $++$ states.

(b) action of $\mathcal{F}_{g_3}$ on $+-$ states.
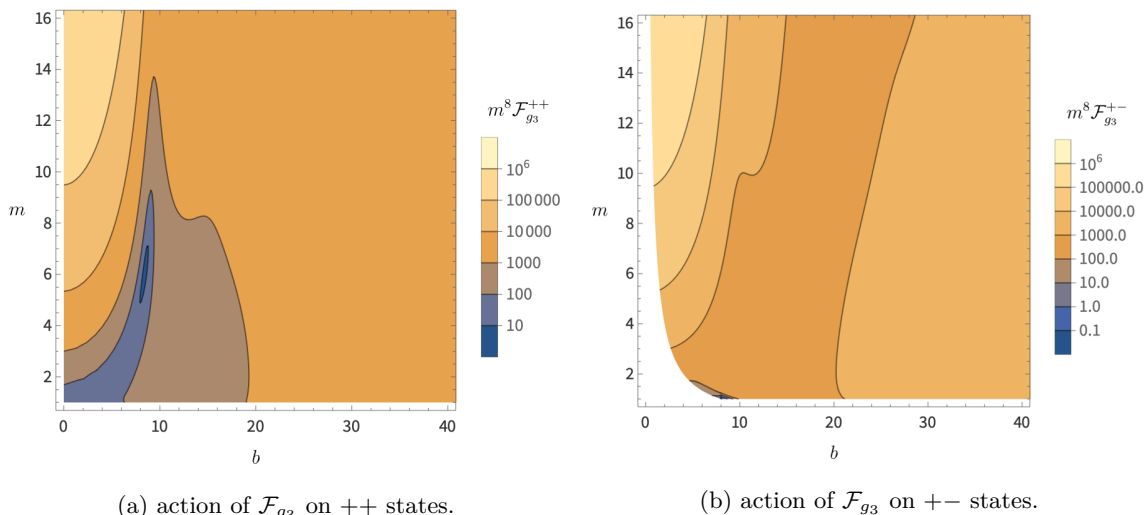
Figure 3. Contour plots which confirm non-negativity of the functional $\mathcal{F}_{g_3}$ giving the upper bound (3.4a), in units where $M = 1$. We scaled the functional by m^{10} to make scaling limits more manifest. The sampling points are detailed in the text. The lower-left corner is blank in the second figure due to the selection rule $J \geq 4$.

(a) action of $\mathcal{F}_{g_4}$ on $++$ states.

(b) action of $\mathcal{F}_{g_4}$ on $+-$ states.
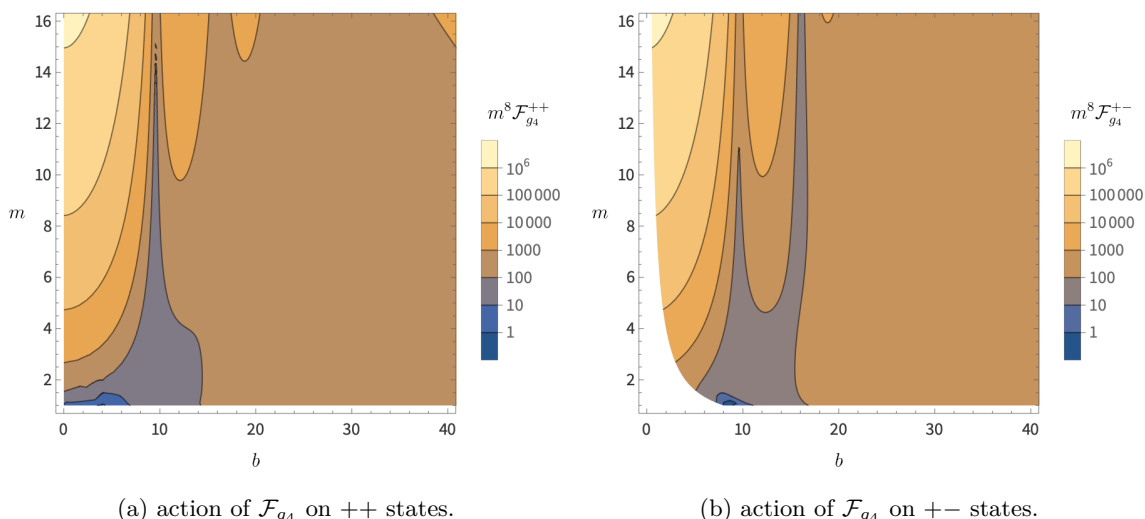
Figure 4. Similar to figure 3: contour plots which confirm non-negativity of $\mathcal{F}_{g_4}$, establishing the bound (3.4b).

the resulting upper bound is vacuous. We thus add a cutoff $\theta(p > m_{\mathrm{IR}})$, which creates negativity (we take $\mathcal{F}_{g_3}$ as example to illustrate):

$$\lim_{b \to \infty} \lim_{m \to \infty} \mathcal{F}_{g_3}[m^2, \tfrac{bm}{2}] \approx \frac{160}{m^4 b^3} - 260 \frac{m_{\mathrm{IR}}^2}{m^4} \frac{J(m_{\mathrm{IR}} b)}{m_{\mathrm{IR}} b} . \tag{3.6}$$

The second term overwhelms the first at $b_{\max} \sim m_{\mathrm{IR}}^{-2/3}$ and creates negative plateau up til $b \sim m_{\mathrm{IR}}^{-1}$, where the Bessel function gets damped. While the total area under the functional is necessarily 0 (because of vanishing at $p = 0$), the behavior past m_{IR}^{-1} is not universal and could be modified by using a smoother cutoff. This is depicted in figure 5, where we contrast the functionals with $m_{\mathrm{IR}} = 0$ and $m_{\mathrm{IR}}/M = 10^{-6}$.
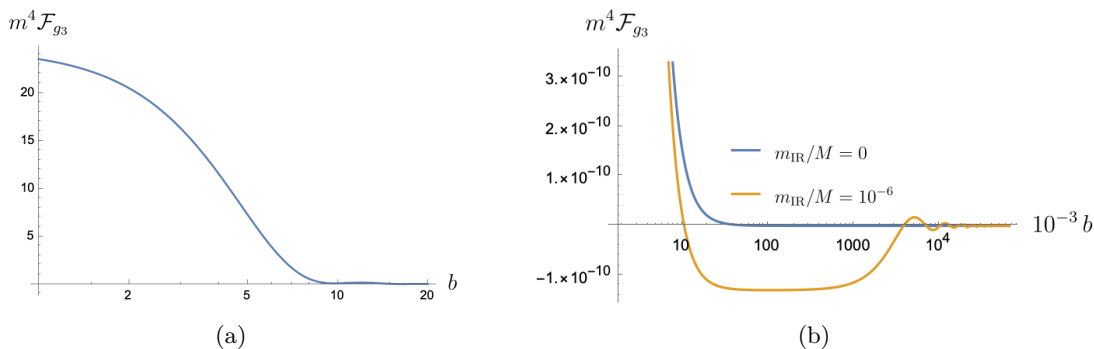
Figure 5. Plots of functionals $\mathcal{F}_{g_3}$ at the scaling limit $J, m \to \infty$ where $b = 2J/m$. (a) confirms the positivity at finite b where the cutoff m_{IR} makes no difference. (b) displays the large b behavior for both $m_{\mathrm{IR}} = 0$ (blue curve) and $m_{\mathrm{IR}}/M = 10^{-6}$ (orange curve). There is a negative plateau in the range $m_{\mathrm{IR}}^{-2/3} \lesssim b \lesssim m_{\mathrm{IR}}^{-1}$ as detailed in the text.

Note that the functionals $\mathcal{F}_{g_3}$ and $\mathcal{F}_{g_4}$ include derivatives of the spin-4 sum rule $B_4^{(1)}$ around the forward limit. For the purposes of bounding $|\widehat{g}_3|$, one can also find functionals that are pure linear combinations of $B_2^{(1)}$ and $B_3^{(1)}$, with no forward limit component.[13] Such functionals may have useful technical applications, since they avoid the subtleties with forward limits discussed in section 2.5.

3.3 Light spin-0 and spin-2 matter fields don't lower the cutoff

In addition to the graviton, we allow for the presence of possible spin-0 and spin-2 light states (for example, Kaluza-Klein modes), whose amplitudes are given by eq. (B.1). They contribute to the low-energy part of sum rules (integral over the arc at $s \sim M^2$):

$$B_2^{(1)}(p^2)|_{\mathrm{matter}} = - \sum_{m_\ell < M} |g_0(m_\ell)|^2 (p^2 - 2m_\ell^2)$$
$$- \sum_{m_\ell < M} \frac{|g_2(m_\ell)|^2}{m_\ell^4} (p^2 - 2m_\ell^2)(m_\ell^2 - 6m_\ell p^2 + 6p^4), \tag{3.7a}$$

$$B_3^{(1)}(p^2)|_{\mathrm{matter}} = \sum_{m_\ell < M} |g_0(m_\ell)|^2 + \sum_{m_\ell < M} \frac{|g_2(m_\ell)|^2}{m_\ell^4}(m_\ell^2 - 6m_\ell p^2 + 6p^4), \tag{3.7b}$$

$$B_4^{(1)}(p^2)|_{\mathrm{matter}} = - \sum_{m_\ell < M} \frac{|g_2(m_\ell)|^2}{m_\ell^4} 12\, p^2. \tag{3.7c}$$

The spin-2 contribution with $m_\ell \to 0$ is proportional to that from the cubic coupling $|\widehat{g}_3|^2$, see (2.30).

---

[13] One example is

$$(1 - p)^2(-154.020p^7 + 2006.03p^6 - 3443.55p^5 - 2322.66p^4 + 5294.87p^3 - 2194.19p^2 + 798.968p)B_2^{(1)}$$
$$+ (3701.82p^6 - 18251.2p^5 + 32684.9p^4 - 25581.3p^3 + 7603.72p^2 - 157.911p + 0.00171)B_3^{(1)},$$

which proves the bound $|\widehat{g}_3|^2 \leq 1598 \log(M/m_{\mathrm{IR}}) - 3211$. This could certainly be improved by including more powers of p.
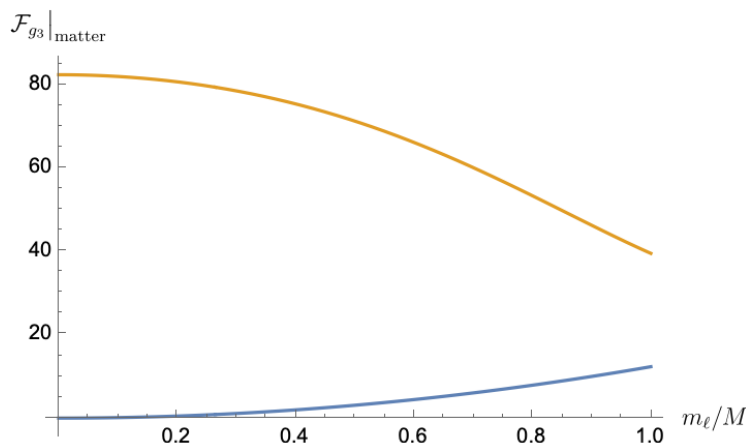
Figure 6. Positivity of functional $\mathcal{F}_{g_3}\big|_{\text{matter}}$ for acting on light matter fields (3.9). We separate out the scalar and spin-2 contributions. The blue line is coefficient of $|g_0(m_\ell)|^2$ and the orange line is coefficient of $|g_2(m_\ell)|^2/m_\ell^4$ in $\mathcal{F}_{g_3}\big|_{\text{matter}}$.

The key feature is that the unknown couplings $|g_J(m_\ell)|^2$ are sign-definite. Therefore, if we take combination of sum rules such that the coefficient of each unknown is positive for $m_\ell \in (0, M)$, then the unknown light and heavy states will all contribute with the same sign leading again to a valid inequality on EFT parameters. For the functionals in eq. (3.3), for example, we get:

$$\mathcal{F}_{g_3}\big|_{\text{matter}} = \sum_{m_\ell < M} |g_0(m_\ell)|^2 \left(\frac{2591}{210} m_\ell^2 + \frac{1}{18} \right) \tag{3.8}$$
$$+ \sum_{m_\ell < M} \frac{|g_2(m_\ell)|^2}{m_\ell^4} \left(\frac{2591}{210} m_\ell^6 - \frac{239}{18} m_\ell^4 - \frac{64573}{1540} m_\ell^2 + \frac{330151}{4004} \right) \, .$$

in units where $M = 1$. These polynomials are positive-definite as demonstrated in figure 6.

Alternatively, we could move the matter contribution in eq. (3.4a) to the left-hand-side, giving:

$$|\widehat{g}_3|^2 + \mathcal{F}_{g_3}\big|_{\text{matter}} \leq \frac{1}{M^8} \left(37.8 \log(M/m_{\text{IR}}) - 45.4 \right) \, . \tag{3.9}$$

We conclude that our bounds on $\widehat{g}_3$ limit the *sum* of squared cubic couplings to *all* light particles below the higher-spin scale M.

It is remarkable that light scalars or spin-2 particles cannot couple strongly to two gravitons. This is very different than for the scalar EFT studied in [14], where the only limit on the interaction strength of light scalars would be the unitarity bound Im $a_J(s) \leq 2$.

This result can be interpreted as follows. In Einstein's gravity, the decay rate of a Kaluza-Klein graviton to two massless gravitons is proportional to an overlap integral $\int dy \sqrt{g} \chi(y) = 0$, which vanishes by orthogonality of eigenfunctions (here y is a coordinate on the internal manifold and $\chi(y)$ is the eigenfunction corresponding to the mode in question). Thus Kaluza-Klein modes only decay through higher-derivative corrections from the higher-dimensional perspective. The $1/M^8$ suppression in (3.9) confirms that such a

suppression exists for *any* massive spin-two particle, irrespective of its microscopic origin. The scale of suppression is controlled the mass of *higher-spin* particles.

3.4 Systematic strategy: improved sum rules

In general, a low-energy gravitational EFT includes an infinite number of contact terms. When evaluated at low-energies, the spin-2 and spin-3 sum rules $B_2(t)$ and $B_3(t)$ are each sensitive to an infinite subset of these contact terms. It is useful to subtract from $B_2(t)$ a linear combination of forward-limits of higher spin sum rules to define "improved" sum rules [15], which are sensitive only to a finite number of higher-derivative corrections.

Ideally, one would like to use only smeared amplitudes in order to construct improved sum rules. In practice, following [15], we instead take a more pragmatic approach, mentioned below (2.22): we use smearing for spin-2 sum rules, where it is strictly necessary due to the graviton pole, and for spin-3. However we allow ourselves to use forward-limit expansions for all higher-subtracted sum rules. This will greatly simplify calculations.

Not all sum rules need to be improved, for example, the $B_{2,3}^{(1)}$ and $B_2^{(3)}$ sum rules are automatically free of higher-dimension contact terms. For the B_2 sum rules, we only need to define

$$B_2^{(2)\,\text{imp}}(p^2) = B_2^{(2)}(p) + \oint_\infty \frac{ds}{4\pi i} \left[-\frac{\partial_t f(p^2 - s, 0)}{s(s+p^2)} + \frac{\partial_t f(0,s)}{s(s-p^2)} - \frac{2s - p^2}{s^2(s-p^2)^2} f(0,s) \right],$$
(3.10a)

$$B_2^{(4)\,\text{imp}}(p^2) = B_2^{(4)}(p) + \oint_\infty \frac{ds}{4\pi i} \left[\frac{p^4(4s - 3p^2)}{s^4(s-p^2)^2} h(s,-s) + \frac{2p^6}{s^3(s^2-p^4)} \partial_t h(s,-s) \right].$$
(3.10b)

The improved spin-2 sum rules have the low-energy contributions

$$-B_2^{(2)\,\text{imp}}(p^2)|_{\text{low, grav}} = \frac{2\pi G}{p^2}|\widehat{g}_3|^2, \qquad -B_2^{(4)\,\text{imp}}(p^2)|_{\text{low, grav}} = -40\pi\widehat{g}_3 p^2, \qquad (3.11)$$

where we have written only the contribution of graviton exchange for brevity. Similarly we can also define improved versions of higher-spin sum rules. An example is:

$$B_4^{(1)\,\text{imp}}(p^2) = B_4^{(1)}(p^2) + \oint \frac{ds}{4\pi i} \left[\frac{p^2}{s(s-p^2)} f(0,-s) - \frac{2p^{12}}{s^5(s^2-p^4)} f(s,-s) \right.$$

$$+ \frac{p^2(s^4 - 2s^3 p^2 - 2s^2 p^4 - 2s p^6 - 2p^8)}{s^5(s+p^2)} f(0,s) + \frac{p^4(s^2 + 2p^4)}{s^4} \partial_t f(0,s)$$

$$\left. - \frac{p^4(s^2 - p^4)}{s^4} \partial_t f(0,-s) - \frac{p^8}{2s^3} \left(\partial_t^2 f(0,-s) + \partial_t^2 f(0,-s) \right) \right], \qquad (3.12)$$

which nicely gives

$$-B_4^{(1)\,\text{imp}}(p^2)|_{\text{low, grav}} = 4\pi G|\widehat{g}_3|^2 + 2g_4. \qquad (3.13)$$

In this section, we apply these improved sum rules (together with some forward-limit sum rules) to derive bounds involving higher-derivative Wilson coefficients and gravity,

using the parameter choices listed in appendix C. Following [15], we consider wavefunctions $\psi_i(p)$ that are polynomials in p

$$\psi_i(p) = \sum_{n=n_{i\min}} c_{\text{in}} p^n \,. \tag{3.14}$$

Our wavefunctions start with specific exponents $n_{i\min}$ that depend on the sum rule. Specifically, we require that all sum rules possess the same large-b behavior in the scaling limit $m \to \infty$ with $b = \frac{2J}{m}$ fixed and large. (One can prove that this is a necessary condition for obtaining a positive functional.) Precisely, the minimal values of n for the improved spin-2 sum rules are

$$(n_{2\min}^{(1)}, n_{2\min}^{(2)}, n_{2\min}^{(3)}, n_{2\min}^{(4)}) = (1, 10, 6, 1) \,. \tag{3.15}$$

Since higher-spin sum rules are subleading at large-m (with fixed b), their wavefunctions can safely start with $n_{\min} = 0$, as long as spin-2 sum rules are included in the functional.

It is worth noting that compact support in p can lead to oscillations at large b, which can potentially hinder positivity. To suppress oscillations, it is useful to make the wavefunctions smoother near $p = 1$ (in units where $M = 1$). We do this by multiplying the wave functions by a power of $(1-p)$. For example, our example functionals in the previous section included factors of $(1-p)^3$ to suppress large-b oscillations. In this section, we consider wavefunctions of the form

$$\psi_i(p) = \sum_{n=n_{i\min}}^{n_{i\min}+n_{\max}} c_{\text{in}}(1-p)^2 p^n \,, \tag{3.16}$$

where we truncate n to $n_{\max}$. Larger values of $n_{\max}$ will correspond to more complicated functionals and stronger bounds.

As mentioned before, to allow for light matter particles, we add additional constraints to the solver SDPB [68, 69] that impose negativity of matter contributions to low-energy sum rules for $m_\ell < M$. In practice, this means including positive matrices that are negatives of the matrices paired with low-energy three-point couplings $g_s(m_\ell), g_s^*(m_\ell)$, and $g_2(m_\ell), g_2^*(m_\ell)$ for $m_\ell < M$. In the MHV sum rules, matrices for matter particles are simply the coefficients of $|g_s(m_\ell)|^2, |g_2(m_\ell)|^2$. In numerics, we discretize m_ℓ to the following values: $m_\ell \in (0, 0.01, \ldots, 1)M$.

3.5 Relation with the CEMZ argument

In [7], CEMZ considered classical scattering of gravitons with various polarizations against a target (for example, a black hole) and argued that a too-large value of $|\hat{g}_3|M^4/G$ would lead to a time advance for one of the polarizations. It is instructive to see how the CEMZ argument is paralled in our formalism.

The basic idea is to consider the B sum rules at large impact parameters $bM \gg 1$. Momentarily ignoring the compact-support constraint $p < M$, we can define impact-parameter sum rules by Fourier transforming in the transverse space

$$\widehat{B}_2^{(1)}(b) = \int \frac{d^2 p}{(2\pi)^2} e^{ip \cdot b} B_2^{(1)}(p) = \frac{1}{2\pi} \int_0^\infty dp \, p J_0(bp) B_2^{(1)}(p) \,. \tag{3.17}$$
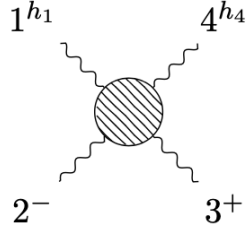
Figure 7. Scattering of gravitons with arbitrary helicity $1^{h_1} \to 4^{-h_4}$ against a $2 \to 3$ "target", as used in the CEMZ-like argument (3.19). Time runs from left to right.

At large b, the integral is highly oscillatory, which suppresses the heavy action (2.32) on states with small J. Acting on states with $J \gg 1$, the $\tilde{d}$ function simplifies according to (A.11), and the integral localizes[14] to

$$\widehat{B}_2^{(1)}(b)\big|_{\text{high}} \to \frac{1}{\pi} \left\langle \left(|c_{J,m^2}^{++}|^2 + |c_{J,m^2}^{+-}|^2 \right) \frac{\delta\left(\frac{2J}{m} - b \right)}{bm^4} \right\rangle . \tag{3.18}$$

This is positive, yielding a constraint on low-energy coefficients $-\widehat{B}_2^{(1)}(b)\big|_{\text{low}}$.

More generally, we can consider scattering of an arbitrary helicity graviton (1 and 4) against a positive-helicity target (2 and 3), see figure 7, which motivates us to consider a matrix of sum rules:

$$\mathbb{B}(p^2) = \oint_{s=\infty} \frac{ds}{s^3} \begin{pmatrix} \mathcal{M}(1^+2^-3^+4^-) & \mathcal{M}(1^+2^-3^+4^+) \\ \mathcal{M}(1^-2^-3^+4^-) & \mathcal{M}(1^-2^-3^+4^+) \end{pmatrix} (s, -p^2) = 0 , \tag{3.19}$$

evaluated in the forward limit. In the center of mass frame, we have

$$\begin{pmatrix} \mathcal{M}(1^+2^-3^+4^-) & \mathcal{M}(1^+2^-3^+4^+) \\ \mathcal{M}(1^-2^-3^+4^-) & \mathcal{M}(1^-2^-3^+4^+) \end{pmatrix} \sim s^4 \begin{pmatrix} f(s,t) & \widehat{p}^4 g(s,t) \\ \widehat{p}^{*4} \overline{g}(s,t) & f(s,t) \end{pmatrix} , \tag{3.20}$$

where we have defined $\widehat{p} = p_1 + ip_2$, where p_i are the components of transverse momentum transfer. Thus, we recognize the matrix elements of eq. (3.19) as our forward $B_2^{(1)}$, $B_2^{(3)}$ sum rules

$$\mathbb{B}(p^2) = \begin{pmatrix} \frac{1}{2} B_2^{(1)}(p^2) & \widehat{p}^4 B_2^{(3)}(p^2) \\ \widehat{p}^{*4} B_2^{(3)}(p^2)^* & \frac{1}{2} B_2^{(1)}(p^2) \end{pmatrix} . \tag{3.21}$$

Let us transform to b space using (3.18) for the diagonal terms. For the off-diagonal terms, we use

$$\widehat{B}_2^{(3)}(b) = \int \frac{d^2 p}{(2\pi)^2} \widehat{p}^4 e^{ip \cdot b} B_2^{(3)}(p) = \frac{\beta^2}{2\pi} \int_0^\infty dp \, p^5 J_4(bp) B_2^{(3)}(p) , \tag{3.22}$$

---

[14]A way to see this is using the identity $\int_0^\infty dp \, p J_\nu(xp) J_\nu(yp) = \delta(x-y)/x$ for $\nu > -1/2$.

where $\beta = \widehat{b}/\widehat{b}^*$ with $\widehat{b} = b_1 + b_2$. We find the following heavy contribution:

$$\mathbb{B}(b)|_{\text{high}} \propto \left\langle \left[\begin{pmatrix} c^{++}_{J,m^2}\beta^* \\ c^{-+}_{J,m^2}\beta \end{pmatrix} \begin{pmatrix} c^{++}_{J,m^2}\beta^* & c^{-+}_{J,m^2}\beta \end{pmatrix}^* + \begin{pmatrix} c^{+-}_{J,m^2}\beta^* \\ c^{--}_{J,m^2}\beta \end{pmatrix} \begin{pmatrix} c^{+-}_{J,m^2}\beta^* & c^{--}_{J,m^2}\beta \end{pmatrix}^* \right] \frac{\delta\left(\frac{2J}{m}-b\right)}{bm^4} \right\rangle, \tag{3.23}$$

which is positive-definite, since the bracket is (by construction) a positive-definite matrix. On the other hand, evaluating the low-energy contribution in eqs. (2.7) we find

$$-\mathbb{B}(b)|_{\text{low}} = 4G \begin{pmatrix} \log[1/(bm_{\text{IR}})] & \frac{24\widehat{g}_3}{\widehat{b}^4} \\ \frac{24\widehat{g}_3^*}{\widehat{b}^{*4}} & \log[1/(bm_{\text{IR}})] \end{pmatrix}. \tag{3.24}$$

The eigenvalues of this matrix precisely reproduce the eikonal time delays given in eq. (3.22) of [7] (with $\gamma^{\text{there}} = 8\widehat{g}_3^{\text{here}}$). It is instructive to see how these calculations relate. In that reference the eikonal phase is computed from the Fourier transform of the non-analytic parts of the amplitude

$$\delta(s,b) = \frac{1}{2s} \int \frac{d^2p}{(2\pi)^2} e^{ip\cdot b} \begin{pmatrix} \frac{8\pi Gs^2}{p^2} & \frac{4\pi Gs^2\widehat{p}^4}{p^2} \\ \frac{4\pi Gs^2\widehat{p}^{*4}}{p^2} & \frac{8\pi Gs^2}{p2} \end{pmatrix}. \tag{3.25}$$

A classical time delay (matrix) is then extracted from the energy derivative: $t = \frac{\partial}{\partial E}\delta(E^2, b)$. We see that the energy-growing part of the time delay is precisely what our matrix of sum rules captures:

$$t = E\,\mathbb{B}(b)|_{\text{low}}. \tag{3.26}$$

Now the dispersive sum rules state that (3.23) equals (3.24). In particular, they imply that the energy-growing part of time delays (in the linear regime) must be positive.

The argument of ref. [7] is concluded by noting that the low-energy calculation (using graviton exchange) is only valid then $b \gg M^{-1}$ where M is the mass of heavy states. Thus we should only impose positivity of (3.24) in that range, which gives the parametric bound:

$$|\widehat{g}_3| \lesssim \frac{1}{M^4} \log[1/(bm_{\text{IR}})]. \tag{3.27}$$

Ref. [7] further argued that an infinite tower of higher-spin states needed to appear.

This discussion highlights that CEMZ constraints are *built into* dispersive sum rules, they are a subset of the functionals enumerated in the preceding subsection. Namely, this subset consists of spin-2 functionals $B_2(p)$ integrated against wavepackets that are peaked at impact parameters b. The CEMZ requirement $b \gg 1/M$ then has a clear origin in our compact support property $p < M$ (see section 2.3).

Despite the fact that the same mathematics appear, the physical assumptions are quite distinct. Ref. [7] considered very large center of mass energy, where the amplitude exponentiates, whereas we consider $Gs \ll 1$ where the tree-level approximation is sufficient. (Appendix D of [7] also presented an argument valid in the linear regime and related to

spin	$8\pi G\widehat{g}_3 M^2$	$g_4 M^4$	$g_5 M^6$	$g_6 M^8$	$g_6' M^8$
0	$+\frac{1\times 2}{315}$	$\frac{16}{1575}$	$\frac{64}{10395}$	$\frac{192}{35035}$	$\frac{64}{135135}$
$\frac{1}{2}$	$-\frac{2\times 2}{315}$	$\frac{58}{1575}$	$\frac{944}{51975}$	$\frac{32}{1911}$	$\frac{1856}{675675}$
1	$+\frac{3\times 2}{315}$	$\frac{248}{1575}$	$\frac{848}{17325}$	$\frac{258304}{4729725}$	$\frac{13856}{675675}$
$\frac{3}{2}$	$-\frac{4\times 2}{315}$	$\frac{1676}{1575}$	$\frac{160}{2079}$	$\frac{434272}{1576575}$	$\frac{32192}{135135}$
2	$+\frac{5\times 2}{315}$	$\frac{5368}{315}$	$-\frac{31888}{10395}$	$\frac{6024448}{1576575}$	$\frac{800288}{135135}$

Table 1. One-loop contributions to low-energy parameters from heavy particles of various spins, divided by an overall factor $G^2 N$ where N is the number of particles of the given spin circulating in the loop. Extracted from [48].

the "chaos bound" [55].) The upside of using large center of mass energies there was that the acausality becomes classical and macroscopic, which led to transparent "grandparent paradoxes". The downside is that the cutoff is imprecise, $b \ll 1/M$: such a method gives parametric bounds, in contrast with the precise cutoff $p < M$ which yields sharp bounds.

In addition, the fact that we scatter waves rather than particles enable the precise energy resolution that is needed to probe the mass and couplings of states above the cutoff. One might say that time advances constitute a classical statement of causality, while crossing symmetry and analyticity of the S-matrix provide a quantum version.

4 Results

In this section we will present our results for bounds on EFT modifications of Einstein gravity.

4.1 Comparison with model amplitudes

Following ref. [48] it will be instructive to compare our bounds with explicit models, with higher-dimension operators arising from integrating out a loop of massive particles of mass m and spin two or less. For this one loop process, the mass of the lightest higher-spin states, which are two-particle states, is $M = 2m$. The resulting Wilson coefficients are shown in table 1.

Note that the complex coupling $\widehat{g}_3$ (which enter non-MHV amplitudes) vanishes for supersymmetric spectra; correspondingly, the first column in the table is proportional to the numbers of degrees of freedom and a fermionic sign. Other complex couplings follow the same pattern and it thus suffices to record the contribution from a heavy spin-0 particle:

$$\left\{\widehat{g}_4 M^2, \ \widehat{g}_5 M^4, \ \widehat{g}_6 M^8, \ \widehat{g}_6' M^8, \ \widehat{g}_6'' M^8\right\}_{\text{spin } 0} = G^2 N \left\{\frac{8}{945}, \frac{4}{495}, \frac{4}{3003}, \frac{3488}{225225}, \frac{512}{1576575}\right\}. \tag{4.1}$$

It is important to note that since we neglect light graviton loops, the effect of integrating out a particle are suppressed by a power of $GM^2 \ll 1$ compared with the effects we focus on, and are thus beyond the accuracy of our bounds. Nonetheless, the comparison with

these models becomes physically meaningful (larger than neglected graviton loops) if a large number N of heavy particles circulate. In this sense, the upper bounds on couplings recorded below in figure 9 can be interpreted as bounds on the number of species above mass M: $N \lesssim \# \frac{M_{\text{pl}}^2}{M^2} \log(m/m_{\text{IR}})$, as far as their effects on low energies are concerned. This has the expected dependence on M_{pl} from "species bounds," although our bounds have an extra factor of an infrared logarithmic due to the nature of the probes we are using.

In addition, following [48], we will consider three models of string theory:

$$f = \frac{8\pi G}{stu} \frac{\Gamma\left(1 - \frac{s}{M^2}\right)\Gamma\left(1 - \frac{t}{M^2}\right)\Gamma\left(1 - \frac{u}{M^2}\right)}{\Gamma\left(1 + \frac{s}{M^2}\right)\Gamma\left(1 + \frac{t}{M^2}\right)\Gamma\left(1 + \frac{u}{M^2}\right)} \times X^p \,, \tag{4.2}$$

where $X = 1 - \frac{su}{M^2(t+M^2)}$ and $p = 0, 1, 2$ representing respectively the superstring theory, the heterotic string, and the bosonic string. In all cases M coincides with the mass of the first spin-4 particle exchanged between gravitons.

For the string models, we extract low-energy parameters by matching with the low-energy expansion (2.7) and subtracting the low-energy poles. For the bosonic string, we additionally subtract a tachyon pole $\frac{1}{t+M^2}$. (Because of the tachyon, the bosonic string model is not fully "physical", however after subtracting this pole it still has a positive heavy spectral density.) There is an ambiguity of whether we include the spin-0 and spin-2 contributions to the $t = M^2$ pole as part of f_{matter} or as part of the "heavy" contributions: our bounds are valid in either case. In the plots below the string models thus span extended regions, depending on what fraction $0 \leq \gamma_i \leq 1$ of each pole we choose to subtract. For example, for the heterotic string we find

$$g_4|_{\text{hs}} = 8\pi G\left(1 + 2\zeta_3 - \frac{11}{12}\gamma_0 - \frac{1}{12}\gamma_2\right), \qquad g_5|_{\text{hs}} = -8\pi G\left(1 + \frac{11}{12}\gamma_0 + \frac{1}{12}\gamma_2\right). \tag{4.3}$$

4.2 Bounds involving R^3, R^4 and gravity

In the preceding section, we provided example functionals giving upper bounds (3.4) on $|\widehat{g}_3|^2$ and g_4 in terms of gravity. As we emphasized, these bounds were not optimal, and we get stronger ones by using the numerical parameter choices in appendix C. Our optimal bounds are:

$$|\widehat{g}_3|^2 M^8 \leq 24.9 \log(M/m_{\text{IR}}) - 27.6 \,, \tag{4.4}$$

$$\frac{g_4 M^6}{8\pi G} \leq 12.3 \log(M/m_{\text{IR}}) - 13.5 \,. \tag{4.5}$$

To obtain these, we included all improved sum rules B_2^{imp} and B_3^{imp} with $n_{\text{max}} = 6$, and we included additional $\partial_{p^2}^q B_4^{(1)\,\text{imp}}(0)$ up to $q = 2$ to get the bound on g_4.

A finer way to present the constraint is to carve out the allowed space in the three EFT parameters $|\widehat{g}_3|^2$, g_4 and G, as shown in figure 8. These were are computed by using all improved B_2 and B_3 for $n_{\text{max}} = 5$ and additional forward-limit contributions from $\partial_{p^2}^q B_4^{(1)\,\text{imp}}(0)$ up to $q = 2$.

A special limit of the bound is the dashed line in figure 8 which is tangent to the allowed region near origin; from its slope we find numerically that

$$\frac{g_4}{8\pi G} \geq 0.26 |\widehat{g}_3|^2 M^2 \,. \tag{4.6}$$
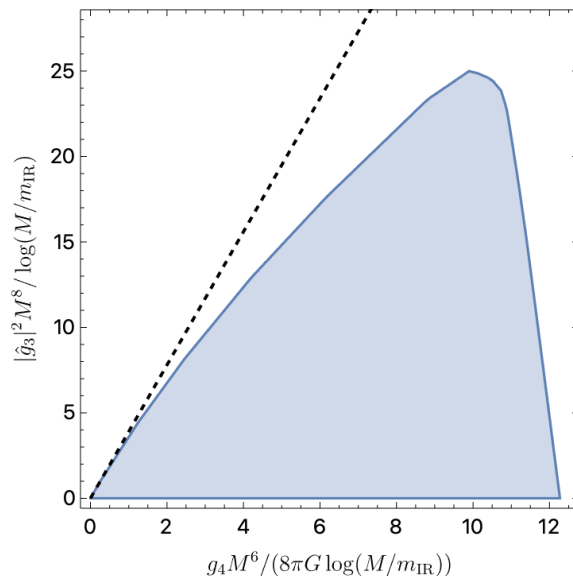
Figure 8. Allowed region for $|\widehat{g}_3|^2$ and g_4 in terms of Newton's constant and the spin-4 mass gap M. Note that both axes are rescaled by an infrared logarithm $\log(M/m_{\mathrm{IR}})$. Manifestly, both $|\widehat{g}_3|^2$ and g_4 obey two-sided bounds; a nonvanishing cubic coupling $\widehat{g}_3$ requires a nonvanishing quartic g_4. The dashed line gives the bound eq. (4.6).

This is effectively equivalent to the bound $\frac{g_4}{8\pi G} \geq \frac{1}{4}|\widehat{g}_3|^2 M^2$ reported in (6.13) of [48] using forward-limit bounds of spin $k \geq 4$. This bound indicates that it is not possible to turn on a cubic coupling without having a quartic coupling as well.

It is instructive to see the impact of allowing light spin-0 and spin-2 matter fields on the bounds. If we assume that such particles are absent, we of course obtain stronger bounds, as shown in figure 9, where we used the same space of functionals. Notice both regions with or without light matter share the same tangent line (4.6).

Our approach can also compute bounds for mixed amplitudes, beyond MHV. In figure 10, we derive bounds for Re $\widehat{g}_4$ (i.e., in all positive helicity configuration) and g_4 in terms of gravity. In particular, we instead use $n_{\max} = 6$ to help convergence of functionals built from combining $B_{2,3}^{\mathrm{imp}}$ and forward-limit of improved sum rules $\partial_{p^2}^q B_4^{(2)\,\mathrm{imp}}$ up to $q = 2$. Although we consider only Re $\widehat{g}_4$, this bound really applies to the magnitude $|\widehat{g}_4|$ since from the perspective of the graviton scattering amplitude the overall phase of a complex coupling can be removed by a little-group rotation and thus cannot be constrained (only relative phases between couplings can).
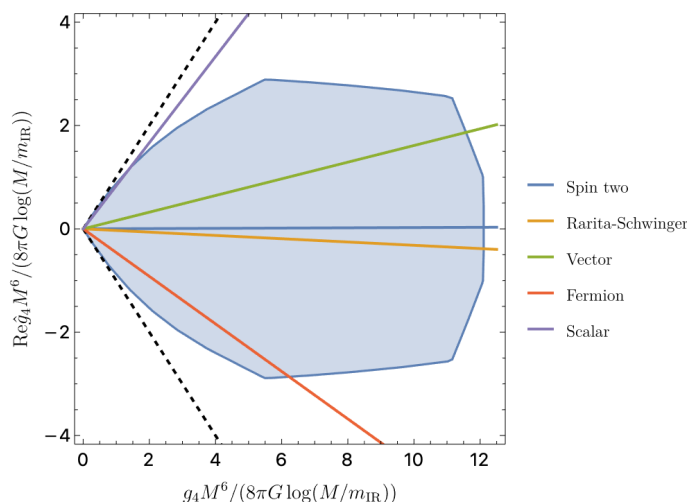
The dashed lines display the positivity constraints (2.37), which we reproduce here:

$$g_4 \pm |\widehat{g}_4| \geq 0\,. \tag{4.7}$$

We see that the allowed region is much smaller than the cone (2.37), which appears to be tangent to the allowed region. In particular, a theory which would saturate one of these inequalities is ruled out; this would correspond to a theory where either of α or α' in (2.8) is set to zero. (A similar conclusion was reached recently using seemingly different arguments which used only causality within the EFT [70]; it would be interesting to understand the relationship.)
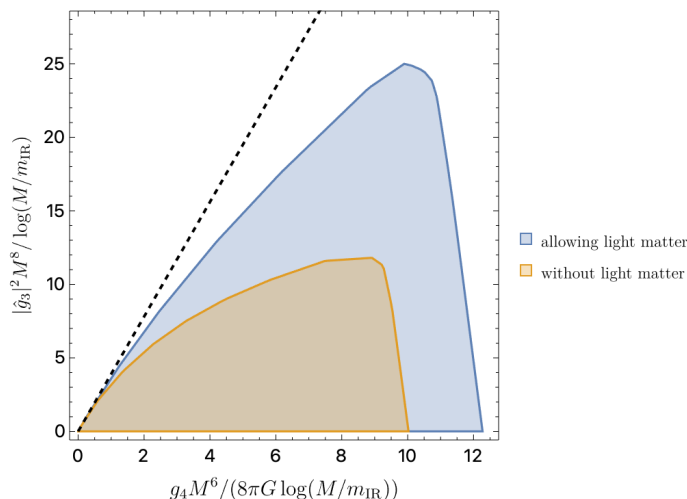
Figure 9. Similar to figure 8: the allowed region for $|\widehat{g}_3|^2$ and g_4 in terms of Newton's constant. For the blue region, light spin-0 and spin-2 massive states are allowed in the spectrum, while the smaller orange region applies to theories that have no such light fields. Both regions share the same tangent (4.6).
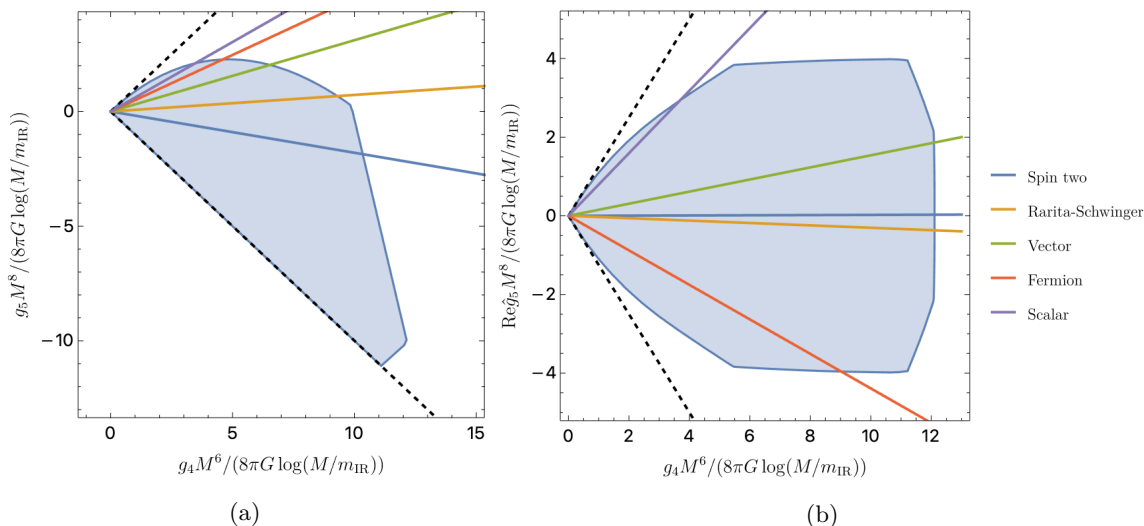

Figure 10. Allowed region for Re $\widehat{g}_4$ and g_4 in terms of Newton's constant, the spin-4 mass gap M, and an infrared cutoff m_{IR}, with light spin-0 and spin-2 matter fields allowed. The dashed lines show the positivity bounds (4.7). Bounds obtained from improved spin-2, 3 sum rules and $B_4^{(2)\,\text{imp}}$ at the second derivatives order. The lines display the loop amplitudes of subsection 4.1.

The solid lines in the figure 10 display the loop amplitudes of subsection 4.1 as a function of N, assuming a large number of species N. As visible from the plot, theories with $N \gg \frac{M^2}{G}$ are excluded.

4.3 Bounds involving $D^2 R^4$

Moving to the next derivative order, we apply spin-2 sum rules to relate g_5 to g_4 and G, see figure 11a. Of course, we already know from eq. (2.40) that g_5/g_4 has two-sided bounds,

Figure 11. Allowed region for (a) g_5 and g_4 and (b) Re $\widehat{g}_5$ and g_4 in units of Newton's constant, the spin-4 mass gap M, and an infrared cutoff m_{IR}, with light spin-0 and spin-2 matter fields allowed. The dashed lines show the positivity bounds (4.8) and (4.9) respectively. Solid lines display the loop amplitudes of subsection 4.1.

and that $g_4/G \log$ is bounded, and therefore we expected two-sided bounds on $g_5/G \log$ as well. Similarly we can compute bounds on non-MHV couplings, i.e., for $\widehat{g}_5$ in figure 11b. We follow the same strategy as for $\widehat{g}_4$ above to project onto the parity eigenstates and thus present bounds for Re $\widehat{g}_5$, although the bounds really apply to $|\widehat{g}_5|$.

An interesting aspect of these plots is that the allowed regions are rather smaller than a cone. The dashed lines in figure 11a are tangent to the plot at origin, which reproduces the positive bounds (2.40):

$$g_4 \pm g_5 M^2 \geq 0 \,. \tag{4.8}$$

Similarly, the dashed lines that are tangent to figure 11b at the origin correspond to simple positive bounds

$$g_4 \pm 0.8|\widehat{g}_5 M^2| \geq 0 \,. \tag{4.9}$$

Curiously, this bound appears stronger than what we could derive from forward limit functionals.

To compute bounds plotted in figures 11a and 11b, we truncated the space of functionals to $n_{\max} = 5$ built from improved sum rules to from spin-2 to spin-6, i.e., B_i^{imp} with $i = 2 \ldots 6$. In the former case we also include forward-limit of sum rules $\partial_t^k B_4^{(1)\,\text{imp}}(0)$ and $\partial_t^k B_5^{(1)\,\text{imp}}(0)$ and in the latter $\partial_{p^2}^q B_4^{(1)\,\text{imp}}(0)$ and $\partial_{p^2}^q B_4^{(2)\,\text{imp}}(0)$, with up to $q = 4$ to guarantee the large J behaviour of functionals are positive. Other detailed parameter choices are listed in table 2.

4.4 Bounds involving $D^4 R^4$ and low spin dominance

In section 2.5 we reviewed how expanding higher-spin sum rules around forward-limit produces homogeneous bounds involving e.g., g_6'/g_6, which gives eq. (2.43), which we
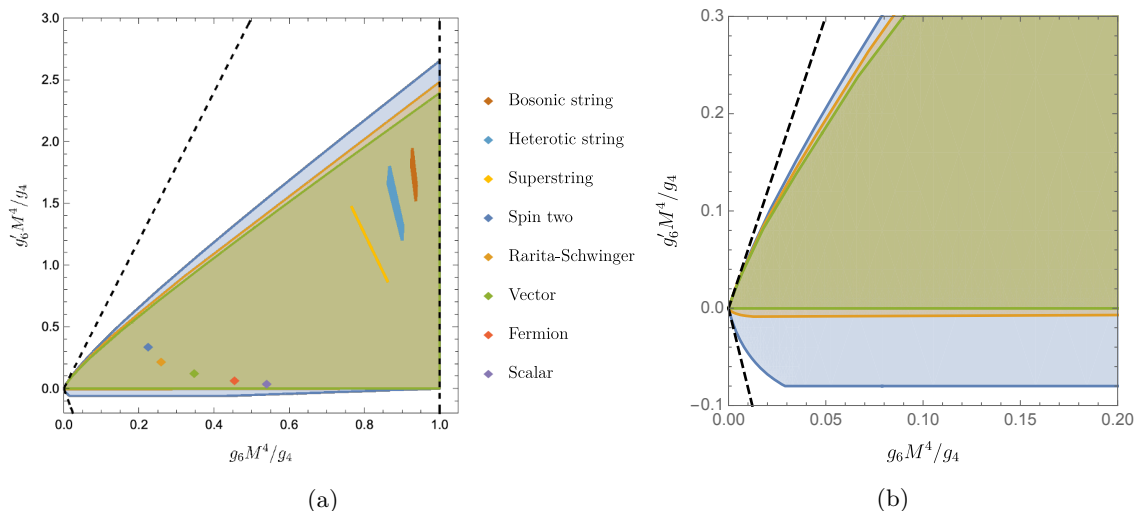
Figure 12. Allowed regions for g_6 and g_6', normalized by the quartic self-coupling, g_4. The blue, orange and olive regions show increasing derivative orders 6, 7 and 15 respectively. On the left we superimposed the values realized in the models of subsection 4.1. (b) is a zoom near the origin of the allowed region, showing the rapid convergence with derivative order. Although negative values of g_6' are excluded asymptotically, at any finite derivative order the boundary appears to be tangent to lines of slope 6 and $-\frac{90}{11}$.

reproduce here:

$$-\frac{90}{11} \leq \frac{g_6'}{g_6} \leq 6 \qquad \text{(using forward limits and a single null constraint)}. \tag{4.10}$$

An important observation made in [48] was that the space of couplings spanned by the theories in section 4.1, a.k.a "the theory island", is much smaller than that given by such homogeneous bounds. In [48], in order to approach the theory island, the authors propose an additional assumption called low-spin-dominance (LSD), which is a constraint on possible UV spectra stating that higher-spin states are suppressed compared to low-spin states. Quantitively, for MHV amplitudes, LSD implies

$$\text{LSD}: \qquad |c_{4,m^2}^{+\pm}| \geq \alpha |c_{J>4,m^2}^{+\pm}|^2, \tag{4.11}$$

where $\alpha \geq 1$ was used to parameterize the size of suppression of higher-spin states. By increasing α, ref. [48] pushes the bounds asymptotically to

$$0 \leq \frac{g_6'}{g_6} \leq 2 \qquad \text{(assuming LSD)}, \tag{4.12}$$

thus narrowing down the space of couplings to that spanned by the aforementioned theories.

In this paper we do not assume LSD. However, by considering inhomogeneous bounds involving g_6/g_4 and g_6'/g_4 of increasing derivative orders n (meaning null constraints having up to the same scaling dimension as the coupling g_n), we find that we can further narrow down the space of couplings as shown in figure 12.[15] As we increase the number of null

---

[15]This result was found concurrently in ref. [71], which also studies inhomogeneous bounds of the form $g_n M^{2(n-m)}/g_m$, and makes similar observations. We thank the authors of that paper for sharing their draft with us and coordinating submission.
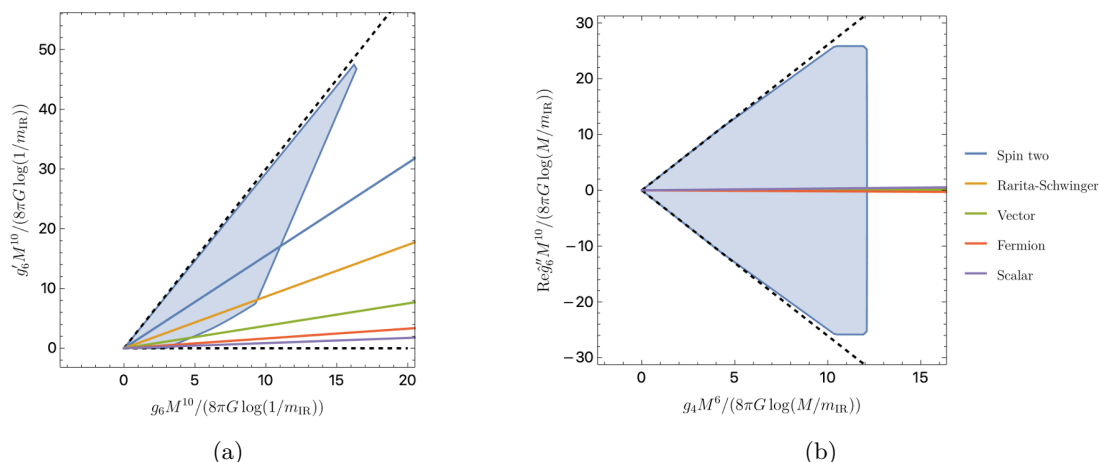
Figure 13. Allowed regions for (a) g_6 and g_6' and (b) Re $\widehat{g}_6''$ and g_4, normalized by gravity, and allowing light matter. The left panel rules out negative values of g_6'. Moving along the model lines corresponds to changing the number N of particles in the loop. Curiously, all considered models have very small values of $\widehat{g}_6''$.

constraints at higher derivative order, we observe that g_6' is approaching $g_6' \geq 0$. We can thus claim that positivity of g_6' holds asymptotically, which agrees with the prediction of LSD [48]. On the rightmost edge $g_6 M^4/g_4 = 1$ we find the absolute upper bound $g_6' M^4/g_4 \lesssim 2.38$, which is significantly closer to the ratio predicted by LSD.

We can do even better by considering impact-parameter bounds on g_6 and g_6' normalized by gravity, which are the main novelty of this paper. This can be computed by using B_i^{imp} with $i = 2 \dots 6$ with $n_{\max} = 8$, with the result shown in figure 13a. Surprisingly, we find, as we include spin-2 sum rules, we have the dashed line tangent to the plot at origin that gives bounds

$$0 \leq \frac{g_6'}{g_6} \leq 3 \qquad \text{(our bounds, using spin } k \geq 2 \text{ sum rules)} . \tag{4.13}$$

This lower bound is exactly the one predicted by LSD [48]. Furthermore, although the absolute upper bound is weaker than that predicted by LSD, the whole region is narrower since the upper dashed line is mostly excluded.

As an additional example, we can also provide bounds involving couplings which contribute to the $+ + +-$ helicity configuration. More concretely, we compute bounds on Re $\widehat{g}_6''$ and g_4 in terms of gravity using B_i^{imp} with $i = 2 \dots 7$ together with $\partial_{p^2}^q B_4^{(1)\,\mathrm{imp}}(0)$ and $\partial_{p^2}^q B_4^{(3)\,\mathrm{imp}}(0)$ up to $q = 4$, with the result shown in figure 13b where the dashed lines correspond to

$$g_4 \pm 2.61 |\widehat{g}_6''| M^4 \geq 0 . \tag{4.14}$$

We conclude that the discrepancy between dispersive bounds and the known "theory space" is very much reduced when more sum rules are used. Compared with previous work, our constraints are tighter (without making additional assumptions such as "low spin
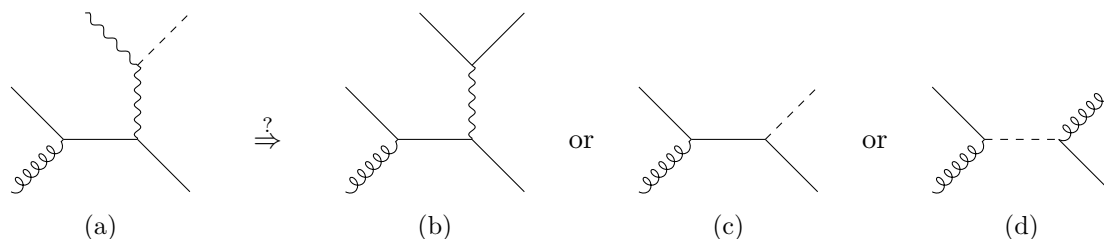
Figure 14. Production of a quark jet plus missing energy from a collision of a quark with a gluon; wavy lines denote graviton and dashed line higher-spin particles. The rapid high-energy growth of (a) could be softened by modifying its various vertices or propagators, however each option opens new processes: (b) if the higher-spin states are composite their components can be produced directly with $\mathcal{M} \sim (M_{\mathrm{pl}})^{-2}$ (c) if other vertical exchanges cancel the graviton contribution they can be produced directly with $\mathcal{M} \sim (M_{\mathrm{pl}})^{-1}$ (d) new horizontal exchanges would lead to resonances with $\mathcal{M} \sim (M_{\mathrm{pl}})^{0}$.

dominance") mainly as a result of including more functionals and considering inhomogeneous bounds. The later greatly helps since tangent slopes near the origin in figure 12 converge more slowly than other constraints. Given the tendency of higher-derivative coefficients to grow geometrically, we expect even more dramatic reductions in the allowed volume at higher derivative orders, although we have not studied those systematically.

4.5 Can higher-spin states be hidden from the Standard Model?

This subsection is less rigorous than the rest of this paper; we limit ourselves to non-exhaustive arguments and order-of-magnitude estimates.

What do collider searches tell us about higher spin particles, of the kind that can lead to modifications of GR? Heuristically, because gravity is universal and couples to all matter, one might expect that modifications to it also couple to everything. Indeed, many specific scenarios of modified gravity, such as string theory models, predict resonances that couple directly to Standard Model matter. The non-observation of such resonances impose strong constraints on the string scale: $M \gtrsim 7.7\,\mathrm{TeV}$, and on many other scenarios as well [72, 73]. Can model-independent constraints be made on potential low-scale modifications to GR?

To orient the discussion, let us imagine a (very hypothetical) scenario where the dynamics of a $10 M_{\odot}$ black hole, of size $L \sim 30\mathrm{km} \sim 1/(10^{-11}\mathrm{eV})$, were somehow observed to differ by more than 10^{-20} from GR's prediction. (Such a signal strength is orders-of-magnitude weaker than considered in either LIGO or EHT contexts [74, 75], but has been chosen to illustrate collider constraints.) An effective field theorist might try to attribute this to higher-derivative terms of size:[16]

$$\widehat{g}_3 \sim \frac{10^{-20}}{(10^{-11}\mathrm{eV})^4} \quad \text{or} \quad \frac{\widehat{g}_4}{16\pi G} \sim \frac{10^{-20}}{(10^{-11}\mathrm{eV})^6}\,, \tag{4.15}$$

---

[16]We recall from the introduction that "fifth forces" which do not grow with energy, involving for example direct couplings of matter to new spin 0 and spin 1 particles, are unconstrained by our arguments. We thus omit this possibility here.

the latter being only relevant if the former vanishes. Note that g_5 and higher-derivative couplings can never dominate over g_4, by (2.40). The bounds (4.4) indicate that light higher-spin states with $M < 10^{-6}$eV or $M < 10^{-8}$eV would then have to exist, depending on whether $\widehat{g}_3$ or $\widehat{g}_4$ dominates, respectively.[17] Could this be compatible with null results from collider searches?

Temporarily treating higher-spin states as point-like particles (left-panel of figure 14), there are two possible index structures for their couplings to two gravitons. These have respectively eight and four derivatives, corresponding to same- and opposite-helicity of the two gravitons. The production cross-section of the higher-spin particle at a hadron collider, in association with a graviton and a jet (to make missing momentum to trigger on) are then

$$|\mathcal{M}|^2_{pp\to\text{jet}+\not{E}}|_{(a)} \sim \begin{cases} \alpha_s \frac{\widehat{s}^8}{M_{\text{pl}}^4 M^{12}}, & \text{if } |c_{J=4}^{++}|^2 \neq 0, \\ \alpha_s \frac{\widehat{s}^4}{M_{\text{pl}}^4 M^4}, & \text{if } |c_{J=4}^{++}|^2 = 0. \end{cases} \tag{4.16}$$

Here $\widehat{s}$ is the partonic center of mass energy, and we have included a factor of α_s to create the jet. For crude estimates, we will consider that missing-energy searches [76, 77] exclude amplitudes of size $|\mathcal{M}| \gtrsim 1$ when $\sqrt{\widehat{s}} \sim$ TeV. Thus, despite the strong suppression by $M_{\text{pl}} \sim 10^{15}$TeV, we see that in the first scenario $M \lesssim$ MeV is clearly ruled out. However, the analysis in this paper does not rule out the option that $|c^{++}|^2 = 0$, which is the boundary of slope -1 in figure 11a; but in this case there is still a tension with collider searches if $M \lesssim 10^{-3}$eV. We conclude that if higher-spin particles are point-like, colliders and our bounds easily exclude effects of the size (4.15); effects on larger distances would be even smaller.

Now, the new states don't have to be point-like: the above amplitudes could be softer. However, softening mechanisms generically open up other production mechanisms that are also constrained, as in the stringy example discussed above. These are depicted in figure 14.

For example, the higher-spin states could be two-particle states of some light fields that couple to us only through gravity, which would soften the top vertex. However since two-graviton couplings in this scenario are small $\sim 1/M_{\text{pl}}^2$, long-distance effects are negligible unless there are a very large number of such light fields (for the same reason that quantum effects from Standard Model loops are extremely small for simple observables around macroscopic black holes [78]). Specifically, in order to have a 10^{-20} effect at L, at a minimum $N \gtrsim 10^{60}$ of them would be needed (such that $N/(M_{\text{pl}}^2 L^2) \sim 10^{-20}$; fields with mass $m < L^{-1}$ can be treated as effectively massless in this estimate). This would lead to a very low "cutoff" $M_{\text{pl}}/\sqrt{N}$ above which gravity becomes strongly coupled, and values $N > 10^{32}$ are typically not considered for this reason [79]. Besides possible cosmological constraints, let us simply mention that even with the minimal coupling (b), since $(\text{TeV}/M_{\text{pl}})^4 \sim 10^{-60}$, for such N even current colliders could have detected a production cross-section

$$|\mathcal{M}|^2_{pp\to\text{jet}+\not{E}}|_{(b)} \sim \alpha_s \frac{\widehat{s}^2}{M_{\text{pl}}^4} \times N. \tag{4.17}$$

---

[17]In this subsection we ignore infrared logarithms since $(\log M R_{\text{Universe}})^{1/4} \lesssim 3$.

Other softening mechanisms may modify other ingredients in figure (a). For example new higher-spin states could be exchanged between Standard Model fields and the higher-spin particle. But these could then be produced directly with a four-derivative graviton-strength coupling as in (c):

$$\left.|\mathcal{M}|^2_{pp\to\text{jet}+\not{E}}\right|_{(c)} \sim \alpha_s \frac{\widehat{s}^6}{M_{\text{pl}}^2 M^4}, \tag{4.18}$$

detectable if $M \lesssim 10\text{keV}$. This amplitude could again be softened, but again the cure is worse than the disease since this almost certainly requires new states exchanged on the horizontal propagator, leading to string-like resonances (d) as discussed above.

These estimates, if correct, challenge the notion that modifications of gravity at large scales could be hidden from colliders. In the considered scenarios, the amplitudes to produce higher-spin states either grow rapidly with energy, or get cancelled by new particles that would be themselves detectable. We hope that more robust and model-independent statements will be obtained in the future.

5 Conclusions

In this paper, we analyzed constraints on low-energy graviton-graviton scattering, assuming that causality and other basic principles apply at all energies. Theoretically, graviton scattering is an ideal way to probe potential modifications of Einstein's theory of gravity. At the same time, causality of a scattering process translates into well-understand mathematical properties like analyticity. This implies Kramers-Kronig-type dispersion relations, which express low-energy observables in terms of unknown but positive absorption probabilities at high energies, and which we have systematically analyzed.

Our key result (4.4) is simple to state: if a higher-derivative correction were measured, corresponding to $\mathcal{L} = \frac{1}{16\pi G}(R + r_0^4\text{Riem}^3 + \ldots)$, and if Nature respects causality as we understand it, then a spin-4 particle must exist whose Compton wavelength is at least as long as the length r_0: $M^{-1} > r_0$.

The notion that higher-derivative corrections can come from heavy particles may not surprise an effective field theorist. The scaling with M_{pl} of our bounds is however interesting. An effective field theorist might argue that the apparent breakdown of the derivative expansion at the length r_0 suggests the existence of new states with mass $\sim r_0^{-1}$ or lighter, but their friend could have objected that the states could be much heavier: the required three-point couplings respect unitarity bounds as long as $M \lesssim (\frac{M_{\text{pl}}^2}{r_0})^{\frac{1}{3}}$. We excluded the second option: new higher-spin states *must* exist with mass $M \lesssim r_0^{-1}$ or lighter. The M_{pl} scaling of our bounds means that heavy states couple to two gravitons with a strength that never exceeds the three-graviton coupling of Einstein's theory.

This means that the effective field theory approach to gravity can never describe modifications to Einstein's theory that are not parametrically small: in any situation where an EFT is justified at the distance L ($ML \gg 1$), corrections are small ($r_0/L \ll 1$).

Our bounds on cubic couplings are qualitatively similar to the parametric bounds obtained in [7], with the important novelty that we find sharp bounds with precise numerical coefficients. As discussed in section 3.5, this is done by imposing a quantum notion of

causality based on commutators and crossing symmetry, rather than a classical one based on time advances. Technically, we apply dispersion relations to a physical region where transverse momenta are below the cutoff M but energies are above M (but still much less than $M_{\rm pl}$, so that tree-level diagrams give a good approximation). The finite momentum transfer also allows us to constrain local (quartic) and non-local (cubic) self-interactions on the same footing.

Our bounds assume that new states satisfy $M \ll M_{\rm pl}$. They are expected to be valid modulo relative corrections $\sim \frac{M^2}{M_{\rm pl}^2}$ from loop effects within the EFT. It would be interesting to explicitly compute some of these corrections.

We treated the graviton as exactly massless. By continuity, we expect our bounds to also control the scattering of transversely polarized gravitons with $m \ll M$. For massive gravitons, one however anticipates much stronger constraints coming from the study of longitudinal polarizations, which do not have a smooth $m \to 0$ limit (see for example [35, 80, 81]).

Our results significantly reduce the gap observed in [48] between dispersive bounds and "theory space", as further discussed in section 4.4. Yet, it is noteworthy that our bounds are not saturated by known theories. This could indicate that stronger bounds are still possible; this is of course also suggested by the fact that all our bounds involving G contain infrared logarithms. It will be interesting to find a way to obtain infrared-safe bounds, and to further investigate the gap between theory space and dispersive bounds.

As explained in [39], dispersive bounds for scalar amplitudes in flat space can be lifted to AdS and imply corresponding bounds on OPE coefficients of holographic CFTs. Physically, the same must be true for graviton scattering as well. For example, (4.4) translates to a bound on three-point OPE coefficients of stress tensors in holographic 3d CFTs. In AdS$_4$, infrared logarithms are simply replaced by $\log(M R_{\rm AdS})$.

The notion of causality that we used in the paper applies to scattering processes in an otherwise empty and flat region of space. Physically, since we mainly probe energies of order M and impact parameters $b \sim M^{-1}$, we expect our conclusions to remain valid in any spacetime in which a flat patch of radius $\gg M^{-1}$ exists, including our own. (This is the reason for the agreement between bounds in flat space and AdS.) Evidently, studying energies above the EFT cutoff was essential to probe the mass and couplings of higher-spin states. An important question is whether some of our bounds can be understood from thought experiments within the EFT itself.

From the generic collider estimates given in section 4.5, assuming causality as we understand it, we find it difficult to imagine that higher-derivative terms could be visible at macroscopic distances. A better understanding of interactions between gravitational higher-spin states and matter should lead to more precise constraints.

In summary, experimental verifications of Einstein's theory test a basic question: can signals travel faster than the speed of light?

Acknowledgments

A Spinning partial waves and Wigner-d functions

The Wigner-d functions are defined as

$$d^J_{h,h'}(x) = \mathcal{N}^J_{h,h'} \left(\frac{1+x}{2}\right)^{\frac{h+h'}{2}} \left(\frac{1-x}{2}\right)^{\frac{h-h'}{2}} {}_2F_1\left(h-J, J+h+1; h-h'+1; \frac{1-x}{2}\right), \quad (A.1)$$

where we defined the normalization

$$\mathcal{N}^J_{h,h'} = \frac{1}{\Gamma(h-h'+1)} \sqrt{\frac{\Gamma(J+h+1)\Gamma(J-h'+1)}{\Gamma(J-h+1)\Gamma(J+h'+1)}}, \quad (A.2)$$

so that $d^J_{h,h'}(1) = \delta_{h,h'}$. In the main text we will use instead the helicity-stripped d-functions

$$\tilde{d}^J_{h',h}(x) = \mathcal{N}^J_{h,h'} \, {}_2F_1\left(h-J, J+h+1; h-h'+1; \frac{1-x}{2}\right). \quad (A.3)$$

Note that $\tilde{d}^J_{0,0}(x)$ is the familiar Legendre polynomial. Some useful properties of these functions are the symmetries

$$d^J_{-h-h'}(x) = d^J_{h'h}(x), \qquad d^J_{h',h}(x) = (-1)^{h-h'} d^J_{h,h'}(x), \quad (A.4)$$

and their transformation under parity

$$d^J_{h,h'}(-x) = (-1)^{J+h} d^J_{h,-h'}(x), \quad (A.5)$$

as well as orthogonality

$$\int_{-1}^{1} dx \, d^J_{h,h'}(x) \, d^{J'}_{h,h'}(x) = \frac{2}{2J+1} \delta_{JJ'}. \quad (A.6)$$

Physically, the decomposition (2.13) corresponds to exchanging heavy states X with the following three-point amplitudes, which are uniquely fixed by Poincaré symmetry [83]:

$$\mathcal{M}(1^+, 2^-, 3^X) = c^{+-}_{J,m^2} \sqrt{16\pi m^2} \, \langle \mathbf{13}\rangle^{J-4} \langle \mathbf{23}\rangle^{J+4} \langle 12\rangle^{-J} \quad (A.7)$$

$$\mathcal{M}(1^+, 2^+, 3^X) = c^{++}_{J,m^2} \sqrt{16\pi m^2} \langle \mathbf{13}\rangle^J \langle \mathbf{23}\rangle^J \langle 12\rangle^{-2-J} [12]^2. \quad (A.8)$$

We can then glue two three-point amplitudes by summing over immediate polarizations to conclude the partial wave expansion eq. (2.11) (where $a_J^h(s)$ is replaced by eq. (2.13)). For example, for $\mathcal{M}^{+--+}$, we have [83]

$$\sum_X \mathcal{M}(1^+, 2^-, 5^X)\mathcal{M}(3^-, 4^+, 5^{X^*})$$

$$= |c_{J,m^2}^{+-}|^2 16\pi m^2 \langle 12\rangle^{-J}\langle 34\rangle^{-J} \times \sum_j \frac{(J+4)!^2(J-4)!^2}{j!(J+4-j)!^2(j-8)!}\langle 13\rangle^j\langle 14\rangle^{J+4-j}\langle 23\rangle^{J+4-j}\langle 24\rangle^{j-8}.$$

(A.9)

We can then convert to the center-of-mass frame where

$$\lambda_1 = \sqrt{m}\begin{pmatrix}1\\0\end{pmatrix}, \quad \lambda_2 = \sqrt{m}\begin{pmatrix}0\\1\end{pmatrix}, \quad \lambda_3 = i\sqrt{m}\begin{pmatrix}\sqrt{\frac{1+x}{2}}\\\sqrt{\frac{1-x}{2}}\end{pmatrix}, \quad \lambda_4 = \sqrt{m}\begin{pmatrix}\sqrt{\frac{1-x}{2}}\\-\sqrt{\frac{1+x}{2}}\end{pmatrix}.$$

(A.10)

Using this parameterization, eq. (A.9) can then be analytically summed to Wigner-d functions.

We will also be interested in the behavior at large J, taken with fixed impact parameter $b = 2J/m$, which is given by

$$\lim_{J,m\to\infty} \tilde{d}_{h,h'}^J\left(1 - \frac{2p^2}{m^2}\right) = \frac{J_{h-h'}(bp)}{(p/m)^{h-h'}},$$

(A.11)

where $J_{h-h'}(bp)$ on the right-hand side is a Bessel function of the first kind.

The following are also useful when expanding around the forward limit, which corresponds to expanding x around 1

$$\frac{d^n}{dx^n}\tilde{d}_{0,0}^J(x)\Big|_{x=1} \geq 0, \qquad \frac{d^n}{dx^n}(\tilde{d}_{4,4}^J(x) - (-1)^J\tilde{d}_{4,-4}^J(x))\Big|_{x=1} \geq 0.$$

(A.12)

B Amplitudes from matter and Kaluza-Klein exchanges

In this appendix, we record the contribution from light "matter" fields to the low-energy amplitudes (2.7). If the light matter fields have discrete masses m_ℓ, these amplitudes are rational functions whose poles match the spin-0 and spin-2 contributions to the above partial waves:

$$f_{\text{matter}}(s,u) = \sum_{m_\ell < M}\frac{|g_0^{++}(m_\ell)|^2}{m_\ell^2 - t} + \sum_{m_\ell < M}\frac{|g_2^{++}(m_\ell)|^2(1 - 6su/m_\ell^4)}{m_\ell^2 - t}.$$

(B.1)

To get the last factor, for example, we used the $J = 2$ partial wave (2.15b) on the $t = m_\ell^2$ pole:

$$\tilde{d}_{0,0}^2\left(1 + \frac{2s}{m_\ell^2}\right) = 1 + \frac{6s(s + m_\ell^2)}{m_\ell^4} \simeq 1 - \frac{6su}{m_\ell^4} \qquad \text{when } t = m_\ell^2.$$

(B.2)

Away from the pole, the spin-2 exchange (B.1) is unique up to contact interactions, which are, to the same derivative order, either linear in t or constant. (That is, different choices

of writing (B.1) could shift the g_4 and g_5 coefficients in (2.7) by multiples of $|g_2^{++}(m_\ell)|^2$.)
The above choice will lead to the amplitude with the best possible Regge growth (spin-4 at fixed-s, and spin-2 at fixed-t), and to the simplest sum rules, but other choices would not significantly change the analysis.

In reality, since they can decay (in particular, to two gravitons), the light fields cannot give sharp poles but at best resonances; this simply replaces the sum by an integral:

$$\sum_{m_\ell < M} \frac{|g_J^{++}(m_\ell)|^2}{m_\ell^2 - t}(\cdots) \mapsto 16\pi(2J+1) \int_0^{M^2} \frac{dm_\ell^2}{2\pi m_\ell^8} \frac{|c_{J,m_\ell^2}^{++}|^2}{m_\ell^2 - t - i0}(\cdots), \qquad (B.3)$$

where the partial waves $c_{J,m_\ell^2}^{++}$ are normalized as in (2.15).

Only same-helicity couplings g_2^{++} appear in the above due to angular momentum rules, which forbid a particle of spin < 4 from decaying to a $(+-)$ two-graviton state. For the same reason, the single minus amplitudes do not receive contributions

$$g_{\text{matter}}(s, u) = 0. \qquad (B.4)$$

Finally, the all-plus amplitudes are

$$h_{\text{matter}}(s, u) = \sum_{m_\ell < M} g_0^{++}(m_\ell)^2 \left(\frac{s^4}{m_\ell^2 - s} + \frac{t^4}{m_\ell^2 - t} + \frac{u^4}{m_\ell^2 - u} \right) \qquad (B.5a)$$

$$+ \sum_{m_\ell < M} g_2^{++}(m_\ell)^2 \left(\frac{s^2(s^2 - 6tu)}{m_\ell^2 - s} + \frac{t^2(t^2 - 6su)}{m_\ell^2 - t} + \frac{u^2(u^2 - 6st)}{m_\ell^2 - u} \right). \quad (B.5b)$$

It is important here that each numerator is quartic in Mandelstam invariants, so as to cancel the denominator in (2.3). Contact ambiguities have been fixed to avoid high-energy growth with Regge spin greater than 2.

C Details on numerics

In this appendix, we give details on our numerical implementation of dispersive bounds. We mostly follow [15], though some new complications arise when considering spinning external particles (as opposed to scalars). Most of the issues are not specific to 4d, so we phrase our discussion in general spacetime dimensions.

A generic dispersive optimization problem takes the following form:

$$\text{maximize} \quad \sum_i \int_0^M dp\, \psi_i(p)\, B_i(p^2)\Big|_{\text{low}}$$

$$\text{such that} \quad \sum_i \int_0^M dp\, \psi_i(p) B_i(p^2)[m^2, J, \lambda] \succeq 0 \quad \forall m > M, \text{ reps } \rho = [J, \lambda]. \qquad (C.1)$$

The $B_i(p^2)$ are dispersive sum rules, labeled by an index i, evaluated at $t = -p^2$. The wavefunctions $\psi_i(p)$ are chosen to project out all but a subset of couplings in the low-energy contribution $B_i(p^2)\big|_{\text{low}}$. Here, ρ runs over little-group representations of massive particles

that can appear in intermediate channels. We can associate to each ρ a Young diagram $[J, m_2, \ldots, m_n]$ where $n = \lfloor \frac{d-1}{2} \rfloor$. We denote $\lambda = [m_2, \ldots, m_n]$ and write $\rho = [J, \lambda]$. Aside from a few special cases at small J, the ρ's fall into families where J can be arbitrarily large, while λ remains fixed. For each ρ-family, the heavy density $B_i[p^2, m^2, J, \lambda]$ is an $n \times n$ matrix, where n (which depends on λ) is the number of three-point structures for two external particles producing an internal ρ-particle.

Following [15], we choose a polynomial basis for the wavefunctions $\psi_i(p)$:

$$\psi_i(p) = \sum_{n=0}^{N} a_{i,n} p^n. \tag{C.2}$$

Then (C.1) becomes a semidefinite program with decision variables $a_{i,n}$ and an infinite number of positivity constraints labeled by m, ρ. We truncate to a finite number of constraints using a combination of discretization and polynomial approximations. Specifically, we split (m, ρ)-space into the following four regimes:

Fixed J, fixed m. Choosing some $J_{\max}$, we impose positivity at the finite set of representations $\rho = [J, \lambda]$ with $J \le J_{\max}$ and masses $m^2 = 1/(1-x)$, where $x \in \{0, \delta x, 2\delta x, \ldots, \lfloor \frac{1}{\delta x} \rfloor \delta x\}$. Our detailed parameter choices are listed in table 2.

Large J, fixed m. Unlike in the case of external scalar particles [15], we found that it is important to explicitly control the large-J limit of our functionals at fixed m. To do so, we compute a series expansion of the integrated heavy density at large J

$$\int_0^1 dp\, p^n B_i[p^2, m^2, J, \lambda] \sim \#J^r + \#J^{r-1} + \ldots. \tag{C.3}$$

Truncating this expansion to a finite number of terms, and multiplying by an appropriate power of J, we obtain a polynomial approximation for the heavy density as a function of J. We discretize m as before and, for each m, impose positivity of the resulting matrix polynomial of J on the interval $[J_{\max}, \infty)$.

Large m, fixed $b = \frac{2J}{m}$. An important set of positivity constraints come from the impact parameter scaling limit $m \to \infty$ with fixed $b = \frac{2J}{m}$. In this regime, the heavy densities behave as

$$\lim_{m \to \infty} B_i\left[p^2, m^2, \frac{bm}{2}, \lambda\right] \sim \frac{\mathcal{J}_{i,\lambda}(pb)}{m^{2k_i}}, \tag{C.4}$$

where k_i is the Regge spin of the i-th sum rule. Consequently, only the lowest-spin sum rules contribute in this limit. Thus, in the scaling limit, we set all higher-spin sum rules to zero and replace the lowest-spin sum rules with their approximations (C.4). The matrices $\mathcal{J}_{i,\lambda}(pb)$ have entries that involve Bessel functions and their derivatives. Following [15], we choose a cutoff B and impose positivity at a discrete set of $b \le B$, $b \in \{\epsilon_b, \epsilon_b + \delta_b, \ldots, \epsilon_b + \lceil \frac{B - \epsilon_b}{\delta_b} - 1 \rceil \delta_b\}$.

Large m, large b. As in [15], we must take care to impose positivity in the impact parameter scaling limit for $b > B$. Let us review the trick used there. For scalar scattering, the heavy density $C^{\text{imp}}_{2,-p^2}[m^2, J]$ is simply a 1×1 matrix, with scaling limit

$$\lim_{m \to \infty} C^{\text{imp}}_{2,-p^2}[m^2, \tfrac{mb}{2}] = \frac{2\Gamma(\frac{d-2}{2})}{m^4} \frac{J_{\frac{d-4}{2}}(bp)}{(bp/2)^{\frac{d-4}{2}}} \, . \tag{C.5}$$

Integrating against a wavefunction, we have

$$\int_0^1 dp\, \psi(p) \frac{J_{\frac{d-4}{2}}(bp)}{(bp/2)^{\frac{d-4}{2}}} = A(b) + B(b) \cos\left(b - \frac{\pi(d-1)}{4}\right) + C(b) \sin\left(b - \frac{\pi(d-1)}{4}\right) , \tag{C.6}$$

where $A(b), B(b), C(b)$ have well-behaved asymptotic expansions in inverse powers of b. The large-b behavior of $A(b)$ can be determined from the small-p behavior of the integral on the left-hand side. The oscillating terms $B(b)$ and $C(b)$ come from the other endpoint of the integral $p = 1$.

The key idea is to replace (C.6) with

$$A(b) + B(b) \cos\phi + C(b) \sin\phi \tag{C.7}$$

and impose positivity for $b > B$ and *arbitrary* ϕ — not just for $\phi = b - \frac{\pi(d-1)}{4}$. Although this is a stronger condition than positivity of (C.6) (and thus, could lead to weaker bounds), we expect it to be a good approximation at large b, where $b - \frac{\pi(d-1)}{4}$ is a rapidly-varying phase relative to $A(b), B(b), C(b)$.

To see how to impose positivity of (C.7), we write

$$A(b) + B(b) \cos\phi + C(b) \sin\phi = \begin{pmatrix} \cos\frac{\phi}{2} & \sin\frac{\phi}{2} \end{pmatrix} M(b) \begin{pmatrix} \cos\frac{\phi}{2} \\ \sin\frac{\phi}{2} \end{pmatrix} , \tag{C.8}$$

where

$$M(b) = \begin{pmatrix} A(b) + B(b) & C(b) \\ C(b) & A(b) - B(b) \end{pmatrix} . \tag{C.9}$$

Positivity of (C.8) for arbitrary ϕ is equivalent to the statement that $M(b)$ is positive-semidefinite:

$$M(b) \succeq 0 \, . \tag{C.10}$$

We can now expand $M(b)$ at large b, and approximate it as a power of b times a 2×2 matrix polynomial of b. We then impose positivity of this matrix polynomial for $b \geq B$.

Let us now consider the case of interest for this work, where the heavy density is a matrix. In the impact parameter scaling limit, the integral against wavefunctions $\psi_i(p)$ has the same form as before:

$$\sum_i \int_0^1 dp\, \psi_i(p) \mathcal{J}_{i,\lambda}(pb) = A_\lambda(b) + B_\lambda(b) \cos\left(b - \frac{\pi(d-1)}{4}\right) + C_\lambda(b) \sin\left(b - \frac{\pi(d-1)}{4}\right) , \tag{C.11}$$

where now $A_\lambda(b), B_\lambda(b), C_\lambda(b)$ are $n \times n$ matrices (n depends on λ). As before, we replace $b - \frac{\pi(d-1)}{4}$ with ϕ and seek to impose positivity for arbitrary ϕ. In other words, we would like to impose

$$\left(\vec{v}\cos\tfrac{\phi}{2} \;\; \vec{v}\sin\tfrac{\phi}{2}\right) M_\lambda(b) \begin{pmatrix} \vec{v}\cos\tfrac{\phi}{2} \\ \vec{v}\sin\tfrac{\phi}{2} \end{pmatrix} \geq 0 \qquad \text{for all } \phi \in \mathbb{R},\ \vec{v} \in \mathbb{R}^n\,, \tag{C.12}$$

where $M_\lambda(b)$ is a $2n \times 2n$ matrix of the form (C.9) with $n \times n$ block entries.

Note that we can freely rescale the vectors $\vec{v}$ and $(\cos\tfrac{\phi}{2}, \sin\tfrac{\phi}{2})$ without changing the positivity conditions (C.12). Thus, we can think of $\vec{v}$ as parametrizing a point $[v_1 : \cdots : v_n] \in \mathbb{RP}^{n-1}$, and $(\cos\tfrac{\phi}{2}, \sin\tfrac{\phi}{2})$ as parametrizing a point $[w_1 : w_2] = [\cos\tfrac{\phi}{2} : \sin\tfrac{\phi}{2}] \in \mathbb{RP}^1$. In this language, (C.12) is equivalent to imposing that $M_\lambda(b)$ is positive on the image of the Segre embedding

$$\sigma : \mathbb{RP}^1 \times \mathbb{RP}^{n-1} \to \mathbb{RP}^{2n-1}$$
$$\sigma : ([w_1 : w_2], [v_1 : \cdots : v_n]) \mapsto [w_1 v_1 : \cdots : w_1 v_n : w_2 v_1 : \cdots : w_2 v_n]\,. \tag{C.13}$$

A theorem of [84] states that if a quadratic form Q is nonnegative on a variety $X \subseteq \mathbb{RP}^{k-1}$ of minimal degree, then Q is a sum of squares of linear forms (and hence represented by a positive semidefinite matrix on $\mathbb{R}^k$). X has minimal degree if it is nondegenerate (not contained in a hyperplane) and $\deg(X) = 1 + \mathrm{codim}(X)$. Fortunately, the image of the Segre embedding σ has precisely these properties: it is nondegenerate, has degree n, and has codimension $n - 1$.[18] Hence, we conclude that (C.12) is equivalent to the statement that $M_\lambda(b)$ is a positive-semidefinite $2n \times 2n$ matrix:

$$M_\lambda(b) \succeq 0\,. \tag{C.14}$$

It is remarkable that such a simple condition captures the necessary positivity conditions, even for $n \neq 1$. We can now proceed as in the scalar case: we expand $M_\lambda(b)$ at large b to $r_{\max}$ subleading orders and approximate it in terms of a matrix polynomial of b. We then impose positivity of this matrix polynomial for $b \geq B$.

Having imposed positivity of the heavy density in these four regimes (fixed m and J, large J and fixed m, large m and fixed b, and large m and b), we find that the resulting functionals are positive in practice for almost all m, J. Violations of positivity come from the functional becoming slightly negative between discretized values of m or b. Such violations can usually be fixed by perturbing the functional slightly (for example by including a small admixture of another nearly positive functional).

Our parameter choices are listed in table 2.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](#)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited. SCOAP[3] supports the goals of the International Year of Basic Sciences for Sustainable Development.

---

[18]In fact, the image of σ is an example of a *rational normal scroll*, which is one of the three families of minimal degree varieties, according to the classification [85].

$J_{\max}$	42
δ_x	1/400
ϵ_b	1/250
δ_b	1/32
B	40
$r_{\max}$	4
non-default SDPB parameters	`--precision=768`

Table 2. Table of parameters used for discretization and polynomial approximations.

References

[1] T. Clifton, P.G. Ferreira, A. Padilla and C. Skordis, *Modified Gravity and Cosmology*, *Phys. Rept.* **513** (2012) 1 [arXiv:1106.2476] [INSPIRE].

[2] P. Bull et al., *Beyond ΛCDM: Problems, solutions, and the road ahead*, *Phys. Dark Univ.* **12** (2016) 56 [arXiv:1512.05356] [INSPIRE].

[3] M. Ishak, *Testing General Relativity in Cosmology*, *Living Rev. Rel.* **22** (2019) 1 [arXiv:1806.10122] [INSPIRE].

[4] P. Bueno, P.A. Cano, V.S. Min and M.R. Visser, *Aspects of general higher-order gravities*, *Phys. Rev. D* **95** (2017) 044010 [arXiv:1610.08519] [INSPIRE].

[5] S. Weinberg, *Photons and Gravitons in S-Matrix Theory: Derivation of Charge Conservation and Equality of Gravitational and Inertial Mass*, *Phys. Rev.* **135** (1964) B1049 [INSPIRE].

[6] S. Weinberg, *Photons and gravitons in perturbation theory: Derivation of Maxwell's and Einstein's equations*, *Phys. Rev.* **138** (1965) B988 [INSPIRE].

[7] X.O. Camanho, J.D. Edelstein, J. Maldacena and A. Zhiboedov, *Causality Constraints on Corrections to the Graviton Three-Point Coupling*, *JHEP* **02** (2016) 020 [arXiv:1407.5597] [INSPIRE].

[8] R. Brustein and Y. Sherf, *Causality Violations in Lovelock Theories*, *Phys. Rev. D* **97** (2018) 084019 [arXiv:1711.05140] [INSPIRE].

[9] Y. Sherf, *Hyperbolicity Constraints in Extended Gravity Theories*, *Phys. Scripta* **94** (2019) 085005 [arXiv:1806.09984] [INSPIRE].

[10] S. Gao and R.M. Wald, *Theorems on gravitational time delay and related issues*, *Class. Quant. Grav.* **17** (2000) 4999 [gr-qc/0007021] [INSPIRE].

[11] C.Y.-R. Chen, C. de Rham, A. Margalit and A.J. Tolley, *A cautionary case of casual causality*, *JHEP* **03** (2022) 025 [arXiv:2112.05031] [INSPIRE].

[12] A. Adams et al., *Causality, analyticity and an IR obstruction to UV completion*, *JHEP* **10** (2006) 014 [hep-th/0602178] [INSPIRE].

[13] C. de Rham, S. Melville, A.J. Tolley and S.-Y. Zhou, *Positivity bounds for scalar field theories*, *Phys. Rev. D* **96** (2017) 081702 [arXiv:1702.06134] [INSPIRE].

[14] S. Caron-Huot and V. Van Duong, *Extremal Effective Field Theories*, *JHEP* **05** (2021) 280 [arXiv:2011.02957] [INSPIRE].

[15] S. Caron-Huot, D. Mazac, L. Rastelli and D. Simmons-Duffin, *Sharp boundaries for the swampland*, *JHEP* **07** (2021) 110 [arXiv:2102.08951] [INSPIRE].

[16] A.J. Tolley, Z.-Y. Wang and S.-Y. Zhou, *New positivity bounds from full crossing symmetry*, *JHEP* **05** (2021) 255 [arXiv:2011.02400] [INSPIRE].

[17] N. Arkani-Hamed, T.-C. Huang and Y.-T. Huang, *The EFT-Hedron*, *JHEP* **05** (2021) 259 [arXiv:2012.15849] [INSPIRE].

[18] L.-Y. Chiang et al., *Into the EFThedron and UV constraints from IR consistency*, *JHEP* **03** (2022) 063 [arXiv:2105.02862] [INSPIRE].

[19] B. Bellazzini et al., *Positive moments for scattering amplitudes*, *Phys. Rev. D* **104** (2021) 036006 [arXiv:2011.00037] [INSPIRE].

[20] A. Sinha and A. Zahed, *Crossing Symmetric Dispersion Relations in Quantum Field Theories*, *Phys. Rev. Lett.* **126** (2021) 181601 [arXiv:2012.04877] [INSPIRE].

[21] L. Alberte, C. de Rham, S. Jaitly and A.J. Tolley, *QED positivity bounds*, *Phys. Rev. D* **103** (2021) 125020 [arXiv:2012.05798] [INSPIRE].

[22] X. Li et al., *Positivity in Multifield Effective Field Theories*, *Phys. Rev. Lett.* **127** (2021) 121601 [arXiv:2101.01191] [INSPIRE].

[23] P. Haldar, A. Sinha and A. Zahed, *Quantum field theory and the Bieberbach conjecture*, *SciPost Phys.* **11** (2021) 002 [arXiv:2103.12108] [INSPIRE].

[24] P. Raman and A. Sinha, *QFT, EFT and GFT*, *JHEP* **12** (2021) 203 [arXiv:2107.06559] [INSPIRE].

[25] J. Henriksson, B. McPeak, F. Russo and A. Vichi, *Rigorous bounds on light-by-light scattering*, *JHEP* **06** (2022) 158 [arXiv:2107.13009] [INSPIRE].

[26] A. Zahed, *Positivity and geometric function theory constraints on pion scattering*, *JHEP* **12** (2021) 036 [arXiv:2108.10355] [INSPIRE].

[27] L. Alberte et al., *Positivity Constraints on Interacting Spin-2 Fields*, *JHEP* **03** (2020) 097 [arXiv:1910.11799] [INSPIRE].

[28] M. Accettulli Huber, A. Brandhuber, S. De Angelis and G. Travaglini, *Eikonal phase matrix, deflection angle and time delay in effective field theories of gravity*, *Phys. Rev. D* **102** (2020) 046014 [arXiv:2006.02375] [INSPIRE].

[29] L. Alberte, C. de Rham, S. Jaitly and A.J. Tolley, *Reverse Bootstrapping: IR Lessons for UV Physics*, *Phys. Rev. Lett.* **128** (2022) 051602 [arXiv:2111.09226] [INSPIRE].

[30] S.D. Chowdhury et al., *Crossing Symmetric Spinning S-matrix Bootstrap: EFT bounds*, *SciPost Phys.* **13** (2022) 051 [arXiv:2112.11755] [INSPIRE].

[31] B. Bellazzini, M. Riembau and F. Riva, *IR side of positivity bounds*, *Phys. Rev. D* **106** (2022) 105008 [arXiv:2112.12561] [INSPIRE].

[32] C. Zhang, *SMEFTs living on the edge: determining the UV theories from positivity and extremality*, *JHEP* **12** (2022) 096 [arXiv:2112.11665] [INSPIRE].

[33] Y.-J. Wang, F.-K. Guo, C. Zhang and S.-Y. Zhou, *Generalized positivity bounds on chiral perturbation theory*, *JHEP* **07** (2020) 214 [arXiv:2004.03992] [INSPIRE].

[34] T. Trott, *Causality, unitarity and symmetry in effective field theory*, *JHEP* **07** (2021) 143 [arXiv:2011.10058] [INSPIRE].

[35] C. de Rham, S. Melville, A.J. Tolley and S.-Y. Zhou, *UV complete me: Positivity Bounds for Particles with Spin*, *JHEP* **03** (2018) 011 [arXiv:1706.02712] [INSPIRE].

[36] C. de Rham, S. Melville, A.J. Tolley and S.-Y. Zhou, *Positivity Bounds for Massive Spin-1 and Spin-2 Fields*, *JHEP* **03** (2019) 182 [arXiv:1804.10624] [INSPIRE].

[37] Z.-Y. Wang, C. Zhang and S.-Y. Zhou, *Generalized elastic positivity bounds on interacting massive spin-2 theories*, *JHEP* **04** (2021) 217 [arXiv:2011.05190] [INSPIRE].

[38] M. Kologlu, P. Kravchuk, D. Simmons-Duffin and A. Zhiboedov, *Shocks, Superconvergence, and a Stringy Equivalence Principle*, *JHEP* **11** (2020) 096 [arXiv:1904.05905] [INSPIRE].

[39] S. Caron-Huot, D. Mazac, L. Rastelli and D. Simmons-Duffin, *AdS bulk locality from sharp CFT bounds*, *JHEP* **11** (2021) 164 [arXiv:2106.10274] [INSPIRE].

[40] S. Weinberg, *The Quantum theory of fields. Volume 1: Foundations*, Cambridge University Press (2005) [DOI:10.1017/CBO9781139644167] [INSPIRE].

[41] T.D. Lee and G.C. Wick, *Negative Metric and the Unitarity of the S Matrix*, *Nucl. Phys. B* **9** (1969) 209 [INSPIRE].

[42] B. Grinstein, D. O'Connell and M.B. Wise, *Causality as an emergent macroscopic phenomenon: The Lee-Wick O(N) model*, *Phys. Rev. D* **79** (2009) 105019 [arXiv:0805.2156] [INSPIRE].

[43] J.F. Donoghue and G. Menezes, *Unitarity, stability and loops of unstable ghosts*, *Phys. Rev. D* **100** (2019) 105006 [arXiv:1908.02416] [INSPIRE].

[44] R.P. Woodard, *Ostrogradsky's theorem on Hamiltonian instability*, *Scholarpedia* **10** (2015) 32243 [arXiv:1506.02210] [INSPIRE].

[45] J.M. Cline, S. Jeon and G.D. Moore, *The Phantom menaced: Constraints on low-energy effective ghosts*, *Phys. Rev. D* **70** (2004) 043543 [hep-ph/0311312] [INSPIRE].

[46] R.E. Cutkosky, P.V. Landshoff, D.I. Olive and J.C. Polkinghorne, *A non-analytic S matrix*, *Nucl. Phys. B* **12** (1969) 281 [INSPIRE].

[47] H. Elvang and Y.-T. Huang, *Scattering Amplitudes*, arXiv:1308.1697 [INSPIRE].

[48] Z. Bern, D. Kosmopoulos and A. Zhiboedov, *Gravitational effective field theory islands, low-spin dominance, and the four-graviton amplitude*, *J. Phys. A* **54** (2021) 344002 [arXiv:2103.12728] [INSPIRE].

[49] P. Bueno, P.A. Cano, J. Moreno and Á. Murcia, *All higher-curvature gravities as Generalized quasi-topological gravities*, *JHEP* **11** (2019) 062 [arXiv:1906.00987] [INSPIRE].

[50] A. Hebbar, D. Karateev and J. Penedones, *Spinning S-matrix bootstrap in 4d*, *JHEP* **01** (2022) 060 [arXiv:2011.11708] [INSPIRE].

[51] E. Pajer, D. Stefanyszyn and J. Supeł, *The Boostless Bootstrap: Amplitudes without Lorentz boosts*, *JHEP* **12** (2020) 198 [*Erratum ibid.* **04** (2022) 023] [arXiv:2007.00027] [INSPIRE].

[52] T. Grall and S. Melville, *Positivity bounds without boosts: New constraints on low energy effective field theories from the UV*, *Phys. Rev. D* **105** (2022) L121301 [arXiv:2102.05683] [INSPIRE].

[53] Z.-Z. Du, C. Zhang and S.-Y. Zhou, *Triple crossing positivity bounds for multi-field theories*, *JHEP* **12** (2021) 115 [arXiv:2111.01169] [INSPIRE].

[54] A. Martin, *Unitarity and high-energy behavior of scattering amplitudes*, *Phys. Rev.* **129** (1963) 1432 [INSPIRE].

[55] J. Maldacena, S.H. Shenker and D. Stanford, *A bound on chaos*, *JHEP* **08** (2016) 106 [arXiv:1503.01409] [INSPIRE].

[56] D. Chandorkar, S.D. Chowdhury, S. Kundu and S. Minwalla, *Bounds on Regge growth of flat space scattering from bounds on chaos*, *JHEP* **05** (2021) 143 [arXiv:2102.03122] [INSPIRE].

[57] K. Häring and A. Zhiboedov, *Gravitational Regge bounds*, arXiv:2202.08280 [INSPIRE].

[58] M. Gell-Mann, M.L. Goldberger and W.E. Thirring, *Use of causality conditions in quantum theory*, *Phys. Rev.* **95** (1954) 1612 [INSPIRE].

[59] C. De Lacroix, H. Erbin and A. Sen, *Analyticity and Crossing Symmetry of Superstring Loop Amplitudes*, *JHEP* **05** (2019) 139 [arXiv:1810.07197] [INSPIRE].

[60] S. Mizera, *Crossing symmetry in the planar limit*, *Phys. Rev. D* **104** (2021) 045003 [arXiv:2104.12776] [INSPIRE].

[61] C.-H. Fu, J.-C. Lee, C.-I. Tan and Y. Yang, *BCFW Deformation and Regge Limit*, arXiv:1305.7442 [INSPIRE].

[62] S.D. Chowdhury et al., *Classifying and constraining local four photon and four graviton S-matrices*, *JHEP* **02** (2020) 114 [arXiv:1910.14392] [INSPIRE].

[63] S. Caron-Huot, *Analyticity in Spin in Conformal Theories*, *JHEP* **09** (2017) 078 [arXiv:1703.00278] [INSPIRE].

[64] D. Simmons-Duffin, D. Stanford and E. Witten, *A spacetime derivation of the Lorentzian OPE inversion formula*, *JHEP* **07** (2018) 085 [arXiv:1711.03816] [INSPIRE].

[65] P. Kravchuk and D. Simmons-Duffin, *Light-ray operators in conformal field theory*, *JHEP* **11** (2018) 102 [arXiv:1805.00098] [INSPIRE].

[66] B. Bellazzini, C. Cheung and G.N. Remmen, *Quantum Gravity Constraints from Unitarity and Analyticity*, *Phys. Rev. D* **93** (2016) 064076 [arXiv:1509.00851] [INSPIRE].

[67] N. Arkani-Hamed, Y.-T. Huang, J.-Y. Liu and G.N. Remmen, *Causality, unitarity, and the weak gravity conjecture*, *JHEP* **03** (2022) 083 [arXiv:2109.13937] [INSPIRE].

[68] D. Simmons-Duffin, *A Semidefinite Program Solver for the Conformal Bootstrap*, *JHEP* **06** (2015) 174 [arXiv:1502.02033] [INSPIRE].

[69] W. Landry and D. Simmons-Duffin, *Scaling the semidefinite program solver SDPB*, arXiv:1909.09745 [INSPIRE].

[70] C. de Rham, A.J. Tolley and J. Zhang, *Causality Constraints on Gravitational Effective Field Theories*, *Phys. Rev. Lett.* **128** (2022) 131102 [arXiv:2112.05054] [INSPIRE].

[71] L.-Y. Chiang et al., *(Non)-projective bounds on gravitational EFT*, arXiv:2201.07177 [INSPIRE].

[72] ATLAS collaboration, *Search for new phenomena in dijet events using $37\,fb^{-1}$ of pp collision data collected at $\sqrt{s} = 13\,TeV$ with the ATLAS detector*, *Phys. Rev. D* **96** (2017) 052004 [arXiv:1703.09127] [INSPIRE].

[73] CMS collaboration, *Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13\,TeV$ and constraints on dark matter mediators and other new particles*, *JHEP* **08** (2018) 130 [arXiv:1806.00843] [INSPIRE].

[74] N. Sennett et al., *Gravitational-Wave Constraints on an Effective Field-Theory Extension of General Relativity*, *Phys. Rev. D* **102** (2020) 044056 [`arXiv:1912.09917`] [INSPIRE].

[75] Z. Carson and K. Yagi, *Asymptotically flat, parameterized black hole metric preserving Kerr symmetries*, *Phys. Rev. D* **101** (2020) 084030 [`arXiv:2002.01028`] [INSPIRE].

[76] ATLAS collaboration, *Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector*, *JHEP* **01** (2018) 126 [`arXiv:1711.03301`] [INSPIRE].

[77] CMS collaboration, *Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **97** (2018) 092005 [`arXiv:1712.02345`] [INSPIRE].

[78] N.E.J. Bjerrum-Bohr et al., *Bending of Light in Quantum Gravity*, *Phys. Rev. Lett.* **114** (2015) 061301 [`arXiv:1410.7590`] [INSPIRE].

[79] G. Dvali, *Black Holes and Large N Species Solution to the Hierarchy Problem*, *Fortsch. Phys.* **58** (2010) 528 [`arXiv:0706.2050`] [INSPIRE].

[80] C. Cheung and G.N. Remmen, *Positive Signs in Massive Gravity*, *JHEP* **04** (2016) 002 [`arXiv:1601.04068`] [INSPIRE].

[81] N. Afkhami-Jeddi, S. Kundu and A. Tajdini, *A Bound on Massive Higher Spin Particles*, *JHEP* **04** (2019) 056 [`arXiv:1811.01952`] [INSPIRE].

[82] J. Ellis, *TikZ-Feynman: Feynman diagrams with TikZ*, *Comput. Phys. Commun.* **210** (2017) 103 [`arXiv:1601.05437`] [INSPIRE].

[83] N. Arkani-Hamed, T.-C. Huang and Y.-T. Huang, *Scattering amplitudes for all masses and spins*, *JHEP* **11** (2021) 070 [`arXiv:1709.04891`] [INSPIRE].

[84] G. Blekherman, G. Smith and M. Velasco, *Sums of squares and varieties of minimal degree*, *Journal of the American Mathematical Society* **29** (2015) 893.

[85] D. Eisenbud and J. Harris, *On varieties of minimal degree (a centennial account)*, in *Algebraic Geometry — Bowdoin 1985, Part 1*, Proceedings of Symposia in Pure Mathematics **46**, American Mathematical Society, Providence, U.S.A (1987) [`DOI:10.1090/pspum/046.1/927946`].