On the uniqueness and consistency of scattering amplitudes

Laurentiu Rodina

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Department of

Physics

Adviser: Nima Arkani-Hamed

June 2017

Abstract

In this dissertation, we study constraints imposed by locality, unitarity, gauge invariance, the Adler zero, and constructability (scaling under BCFW shifts).

In the first part we study scattering amplitudes as the unique mathematical objects which can satisfy various combinations of such principles. In all cases we find that locality and unitarity may be derived from gauge invariance (for Yang-Mills and General Relativity) or from the Adler zero (for the non-linear sigma model and the Dirac-Born-Infeld model), together with mild assumptions on the singularity structure and mass dimension. We also conjecture that constructability and locality together imply gauge invariance, hence also unitarity. All claims are proved through a soft expansion, and in the process we end re-deriving the well-known leading soft theorems for all four theories. Unlike other proofs of these theorems, we do not assume any form of factorization (unitarity).

In the second part we show how tensions arising between gauge invariance (as encoded by spinor helicity variables in four dimensions), locality, unitarity and constructability give rise to various physical properties. These include high-spin no-go theorems, the equivalence principle, and the emergence of supersymmetry from spin 3/2 particles. We also complete the fully on-shell constructability proof of gravity amplitudes, by showing that the improved "bonus" behavior of gravity under BCFW shifts is a simple consequence of Bose symmetry.

Acknowledgements

I first of all thank my advisor Nima Arkani-Hamed for his support, encouragement, for sharing his deep insights on a great variety of things, and for teaching me to always ask the zeroth order questions first. This made possible yet new basic results on tree level scattering amplitudes, where one might have thought there is nothing more to uncover.

I thank David McGady and Jaroslav Trnka for collaborating on parts of this thesis, and Song He and Yu-tin Huang for many great discussions during their visits in Princeton. I owe many thanks to Herman Verlinde for being a reader of this dissertation, and also for all his help and support throughout my time at Princeton. I thank Igor Klebanov and Ed Groth for serving on my FPO committee.

I would also like to thank Anastasia Volovich and Marcus Spradlin for introducing me to the wonderful world of scattering amplitudes, and also all the people who guided my path in physics, especially Ioana Stoica.

I greatly thank my family for their support and love.

To my loving fiancée.

Contents

Chapter 1

Introduction

Scattering amplitudes lie at the heart of quantum field theory, and at the intersection between theory and experiment. We are lucky that scattering amplitudes also posses a rare and valuable property: they are sufficiently simple to compute for many theories, and at the same time they are rich enough to uncover new symmetries and structures, completely hidden by the Lagrangian formulation. In the last few decades the study of amplitudes has greatly developed both computationally and conceptually, transforming scattering amplitudes into table-top experiments for theoretical physics. Much of this progress is due to the 1986 discovery of Parke and Taylor that the very complicated Feynman diagram expansion for gluon scattering amplitudes can be simplified to an amazingly concise formula [1]:

$$A(1^-, 2^-, 3^+, \ldots, n^+) = \frac{\langle 12 \rangle^4}{\langle 12 \rangle \langle 23 \rangle \ldots \langle n-1 n \rangle} \qquad (1.0.1)$$

The existence of this very simple formula eventually lead to the current perspective that scattering amplitudes may have a different origin, or that at least some simpler methods for computing them should exist, based not on Feynman diagrams and Lagrangians, but rather on physical principles and symmetries.

One such approach, which formed the power-house of the modern S-matrix program, is the on-shell BCFW recursion [2]. Here the locality (pole structure) and unitarity (factorization on those poles) are used to recursively build amplitudes. When certain conditions are met (ie., the theory is constructible), the BCFW recursion expresses an amplitude as:

$$A_n(1, 2, \ldots, n) = \sum_{\{L,R\}} \frac{A_{i+1}^L(1, 2, \ldots, i, P) \times A_{n-i+1}^R(-P, i+1, \ldots, n)}{P^2} \qquad (1.0.2)$$

This computational method is immensely more efficient than Feynman diagrams, especially when used with spinor helicity variables in four dimensions, and can be applied to a wide variety of theories. Further probing this approach also helped reveal many new surprising properties of scattering amplitudes, such as the Yangian invariance of $\mathcal{N} = 4$ Yang-Mills amplitudes [3], the polytope interpretation of amplitudes [4], and ultimately the Amplituhedron [5].

In parallel, another remarkable development by Kawai, Lewellen, and Tye in the same year was using string theory to show that graviton amplitudes are tightly connected to gluon amplitudes. Although the theories and Lagrangians describing General Relativity and Yang-Mills look completely different, their scattering amplitudes actually follow a simple relationship: Gravity = (Yang-Mills)^2. To be precise, graviton amplitudes can be expressed in terms of gluon amplitudes through the KLT relations [6] as:

$$M_3(1, 2, 3) = A_3(1, 2, 3)^2$$

$$M_4(1, 2, 3, 4) = s_{12} A_4(1, 2, 3, 4) A_4(1, 2, 4, 3) \qquad (1.0.3)$$

$$M_5(1, 2, 3, 4, 5) = s_{12} s_{34} A_5(1, 2, 3, 4, 5) A_5(2, 1, 4, 3, 5)$$

$$+ s_{13} s_{24} A_5(1, 3, 2, 4, 5) A_5(3, 1, 4, 2, 5)$$

and so on. Further refinement of this idea lead to the BCJ duality [7], or color-kinematic duality, which was also extended to loop level. Using this method high loop gravity calculations become possible for the first time, and revealed very surprising cancellations happening in $\mathcal{N} = 8$ supergravity, leading to a conjecture of its perturbative finitness [8].

Yet another avenue that has recently been revived is that of soft theorems. Initially, Weinberg showed that considerations of gauge invariance, locality and unitarity, when a particle is taken soft, $p_i \rightarrow 0$, fix QED and GR amplitudes to take the form:

$$A_{n+1} \rightarrow S_0 A_n \qquad (1.0.4)$$

In turn, this implies both charge conservation and the equivalence principle [9]. Since then, the theorems have been extended to sub-leading order [10], as well as to Yang-Mills and scalar theories [11]. Most recently, the newly found subleading graviton theorem was found to have implications for black-hole information [12].

Other developments include generalized unitarity [13], scattering equations [14], twistor string theory [15], pure spinors methods [16], to name a few. The common theme of these recent developments, forming the modern S-matrix program, has been moving away from the Lagrangian formulation, and putting basic principles and symmetries in the forefront. The goal of this dissertation is to establish the constraining power and interdependence of various such principles (whether physical or purely mathematical), which we define in the next section.

1.1   Five basic principles of scattering amplitudes

Locality   Broadly speaking, locality means that an object can only be influenced by its immediate surroundings. In position space that means that a Lagrangian can only be a function of fields (and their derivatives) at a single point in spacetime.

For scattering amplitudes in momentum space, locality imposes constraints on the singularity structure, and comes in two versions. The weak version only requires singularities to have a form $1/(\sum_i p_i)^2$. Full locality requires these singularities to correspond to propagators of tree graphs, with momentum conservation holding at each vertex.

Unitarity For scattering amplitudes unitarity is actually dependent on locality: on any singularity, the amplitude must factorize into two parts, both identifiable as lower point amplitudes. Both locality and unitarity are of course manifest in Feynman diagrams.

Gauge invariance At the level of amplitudes, gauge invariance is the property that the amplitude vanishes when sending any $e_i \to p_i$. Diffeomorphism invariance for gravity can also be understood in this way, by separating the two indices of the polarization tensor $e^{\mu\nu} \to e^\mu e^\nu$. In the Feynman diagram approach this property is not manifest - individual diagrams are not gauge invariant, as easily seen from the mass dimension of the numerators. For example, a Yang-Mills amplitude has numerators with mass dimension $n-2$, while gauge invariance in n particles requires a mass dimension of n.

Adler zero The Adler zero [17] is in some ways the equivalent of gauge invariance for theories containing Goldstone bosons. This states that amplitudes must vanish when we take one of the bosons to be soft, $p_i \to 0$. Amplitudes in different theories vanish at different rates. NLSM amplitudes vanish as $\mathcal{O}(z)$, DBI as $\mathcal{O}(z^2)$, and galileon amplitudes as $\mathcal{O}(z^3)$ [18]-[19]. Like gauge invariance, this property in only present in the complete sum of diagrams, again due to simple mass dimension considerations.

Constructability In the usual literature, constructability means that an arbitrary n-point amplitude can be constructed from lower point ones through on-shell recursions, typically BCFW (see for example [20, 21]). For this to be possible a few properties must be satisfied. First, under an $[i, j\rangle$ BCFW shift, which schematically sends some $p_i \to p_i + zq$ and $p_j \to p_j - zq$, the amplitude must vanish for $z \to \infty$. Then, if the amplitude only has simple poles (is local), through the Cauchy theorem it can be rebuilt solely from the residues on those poles. Finally, if the amplitude is unitary, those residues are just products of lower point amplitudes. It very nicely rounds up (and puts to very practical use) the previous principles.

In the first part of the dissertation, we will take constructability in a more abstract sense, and use it to refer only to scaling under BCFW shifts, for lack of a better name of this property. We will see that in this very restricted sense, constructability somehow manages to incorporate gauge invariance.

1.2 Overview of this thesis

In the first part of this thesis, we take an orthogonal approach to the recent on-shell perspective, and propose that gauge invariance is more than a cumbersome tool whose only role is to make locality and unitarity manifest.

In chapters 1 and 2, based on work in collaboration with Nima Arkani-Hamed and Jaroslav Trnka, [22, 23], we prove that Yang-Mills and General Relativity amplitudes are uniquely fixed by gauge invariance in $n-1$ particles, and very mild assumptions on the singularity structure and mass dimension. Specifically, we allow only singularities of the form $1/(\sum_i p_i)^2$, with consecutive momenta in the case of Yang-Mills. It follows that both their locality (cubic propagator structure) and unitarity (factorization) are emergent properties. This very strange phenomenon is also shown to hold for the

scalar non-linear sigma model and Dirac-Born-Infeld amplitudes, by replacing gauge invariance with the Adler zero condition, again for $n - 1$ of the particles.

Next, in chapter 3, based on [24] we extend these ideas even further, and show that locality and correct BCFW behavior in a sufficient number of shifts are also enough to fix the Yang-Mills amplitude. This is accomplished by introducing a new BCFW shift compatible with polarization vectors, which can also be used to recursively build the full amplitude.

In all cases described above, the proofs are carried out through soft expansions, enabling the use of a very simple inductive argument. In the process we end up re-deriving the known leading soft theorems for YM, gravity, NLSM and DBI.

Besides the surprising conceptual implications of both locality and unitarity emerging from gauge invariance, the uniqueness results have two main implications. First, they settle the long-standing conjecture that if an object has the right singularity structure and factorizes correctly, it must be the scattering amplitude. This has been the main method for checking the validity of new results, since direct comparisons with Feynman diagrams are not always feasible. Now, if gauge invariance is present (and it is in most formalisms, like those involving spinor helicity variables, or the Cachazo-He-Yuan scattering equations [14]), simply checking the presence of the correct singularity structure is sufficient to validate any expression. Similarly, it allows a more transparent understanding of the Bern-Carrasco-Johansson gravity squaring relations [7]. Second, these results suggest that a new mathematical definition of the amplitude might exist, complementing the recent amplituhedron program [5], where the goal is to see both locality and unitarity emerge from geometric principles.

In the second part of the dissertation we return to the 4 dimensional on-shell approach, where gauge invariance is automatically included (and therefore ignored)

in the power of spinor helicity variables. In doing so manifest locality and unitarity are lost, and we turn to deriving physical constraints from manually imposing locality and unitarity, as well as constructability.

In chapter 4, based on [25] with David McGady, we show that such constraints can be sharply expressed by using a very simple counting argument. It turns out that simply specifying the helicities of the participating particles is enough to fix the mass dimension of the denominator for any 4-point massless scattering amplitude. Then locality and unitarity force the [mass dimension]^2 to give the number of different factorization channels. The high-spin no go theorems, the emergence of supersymmetry and supergravity from spin 3/2 particles, as well as the unique structure of self-interacting spin 1 and spin 2 particles, all become simple consequences of counting the allowed number of factorization channels.

In chapter 5, based on [26] with David McGady, we investigate the origin of the mysterious "bonus" behavior of gravity amplitudes under BCFW shifts. As mentioned above, BCFW shifts, which roughly deform two momenta as $p_1 \to p_1 + zq$ and $p_2 \to p_2 - zq$, enable the use of recursion relations to construct amplitudes, as long as $A(z) \propto \mathcal{O}(z^{-1})$. However, it has been long known that gravity manifests a stronger $\mathcal{O}(z^{-2})$ behavior, which is hidden by the Lagrangian, but necessary for on-shell consistency. We show that this behavior is a direct consequence of permutation invariance between gravitons, and that the non-adjacent shifts in Yang-Mills theory can be understood in a similar way. This also completes the fully on-shell proof of gravity constructability, initiated in [21].

Chapter 2

Locality and unitarity from singularities and gauge invariance

2.1 Gauge Redundancy

The importance of gauge invariance in our description of physics can hardly be overstated, but the fundamental status of "gauge symmetry" has evolved considerably over the decades. While many older textbooks rhapsodize about the beauty of gauge symmetry, and wax eloquent on how "it fully determines interactions from symmetry principles", from a modern point of view gauge invariance can also be thought of as by itself an empty statement. Indeed *any* theory can be made gauge-invariant by the "Stuckelberg trick"–elevating gauge-transformation parameters to fields–with the "special" gauge invariant theories distinguished only by realizing the gauge symmetry with the fewest number of degrees of freedom.

Instead of gauge symmetry we speak of gauge "redundancy" as a convenient but not necessarily fundamental way of describing the local physics of Yang-Mills and gravity theories. Indeed in the sophisticated setting of quantum field theories and string theories at strong coupling we have seen the crucial importance of understand-

ing gauge symmetries as "redundancies"–for instance, in the famous gauge-gravity duality, it is silly to ask "where is the gauge symmetry?" in the bulk or "where is general covariance" on the boundary; these are merely two differently-redundant descriptions of the same physical system.

If gauge "symmetries" are merely redundancies, why have they been so useful? We can see the utility of gauge-redundancy [27] in the down-to-earth setting of scattering processes for elementary particles even at weak coupling, where we encounter a peculiarity in the Poincare transformation properties of scattering amplitudes. When the momenta of particles are transformed, the amplitude transforms according to the *little group.* Thus e.g. in four dimensions under a Lorentz transformation Λ the amplitude picks up phases $e^{ih\theta(\Lambda,p)}$ for each massless leg of momentum p helicity h, and an $SO(3)$ rotation on the massive particles. On the other hand, the standard formalism of field theory, the amplitudes are computed using Feynman diagrams, which give us "Feynman amplitudes" that are not the real amplitudes, but are instead Lorentz tensors. We contract them with *polarization vectors* to get the actual amplitudes– the polarization vectors are supposed to transform as "bi-fundamentals" under the Lorentz and little groups. For massive particles of any spin, there is a canonical way of associating polarization vectors with given spin states. But this is impossible for massless particles. Say for massless spin 1, we associate $\epsilon_\mu^\pm(p)$ with the $\pm$ polarizations of photons: the ϵ_μ do not transform as vectors under the Lorentz group. Indeed consider Lorentz transformations Λ that map p into itself $(\Lambda p)_\mu = p_\mu$. Then, it is trivial to see that $(\Lambda\epsilon)$ does not equal ϵ in general, rather we find $(\Lambda\epsilon)_\mu = \epsilon_\mu + \alpha(p)p_\mu$. Thus the polarization vector itself does not transform properly as a four-vector, only the full equivalence class $\{\epsilon_\mu|\epsilon_\mu \sim \epsilon_\mu + \alpha(p)p_\mu\}$ is invariant. These are all the "gauge-equivalent" polarization vectors. And so, for the amplitude obtained by contracting with ϵ's to be Lorentz-invariant, we must have that under replacing $\epsilon_\mu \to p_\mu$ the amplitude vanishes; i.e. we must satisfy the "on-shell Ward identity' $p^\mu M_{\mu\ldots} = 0$.

In order to guarantee that the Lorentz tensors $M_{\mu_1 \cdots \mu_n}$ arising from Feynman diagrams from a Lagrangian satisfy this on-shell Ward-identity, the Lagrangian must be carefully chosen to have an (often non-linearly completed) gauge-invariance, which is then gauge-fixed. From the modern point of view, then, gauge symmetry is merely a useful redundancy for describing the physics of interacting massless particles of spin 1 or 2, tied to the specific formalism of Feynman diagrams, that makes locality and unitarity as manifest as possible.

But over the past few decades, we have seen entirely different formalisms for computing scattering amplitudes not tied to this formalism, and here gauge redundancy makes no appearance whatsoever. Instead of polarization vectors that only redundantly describe massless particle states, we can use spinor-helicity variables λ_a, $\widetilde{\lambda}_a$ for the a'th particle, with momentum $p_a^{\alpha\dot\alpha} = \lambda_a^\alpha \tilde\lambda^{\dot\alpha}$. The $\lambda, \tilde\lambda$'s *do* transform cleanly as bi-fundamentals under the Lorentz and little groups; under a Lorentz transformation Λ that maps $(\Lambda p) = p$, we have $\lambda \to t\lambda, \tilde\lambda \to t^{-1}\tilde\lambda$. Thus while the description of amplitude using polarization vectors is gauge-redundant, the amplitude is directly a function of spinor-helicity variables, with the helicities encoded in its behavior under rescaling $M(t_a\lambda_a, t_a^{-1}\tilde\lambda_a) = t_a^{-2h_a} M(\lambda_a, \tilde\lambda_a)$.

With this invariant description of the fundamental symmetries and kinematics of amplitudes at hand, it becomes possible to pursue entirely new strategies for determining the amplitudes. In a first stage, one can speak of a modern incarnation of the S-matrix program, where the fundamental physics of locality and unitarity are imposed to determine the amplitudes from first principles. This has allowed the computation of amplitudes in an enormous range of theories, from Yang-Mills and gravity to goldstone bosons, revealing stunning simplicity and deep new mathematical structures that are completely hidden in the usual, gauge-redundant Feynman diagram formalism. Conversely and more ambitiously, these developments suggest that what we think of as "scattering amplitudes from local evolution in spacetime"

might fundamentally be something entirely different: instead of merely *exploiting* locality and unitarity to determine the amplitudes, we seek "scattering amplitudes" as the answer to very different natural mathematical questions, and only later *discover* that the results are local and unitary. Carrying this program out in full generality for all interesting theories would likely shed powerful new light on a deeper origin for both space-time and quantum mechanics itself.

A step in this direction has been taken with the discovery of the "Amplituhedron" [28], a geometric object generalizing plane polygons to higher-dimensional spaces, whose "volume" computes scattering amplitudes for maximally supersymmetric four-dimensional theories in the planar limit (in particular giving tree-level gluon scattering amplitudes for the real theory of strong interactions relevant for particle collisions at the LHC). In this example we can see concretely how the usual rules of spacetime and quantum mechanics emerge from more primitive principles.

2.2   Role Reversal

In this letter, we will explore aspects of locality and unitarity from a point of view entirely orthogonal to these recent developments. As emphasized above, much of the explosion of progress in understanding scattering amplitudes has taken place precisely by eschewing any reference to gauge-redundancy, and working directly with the physical on-shell amplitudes. Here we instead return to the requirement of on-shell gauge invariance as primary, and consider rational functions built out of polarization vectors and momenta, without making any reference to an underlying Lagrangian, Feynman rules or diagrams of any kind. Surprisingly, we find that with mild restrictions on the form of functions we consider, the requirement of on-shell gauge-invariance alone uniquely fixes the functions to match the tree amplitudes of Yang-Mills theory for spin one and gravity for spin two. There is a similar story determining the amplitudes

11

for goldstone bosons of the non-linear sigma model and the Dirac-Born-Infeld action, where the requirement of on-shell gauge invariance is replaced by an appropriate vanishing of amplitudes in soft-limits.

Suppose that we are handed a rational function of momenta and polarization vectors. What constraints determine this function to correspond to "scattering amplitudes"? One might imagine that both locality and unitarity are crucially needed for this purpose. In other words, we have to assume that this function has only *simple* poles when the sum of a subset $\mathcal{S}$ of the momenta $P_{\mathcal{S}}^{\mu} = \sum_{i \subset \mathcal{S}} p_i^{\mu}$ goes on-shell i.e. the only singularities look like $\sim 1/P_{\mathcal{S}}^2$, *and* that the function factorizes on the poles into the product of lower-point objects on the left and right, with an extra intermediate line. Note that locality and unitarity are intertwined in an interesting way. Factorization on *simple* poles guarantees that (in Lorentzian signature with the Feynman $i\epsilon$'s included) the imaginary part of amplitudes correspond to particle production. But factorization also implies that the singularities must be associated with a graph structure: sitting on a factorization channel, we can seek further singularities to deeper channels, but the longest sequence of poles we can encounter in this way all correspond to the $(n-3)$ propagators of some cubic graph.

The expectation that both locality and unitarity are needed to fix the form of the amplitude comes from our direct familiarity with simple theories of scalars, like ϕ^3 or ϕ^4 theory. If we only impose poles when $P^2 \to 0$ and the usual mass dimensions of amplitudes associated with, say, ϕ^3 theory, nothing forbids the presence of various trivially "illegal" terms of the form e.g. for $n = 5$

$$\frac{1}{(p_1 + p_2)^2 (p_2 + p_3)^2}, \frac{1}{((p_1 + p_2)^2)^2} \tag{2.2.1}$$

The first term has legal simple poles, but in overlapping channels in a way that never arises from Feynman diagrams; thus while at the coarsest level its singularities are

"local poles", it does not correspond to any local spacetime process. The second does not suffer from overlapping poles but has double poles. We can choose to also enforce locality by declaring that our functions can only have the poles corresponding to cubic graphs. If we again imagine objects with the mass dimension corresponding to a ϕ^3 theory, we would get a sum over cubic graphs Γ with some numerical coefficient n_Γ:

$$\sum_\Gamma \frac{n_\Gamma}{D_\Gamma} \tag{2.2.2}$$

where D_Γ is the product of the propagators of the cubic graph Γ. This expression corresponds to the amplitude only if the coefficients n_Γ are all equal but this is obviously not an automatic consequence of our rules. We must demand unitarity– factorization into products of lower amplitudes–to force all the n_Γ to be equal.

Our central claim in this note is that while locality and unitarity must be imposed to determine amplitudes for garden-variety scalar theories, *much* less than this is needed to uniquely fix the function to be "the amplitude" for gauge theories and gravity. In fact, we conjecture that simply specifying that the only singularities occur when the sum of a subset of momenta goes on-shell $P^2 \to 0$, together usual power-counting (which also enforces *non-trivial* gauge invariance) uniquely fixes the function! We will sketch the essential ideas in this note, a more detailed exposition of our proof and other related results can be found in [23].[1]

To begin with, we can enforce only locality, in the form of the location of singularities of the amplitudes. This tells us to only look at functions whose singularities are (powers of) propagator poles appearing in cubic graphs, as we did above in the scalar case. But we don't demand unitarity: we don't ask the poles to be simple, and we don't demand that the function factorizes on the poles. We find that instead the leading non-trivial gauge-invariants with the singularities of cubic graphs

---

[1]Other observations about the surprisingly restrictive power of on-shell gauge invariance have also been made in [29, 30].

are unique in both Yang-Mills and gravity, and give us the amplitude! The necessity of simple poles and factorization–and thus unitarity–therefore follows from locality and gauge invariance. We will sketch a straightforward proof of this fact, which begins by showing that given the poles of cubic graphs, gauge-invariance alone (with no assumption about factorization) fixes the structure of the soft limit of any expressions to reproduce the usual Weinberg soft theorems [31].

But we are making a stronger conjecture, that even the structure of singularities associated with cubic graphs need not be enforced: we need only assume that the singularities occur when $P_S^2 \to 0$. We will consider functions that have at most degree β singularities of this form, that is our most general ansatz is

$$\sum_{\{S_1, \cdots, S_\beta\}} \frac{N_i^{(\alpha)}}{P_{S_1}^2 \cdots P_{S_\beta}^2} \tag{2.2.3}$$

Here $N(\alpha)_i$ is a polynomial in the momenta (and linear in all the polarization vectors), with a total of α momenta in the numerator.

We will now only ask for this expression to be on-shell gauge-invariant. Clearly, even if there are no singularities at all i.e. $\beta = 0$, we can of course trivially build gauge-invariants simply starting with linearized field strengths $f_{\mu\nu} = p_\mu \epsilon_\nu - p_\nu \epsilon_\mu$, and contracting n of these together in any way we like. This would give us a number $\alpha \geq n$ of momenta. These correspond to the amplitudes from local higher-dimension operators. We will thus ask that our functions are *non-trivially* gauge invariant, and so we will demand that $\alpha < n$; the only hope for making gauge-invariants now crucially must use momentum-conservation $p_1^\mu + \cdots p_n^\mu = 0$. It is then easy to see that this is impossible with purely local expressions, and we must allow poles, so $\beta > 0$. Our precise claim is that it is impossible to build a gauge-invariant unless $\alpha = (n-2)$ for gauge-theory and $\alpha = 2(n-2)$ for gravity, and furthermore this is impossible for all $\beta = 0, 1, \cdots, (n-4)$, but that there is a unique gauge-invariant at $\beta = (n-3)$.

14

In fact just demanding gauge-invariance in $(n-1)$ legs suffices to fix the function. This unique object picks out the singularities from cubic graphs and factorizes on poles; locality and unitarity arise from singularities and gauge invariance. While we haven't yet completed a proof of this conjecture, we will show how it works in some non-trivial examples which suggest the structure a proof should take.

All of these statements are made in general D spacetime dimensions: we are simply working with lorentz-invariants multilinear in the polarization vectors, of the form $(\epsilon_i \cdot \epsilon_j)$, $(\epsilon_i \cdot p_j)$ and $(p_i \cdot p_j)$, only satisfying the relations $p_i^2 = 0$, $\epsilon_i \cdot p_i = 0$, and momentum-conservation $\sum_i p_i^\mu = 0$. Thus the gauge-invariance checks where we demand the vanishing of the amplitude upon substituting $\epsilon_i^\mu \to p_i^\mu$ can only follow from these relations.

2.3   Locality, Unitarity and Gauge Invariance

Let us begin by focusing on the tree-level scattering amplitudes in Yang-Mills theory; we will later summarize the precisely analogous statements for gravity. The group structure of gluon amplitudes can be stripped off in trace factors $\mathcal{A}_n = \sum_{\sigma/Z} \mathrm{Tr}(T^{\sigma_1} T^{\sigma_2} \ldots T^{\sigma_n}) A_n(123 \ldots n)$, where A_n is an ordered amplitude which is a gauge invariant cyclic object. All poles in A_n are local cyclic factors, $P_{ij}^2 = (p_i + p_{i+1} + \ldots p_j)^2$. And on these poles A_n factorizes as a product of two ordered amplitudes,

$$\lim_{P^2 \to 0} A_n = \sum_h A_L^{(h_L)} \frac{1}{P^2} A_R^{(h_R)} \tag{2.3.1}$$

were we sum over all internal degrees of freedom h. In practice we can replace the helicity sum over the intermediate line I by $\sum_h \epsilon_{I,h}^\mu \epsilon_{I,-h}^\nu \to \eta^{\mu\nu}$; this differs from the true polarization sum by terms proportional to p_I^μ, p_I^ν which vanish by gauge invariance, when contracted into the lower-point amplitude factors.

The cyclic amplitude A_n can by calculated using color-ordered Feynman rules. For each cubic graph Γ we get

$$D_n^{(\Gamma)} = \frac{N_n^{(\Gamma)}(\epsilon_i, p_j)}{P_{\sigma_1}^2 P_{\sigma_2}^2 \dots P_{\sigma_{n-3}}^2} \qquad (2.3.2)$$

where all the factors $P_{\sigma_a}^2$ in the denominator come from Feynman propagators of cubic diagrams. The numerator is a polynomial in all polarization vectors ϵ_i and $n-2$ momenta p_j, which appear in the scalar products $(p_i \cdot p_j)$, $(p_i \cdot \epsilon_j)$ and $(\epsilon_i \cdot \epsilon_j)$. For diagrams with four point vertices we get fewer than $n-3$ propagators, but they can be also put (non-uniquely) in the cubic form by multiplying both numerator and denominator by some P^2.

Feynman diagrams are designed to make locality and unitarity as manifest as possible, but gauge-invariance is not manifest diagram-by-diagram: we have to sum over all Feynman diagrams to get a gauge invariant expression. The tension between locality, unitarity and gauge invariance is vividly seen in the four-particle amplitude. The color ordered amplitude A_4 is a sum of three Feynman diagrams, schematically written as (ignoring all indices)

$$A_4 \sim \frac{(\epsilon \cdot p)(\epsilon \cdot p)(\epsilon \cdot \epsilon)}{s} + \frac{(\epsilon \cdot p)(\epsilon \cdot p)(\epsilon \cdot \epsilon)}{t} + (\epsilon \cdot \epsilon)(\epsilon \cdot \epsilon) \qquad (2.3.3)$$

Only the sum of all three terms is gauge invariant which can be made manifest once we write A_4 as

$$A_4 \sim \frac{F^4}{st} \qquad (2.3.4)$$

where the numerator is just a (color-ordered) local amplitude. This expression is trivially gauge-invariant but we don't have *manifest* locality and unitarity: we see the *product* of st in the denominator. It is impossible to write the amplitude as a sum over s and t channels in a way that is both Lorentz invariant and gauge invariant.

16

2.4 Unitarity From Locality and Gauge Invariance

Elaborating further on the example from the previous section we can ask what is the minimal number of momenta p_j we need in order to make a polynomial in $\epsilon_1, \dots, \epsilon_n$ gauge invariant. Obviously, if we take n momenta we can always build gauge invariant tensors $\epsilon^{[\mu}p^{\nu]} = (p_\mu \epsilon_\nu - p_\nu \epsilon_\mu)$ and contract n of them in an arbitrary way. But can we make a non-trivial invariant, one which has fewer than n momenta? This has a chance of being possible because of momentum conservation. The first non-trivial case is with $n-2$ momenta p_i. It is easy to see that if we demand the object is just a polynomial we find there exist no gauge invariant, but if we allow poles we certainly find at least one solution which is the amplitude A_n, written as a sum of Feynman diagrams.

Let us now consider a set of all cubic graphs with cyclic ordering of external legs and for each of them we write an expression $\widetilde{D}_n^{(\Gamma)}$ of the form (2.3.2) where the poles in the denominator are dictated by the internal lines of the given graph. Unlike in Feynman diagrams we do not demand the numerator comes from Feynman rules, and therefore we are not imposing unitarity; more invariantly we are not asking the amplitude to actually factorize on factorization channels. Instead we take $N_n^{(\Gamma)}$ to be an arbitrary polynomial of degree $n-2$ in momenta p_j and n polarization vectors ϵ_i. For four points we get,

$$N_4 = \alpha_1(\epsilon_1 \cdot p_2)(\epsilon_2 \cdot p_3)(\epsilon_3 \cdot \epsilon_4) + \alpha_2(\epsilon_1 \cdot p_2)(\epsilon_3 \cdot p_4)(\epsilon_2 \cdot \epsilon_4)$$
$$+ \alpha_3(p_1 \cdot p_2)(\epsilon_1 \cdot \epsilon_2)(\epsilon_3 \cdot \epsilon_4) + \dots \tag{2.4.1}$$

We of course impose $(\epsilon_i \cdot p_i) = 0$ and momentum conservation $\sum_i p_i = 0$. The same structure of numerator is used for the s and t channels, but with different parameters

$\alpha_k^{(1)}$ and $\alpha_k^{(2)}$. Now we consider a sum of all expressions associated with graphs Γ,

$$\widetilde{A}_n = \sum_\Gamma \widetilde{D}_n^{(\Gamma)} \tag{2.4.2}$$

and impose gauge invariance in $n-1$ legs. We claim that this specifies a unique expression which is an n point tree-level amplitude, $\widetilde{A}_n = A_n$. Note we do not have to even check gauge invariance in the nth leg, everything is fixed already.

The proof goes as follows: First, it is easy to show that polynomials with at most k factors of $e_i.p_j$ per term can be gauge invariant in at most k particles, when momentum conservation is not used. If we allow momentum conservation, the same is true, but only for $k < n - 2$. This statement can then be extended from polynomials to general functions $B(p^k)$ containing poles, but which have only k momenta in the numerators. A stronger statement also holds for tensors $B^{\mu\nu}(p^k)$ containing poles, which can be gauge invariant in at most k particles for $k < n - 1$.

Next, we assume inductively that $\widetilde{A}_n(p^{n-2}) = A_n(p^{n-2})$ is unique for the n particle case. Now we take the expansion (2.4.2) for $n + 1$ particles and go to the soft limit of one of the particles, $p_{n+1} \equiv q \to 0$. It is easy to show that gauge invariance in the soft particle requires the leading divergent term to be proportional to the Weinberg soft factor,

$$\widetilde{A}_{n+1} = \left(\frac{\epsilon \cdot p_1}{q \cdot p_1} - \frac{\epsilon \cdot p_n}{q \cdot p_n} \right) B_n(p^{n-2}) + \mathcal{O}(1) \tag{2.4.3}$$

where B_n is the gauge invariant function in n legs with $n-2$ powers of momenta which is $B_n = A_n$ by induction. The important point here is that the soft limit is controlled by the usual Weinberg soft factor *purely as a consequence of gauge invariance*, without any further assumption about factorization.

Now since both A_{n+1} and $\widetilde{A}_{n+1}$ have equal leading pieces, we can consider instead the object $M_{n+1} = \widetilde{A}_{n+1} - A_{n+1}$, which has vanishing leading piece. Then the sub-

leading piece in the soft limit receives no contributions from its leading order[2], and has the form

$$\delta_1 M_{n+1} = \frac{\epsilon^\mu q^\nu B_n^{\mu\nu}(p^{n-2})}{q \cdot p_1} + \frac{\epsilon^\mu q^\nu \overline{B}_n^{\mu\nu}(p^{n-2})}{q \cdot p_n} \qquad (2.4.4)$$

where we omitted the terms with double poles which are directly ruled out by gauge invariance. The tensors $B_n^{\mu\nu}$, $\overline{B}_n^{\mu\nu}$ have $k = n - 2$ and therefore are ruled out. At higher orders, terms in the soft limit always have a form $\delta_p M_{n+1} \sim X_n^{\mu\nu}(p^{n-2})$ for some tensor X, which is then ruled out by gauge invariance, and all these terms must vanish. This implies $M_{n+1} = 0$ to all orders in the soft expansion, so $\widetilde{A}_{n+1} = A_{n+1}$.

It is interesting that in these arguments, it suffices to check gauge invariance only in $(n - 1)$ legs to uniquely fix the answer! This observation explains why the object factorizes on poles. We'd like to determine what our unique gauge-invariant looks like on a factorization channel. Since there is already a unique gauge invariant, only checking invariance on $(n-1)$ legs, we can take "left" and "right" gauge invariants ignoring gauge-invariance on the intermediate line; gluing together these unique objects then gives us something that is gauge-invariant in all n legs, and therefore must match the unique n-pt gauge-invariant on this channel. This shows that gauge-invariants factorize on poles, allowing us to see the emergence of unitarity very directly.

2.5 Locality from Gauge Invariance

We showed that unitarity is a derived property of gluon amplitudes if we demand only locality and gauge invariance. But we can go even further and even remove the requirement of locality. We again consider a sum of terms (2.4.2) but now we give up on the assumption that individual terms (2.3.2) have poles which correspond to cubic diagrams. We just consider any cyclic poles $P_{ij}^2 = (p_i + p_{i+1} + \cdots + p_j)^2$, and even allow powers $(P_{ij}^2)^\#$. The only assumption is that the total number of poles in

---

[2]See also [32] for a discussion of this issue

the denominator (the degree of P^2) is $n-3$. For example, for the $n=5$ case we allow terms of the form

$$\frac{N_5^{(1)}}{s_{12}^2}, \quad \frac{N_5^{(2)}}{s_{12}s_{23}} \tag{2.5.1}$$

While the double (or higher) poles can come from Lagrangians with non-canonical kinetic term, the second term can not be associated with any local interaction as it does not correspond to any "diagram" of particle scattering. The numerator is still an arbitrary polynomial in n polarization vectors ϵ_i and $n-2$ momenta p_j.

We now conjecture that if we simply impose gauge invariance on the general sum of all possible terms with $n-3$ cyclic poles (2.4.2) the only solution is again the n point scattering amplitude A_n. There are no other solutions and all numerators for terms like (3.2.10) are forced to vanish as a consequence of gauge invariance. We have directly checked this conjecture by brute force up to $n=5$, which is already highly non-trivial. We will also give an analytic proof of the analog of this conjecture for the non-linear sigma model up to at $n=8$ points (which is the NLSM analog of $n=5$ for YM theory) below.

2.6 Gravity and BCJ

The story for gravitons is essentially identical. In particular, we can again consider cubic graphs with no ordering of external legs. For each graph we associate an expression (2.3.2). The denominator contains $n-3$ propagators consistent with the cubic graph, but the poles are not restricted to be cyclic sums of momenta anymore. The numerator $N_n^{(\Gamma)}$ is polynomial of degree $2(n-2)$ in momenta p_i, and it also depends on n polarization tensors $\epsilon_{\mu\nu} = \epsilon_\mu \epsilon_\nu$. For example the four point amplitude has a schematic form,

$$\frac{(\epsilon \cdot p)^4 (\epsilon \cdot \epsilon)^2}{s} + \frac{(\epsilon \cdot p)^4 (\epsilon \cdot \epsilon)^2}{t} + \frac{(\epsilon \cdot p)^4 (\epsilon \cdot \epsilon)^2}{u} + (\epsilon \cdot p)^2 (\epsilon \cdot \epsilon)^6 \tag{2.6.1}$$

20

For each diagram we write an ansatz for the numerator $N_n^{(\Gamma)}$ with free parameters and impose the gauge invariance condition in $n-1$ external legs. As a result, we get an unique solution which is the graviton amplitude. Therefore, unitarity emerges from locality and gauge invariance in the same sense as in the Yang-Mills case. The proof is analogous to that case too, using the soft limit to show uniqueness.

For gravity we can also make the stronger statement, that even locality emerges from gauge invariance. Assuming only the $n-3$ poles in the denominator, including multiple poles and with no reference to cubic graphs, we claim that there still emerges a unique gauge-invariant, the amplitude itself.

We can also go back to the gluon case and consider now all possible P^2 poles, not just the ones with cyclic momenta, maintaining the non-trivial power-counting in the numerator, ie. $n-2$ momenta p_j, but now choosing $n-3$ of all possible cubic graph poles. Now imposing gauge invariance we conjecture $(n-2)!$ solutions corresponding to different cyclic orderings of Yang-Mills amplitudes, modulo the relations following from the $U(1)$ decoupling and KK relations.

The uniqueness of gauge-invariants also gives a natural proof for the BCJ relation [33] between the Yang-Mills and gravity amplitudes. If we write the Yang-Mills amplitude in the BCJ form, then for each cubic graph the kinematical numerators satisfy $N_s + N_t = N_u$, if the color factors satisfy the same Jacobi identity, $c_s + c_t = c_u$. Then the gravity amplitude is given by the simple replacement of the color factor by one more power of the kinematical numerator,

$$A_n^{(YM)} = \sum_\Gamma \frac{N_\Gamma c_\Gamma}{D_\Gamma} \quad \rightarrow \quad A_n^{(GR)} = \sum_\Gamma \frac{N_\Gamma^2}{D_\Gamma} \qquad (2.6.2)$$

The reason is very simple. Under a gauge variation, the N_Γ change by some Δ_Γ; the invariance of the full amplitude $\sum_\Gamma c_\Gamma \Delta_\Gamma / D_\Gamma = 0$ can then only be ensured by the Jacobi relations satisfied by c_Γ. But if we now replace c_Γ with some kinematical factor

N_Γ which satisfies the same identities, the gravity-gauge invariance check follows in exactly the same way as for YM. Thus the object with $c_\Gamma \to N_\Gamma$ is a gravitational gauge-invariant with $2(n-2)$ powers of momenta in the numerator; since this object is unique it gives the gravity amplitude.

2.7 Gauge-Invariance $\to$ Soft Limits and Goldstone Theories

We have seen that gauge and gravity amplitudes are much more special than garden-variety scalar theories. But of course famously there is also no good reason to have light scalars to begin with, unless they are goldstone bosons whose mass is appropriately protected by shift symmetries. Recent investigations revisiting some classic aspects of goldstone scattering amplitudes have revealed precisely what is special about these goldstone theories from a purely on-shell perspective. In the case of the non-linear sigma model, soft limit behavior in the form of the Adler zero [17] supplements unitarity and locality in certain cases to completely fix the tree-level S-matrix [34, 35, 36]. In particular, we can ask what is the minimally derivatively coupled theory whose amplitudes have vanishing soft-limit, $A_n = 0$ for $p_j \to 0$. The answer appears to be non-linear sigma model (NLSM). If we demand the quadratic vanishing, $A_n = \mathcal{O}(p^2)$ this uniquely specifies the Dirac-Born-Infeld (DBI) theory and $A_n = \mathcal{O}(p^3)$ gives a special Galileon [34, 37]. The soft limit behavior was then used in the recursion relations to reconstruct the amplitudes in these theories, supplementing locality and unitarity.

In the spirit of our previous statements we can make similar claims for these theories. Like in Yang-Mills we can strip the flavor factor in the NLSM [38] and consider cyclically ordered amplitudes A_n. Now the individual Feynman diagrams

are quartic diagrams $\mathcal{Q}$, and we can write an expression for each of them

$$D_n^{(\mathcal{Q})} = \frac{N_n^{(\mathcal{Q})}(p_j)}{P_1^2 P_2^2 \ldots P_{n/2-2}^2} \tag{2.7.1}$$

Then the poles in (2.7.1) are cyclically ordered and the numerator is degree $n-2$ in momenta. Imposing the soft-limit vanishing then requires summing over all Feynman diagrams as only the full amplitude has this property. Now we forget the Lagrangian and consider a general numerator,

$$N_n^{(\mathcal{Q})}(p_j) = \sum_k \alpha_k \Delta_k \tag{2.7.2}$$

where Δ_k is the product of $n/2 - 1$ terms of the form $s_{ij} = (p_i \cdot p_j)$. Note that if we allow one more s_{ij} factor in the numerator then we could always write an expression which manifestly vanishes in the soft limit. For example, for the six point case one of the Feynman diagrams is

$$D = \frac{(s_{12} + s_{23})(s_{45} + s_{56})}{s_{123}} \tag{2.7.3}$$

and it does not vanish in all soft limits, and no other numerator with two s_{ij} does. If we replace the numerator by $s_{12}s_{34}s_{56}$ we would have manifestly each diagram vanishing.

Now we ask that the numerator $N_n^{(\mathcal{Q})}$ is an arbitrary linear combination of products of $n/2 - 1$ factors s_{ij} with free parameters. The statement is that imposing the soft limit vanishing in $n-1$ legs fixes all coefficients completely and there is a unique expression, which is the n-pt tree-level amplitude in NLSM. The proof for this statement uses the double soft limit where two of the momenta go to zero. In that case the amplitude does not vanish but rather gives a finite expression, and in some sense it is an analogue of the Weinberg soft factor for the Yang-Mills and gravity. One can

then prove the statement in a similar way to the soft limit argument for gluons and gravitons. The soft limit and locality then imply unitarity of goldstone amplitudes.

The stronger claim is that we do not have to consider quartic graphs, but rather take any expression with $n/2 - 2$ factors in the denominator (allowing double poles, and non-diagrammatic combinations of poles) and at most $n/2 - 1$ terms s_{ij} in the numerator. Then only imposing the soft limit again fixes the result uniquely, and we can see both locality and unitarity arising vanishing in the soft limit. We will give evidence for this in the next section.

We can make analogous claims for the DBI and special Galileon. Now the power-counting of the numerator is $n - 2$, resp. $3n/2 - 3$ factors s_{ij} and $n/2 - 2$ poles in the numerator. We have to consider all quartic graphs with no ordering. Imposing the $\mathcal{O}(p^2)$, resp. $\mathcal{O}(p^3)$ vanishing in the soft limit of $n - 1$ legs fixes the numerators uniquely to be the numerators of corresponding Feynman diagrams, and we get the amplitude as the only soft limit (with certain degree) vanishing object. The stronger statement again removes the requirement of single poles associated with quartic diagrams and we only consider the correct number $n/2 - 2$ poles P^2 with no restrictions.

2.8   Evidence for the strong conjecture

We have made two distinct claims: the first is that locality (in the form of the pole structures of cubic graphs), together with numerator power-counting, uniquely fixes the result when gauge-invariance/soft limits are imposed.

But we have also made a more striking conjecture, where we don't even impose locality, only ask that singularities are made of up to $(n - 3)$ "P^2" poles, without asking that these poles are associated with graphs at all. And we demand the non-trivial number of momenta in the numerator which prohibits a trivial solution such as the powers of $(p_\mu \epsilon_\nu - p_\nu \epsilon_\mu)^n$ for gauge invariance of spin s, and the products of

$\prod_j s^{\sigma}_{j\,j+1}$ for the soft limit $\mathcal{O}(p^{\sigma})$. The claim is that the result is still unique: locality and unitarity arise from (non-trivial) gauge-invariance/soft limits.

We do not currently have a proof of this conjecture, but if it is true, we suspect that the mechanism behind it should be the same for gluons, gravitons and goldstone theories. We will therefore confirm the conjecture for the case of the NLSM here; the way the graph structure emerges "out of thin air" is already quite suggestive for what might be going on at general n.

For the 6pt NLSM amplitude there are only three poles s_{123}, s_{234}, s_{345} which can appear in the denominator, and there is always just one of such factor. Therefore, locality here is directly imposed as we do not have any double poles or overlapping poles. The first non-trivial case to test our conjecture is then 8pt. The general ansatz is given by five different types of terms with two poles,

$$\widetilde{A}_8 = \frac{N_8^{(a)}}{s_{123}s_{456}} + \frac{N_8^{(b)}}{s_{123}s_{567}} + \frac{N_8^{(c)}}{s_{123}s_{345}} + \frac{N_8^{(d)}}{s_{123}s_{234}} + \frac{N_8^{(e)}}{s_{123}^2} \qquad (2.8.1)$$

Only the first two terms correspond to quartic graphs as the last three terms are not in the Feynman expansion as they violate locality. We will show that just soft limit vanishing forces $N_8^{(c)} = N_8^{(d)} = N_8^{(e)} = 0$, or more precisely we can rewrite everything in terms of first two terms. Then we are left with the terms associated with quartic graphs only when the double soft limit argument can be applied to fix the answer uniquely to be an 8pt amplitude in NLSM.

The numerator is degree 6 in momenta, ie. degree 3 in invariants s_{ij}. It is convenient to use the cyclic basis,

$$s_{12}, \ldots, s_{81}, s_{123}, \ldots, s_{812}, s_{1234}, \ldots, s_{4567}. \qquad (2.8.2)$$

In the soft limit $p_8 \to 0$ these terms go to the 7pt cyclic basis made of s_{12}, s_{123} and the cyclic images. Two of the terms $s_{78}, s_{81} \to 0$, the other nine terms s_{23}, s_{34}, s_{45},

25

s_{56}, s_{234}, s_{345}, s_{456}, $s_{2345} \to s_{671}$, $s_{3456} \to s_{712}$ stay the unique basis elements, while the remaining become degenerate (2-to-1 map).

$$s_{12}, s_{812} \to s_{12}; \ s_{1234}, s_{567} \to s_{567}; \ s_{4567}, s_{123} \to s_{123}$$

$$s_{678}, s_{67} \to s_{67}; \ s_{812}, s_{12} \to s_{12} \tag{2.8.3}$$

Analogously for all other soft limits. It is very easy to show that $N_8^{(e)} = 0$, or the corresponding term can be absorbed into first two terms in case we cancel one power of s_{123}. In the proof we critically use the relation between 7pt and 8pt basis of kinematical invariants. In the soft limit $\widetilde{A}_8 = 0$ and therefore the 7pt expression must vanish identically. Because the last term in (2.8.1) is the only term with that particular double pole in s_{123} we apply different soft limits and demand that this term cancels or becomes degenerate with other terms.

For soft limits in momenta p_2, p_5, p_6, p_7 the term s_{123} is a unique basis elements also in the 7pt basis, and there is no way how to cancel a double pole. Therefore, the numerator $N_8^{(e)}$ must simply vanish in all these four soft limits. For other four soft limits s_{123} becomes degenerate with other kinematical invariants: with s_{23} for $p_1 \to 0$, with s_{12} for $p_3 \to 0$, with s_{1234} for $p_4 \to 0$ and with s_{4567} for $p_8 \to 0$. Therefore, either the numerator again vanishes or it is proportional to s_{23} for $p_1 \to 0$, s_{12} for $p_3 \to 0$ etc. It is easy to show that there is no such numerator $N_8^{(e)}$ which satisfies all these constraints. As a result, $N_8^{(e)} \sim s_{123}$ killing a double pole and being degenerate with first two terms. The proofs for vanishing of $N_8^{(d)}$ and $N_8^{(c)}$ have the same flavor.

It is likely that this sort of reasoning can be generalized to any n. Ultimately all statements about numerators $N_n^{(p)}$ are translated to properties of basis elements of kinematical invariants. The set of all cyclic invariants form a basis for any n and they smoothly go to $n-1$ point basis in the soft limit. It seems plausible that some clever bookkeeping along the above lines can be done to prove the statement in general.

2.9 Outlook

There are a number of avenues for further exploration suggested by this work. One obvious question has to do with space-time dimensionality: all of our analyses find objects that would be gauge-invariant in *any* number of dimensions. But could there be functions that are only gauge-invariant in a specific dimension d? In a specific space-time dimensionality d, there are further "gram determinant" conditions that arise from the fact that any number $k > (d+1)$ of momenta/vectors must be linearly dependent. Could it be that there are objects whose gauge-variation is proportional to gram determinant conditions in a specific number of dimensions, and so would be gauge-invariant in those dimensions but not otherwise? It is overwhelmingly likely that the answer to this question is "no"–*all* non-trivial gauge-invariants are the ones that exist in all numbers of dimensions. This is certainly a fascinating feature of amplitudes for fundamental (parity-invariant) theories like Yang-Mills and General Relativity, and it would be nice to prove it directly along the lines of this note.

Resolving this issue about dimension-dependence would also settle a natural question posed by thinking about scattering amplitudes, the pursuit of which led directly to this work. Suppose we are given all the scattering amplitudes in some theory. These are "boundary observables in flat space"–they can be measured by experiments not in the interior of spacetime, but out at infinity. Given only this information, how could we discover the description of the physics in terms of local quantum evolution through the interior of the space-time? We can ask this question already at tree-level. We often say that the "locality" and "unitarity" of amplitudes is reflected in the location of their poles (locality) and the factorization of the poles on these poles (unitarity). But what we colloquially mean by these concepts is much more detailed that this–we would like to see that the amplitudes arise from local rules of particles moving and colliding at points in spacetime. Thus most prosaically, given the final amplitudes,

we would like to know: how could we discover that they can be computed by the Feynman diagrams of a local theory?

As a trivial first step, we have to compare apples to apples. As we stressed in our introductory remarks the amplitudes are not Lorentz tensors, but Feynman amplitudes are. It is however trivial to associate on-shell amplitudes, written in terms of spinor-helicity variables, in terms of gauge-invariant Lorentz tensors. We can "rationalize" any expression for amplitudes so that the poles are mandelstam invariants. Then e.g. an amplitude for a $-$ helicity spin 1 particle would have weight two in it's λ and is thus of the form $\lambda_\alpha \lambda_\beta T^{\alpha\beta}$ for some tensor $T^{\alpha\beta}$. But we can associate $\lambda_\alpha \lambda_\beta$ directly with a gauge invariant field strength; indeed defining $F^{\pm}_{\mu\nu} = F_{\mu\nu} \pm i\tilde{F}_{\mu\nu}$, we have that $F^{-}_{\alpha\dot\alpha\beta\dot\beta} = \lambda_\alpha \lambda_\beta \epsilon_{\dot\alpha\dot\beta}$, and similarly $F^{+}_{\alpha\dot\alpha\beta\dot\beta} = \epsilon_{\alpha\beta}\tilde\lambda_{\dot\alpha}\tilde\lambda_{\dot\beta}$. These expressions can be computed uniformly from $F_{\mu\nu} = p_\mu \epsilon_\nu - p_\nu \epsilon_\mu$, making the familiar choices for the helicity polarization vectors $\epsilon^{+}_{\alpha\dot\alpha} = \tilde\lambda_{\dot\alpha}\xi_\alpha / \langle\lambda\xi\rangle$ for any reference ξ, and similarly for $\epsilon^{-}_{\alpha\dot\alpha} = \lambda_\alpha \tilde\xi_{\dot\alpha} / [\tilde\lambda\tilde\xi]$.

In this way, any Lorentz-invariant expression of appropriate helicity weights can be associated with a gauge-invariant expression made out of field strengths. Consider for instance the 4 particle Parke-Taylor amplitude for $(1^- 2^- 3^+ 4^+)$, we can associate this with a gauge-invariant expression as

$$\frac{\langle 12\rangle^2 [34]^2}{st} \to \frac{(F^{-}_{1\mu\nu}F_2^{-\mu\nu})(F^{+}_{3\alpha\beta}F_4^{+\alpha\beta})}{st} \qquad (2.9.1)$$

The right-hand side is non-vanishing only for this helicity configuration, so by summing over all helicities we construct a gauge-invariant expression that matches the amplitude on all helicity configurations. It is amusing to carry out the exercise of constructing the Feynman amplitude from on-shell helicity amplitudes for the 4 gluon and 4 graviton amplitudes. Of course the individual helicity expressions explicitly involve $\epsilon_{\mu\nu\alpha\beta}$ and thus make sense only in four dimensions. But since the theory is

parity invariant, after summing over all helicities all terms with an odd number of ϵ's cancel. Terms with an even number of ϵ's can be turned into expressions only involving the metric $\eta_{\mu\nu}$ using the fact that $\epsilon_{abcd}\epsilon_{xyzw} = (\eta_{ax}\eta_{by}\eta_{cz}\eta_{dw} \pm \text{permutations})$. This gives us

$$A_4 = \frac{(F_{\mu\nu}F^{\mu\nu})^2 - 2(F_{\mu\nu}F^{\nu\alpha}F_{\alpha\beta}F^{\beta\mu})}{st} \tag{2.9.2}$$

(where we expand $F_{\mu\nu} = F_{1\mu\nu} + \cdots F_{4\mu\nu}$ and we only keep terms linear in the polarization vectors). By construction matches the amplitude in four dimensions, but is a gauge-invariant expression in any number of dimensions.

Thus, starting from on-shell amplitudes, we can trivially construct the gauge-invariant Lorentz-tensor $M_{\mu_1\cdots\mu_n}$ that matches all the helicity amplitudes making appropriate choices for the polarization vectors. We can now ask, how can we see that this object can be computed from local Feynman diagrams? Most naively, one might have expected that the critical properties of the amplitude–location of poles and factorization–would be needed in order to establish this fact. But we now see that much less is needed; even our weakest (and proven) statement about unique gauge-invariants requirements already shows that $M_{\mu_1\cdots\mu_n}$ can be computed from Feynman diagrams. The reason is simply that Feynman diagrams give us a local gauge-invariant, and this object is unique, thus it must match the $M_{\mu_1\cdots\mu_n}$ constructed from scattering amplitudes!

Note that however that complete proof of this fact requires us to show that there are no gauge-invariants special to any particular dimension. [It is trivial to see that no "gram determinant" conditions are possible for four-points, so the above expression is indeed valid in general D dimensions, but this is no longer immediately true starting at five points].

Provided that the absence of dimension-specific invariants can be established, we have found a simple conceptual understanding of a fact that has resisted a transparent understanding for many years. There is an apparently straightforward proof

that "amplitudes that factorize properly" must match feynman diagrams, by using Cauchy's theorem and the BCFW deformation to show that that if functions have the same singularities they must be equal [?]. However, this famously needs a proof of an absence of poles at infinity on the Feynman diagram side, which can only be shown by a relatively indirect argument far afield from on-shell physics [39].

The uniqueness of gauge invariance implies further properties of the S-matrices. In particular, it is trivial to show that it is impossible to have interactions of higher spin particles. The standard modern S-matrix argument relies on the factorization of the 4 point amplitude which is inconsistent in all three channels for $s > 2$ [40, 25]. In our story we do not use factorization, gauge invariance alone implies that any amplitude with cubic graphs needs a Weinberg soft factor, and that is impossible to construct for higher spins.

Our results also illuminate why the CHY construction [41] of YM and gravity amplitudes must match the correct answer, without any detailed analysis of the poles and factorization structure. We simply observe the poles of the CHY formula are local, and the expressions are gauge invariant, with the correct units to match the correct numerator power-counting.

As we have seen the uniqueness of gauge-invariants gives a one-line proof of the passage from color-kinematics satisfying forms of Yang-Mills amplitudes to a gravity amplitude; it would be satisfying if the act of building gauge invariants naturally led to the color-kinematic structure for Yang-Mills to begin with.

Beyond these issues, the main open problem is to prove (or disprove) the strong conjecture about the emergence of the graph structure. Also, we have only looked at trees. Obviously the story at loop level will be much more interesting, and we can't expect uniqueness to follow simply from gauge invariance on external legs, since the particles propagating in the loops will also matter.

It is also natural to conjecture that maximal SUSY, together with degree $(n-3)$ poles, gives same result: diagrams emerge and the result is unique. We know that there is a tension between SUSY and locality/unitarity which is quite similar to the case of gauge invariance. This idea has also been explored in the context of pure spinors (see for instance [16]), where gauge invariance and supersymmetry are merged into BRST invariance.

Finally, while the claims in this note are mathematically non-trivial and certainly have physical content, their ultimate physical significance is not clear. It is intriguing that locality and unitarity can be derived from a redundancy, inverting the usual logic leading to the need for gauge invariance. If this is more than a curiosity, it would be interesting to look for an abstract underlying system that gives rise to an effective description–either exactly or approximately–with a gauge redundancy, from which locality and unitarity emerge in the way we have seen here.

Chapter 3

Uniqueness from gauge invariance and the Adler zero

In this chapter we provide detailed proofs for some of the uniqueness results presented in Ref. [22]. We show that: (1) Yang-Mills and General Relativity amplitudes are completely determined by gauge invariance in $n-1$ particles, with minimal assumptions on the singularity structure; (2) scalar non-linear sigma model and Dirac-Born-Infeld amplitudes are fixed by imposing full locality and the Adler zero condition (vanishing in the single soft limit) on $n-1$ particles. We complete the proofs by showing uniqueness order by order in the single soft expansion for Yang-Mills and General Relativity, and the double soft expansion for NLSM and DBI. We further present evidence for a greater conjecture regarding Yang-Mills amplitudes, that a maximally constrained gauge invariance alone leads to both locality and unitarity, without any assumptions on the existence of singularities. In this case the solution is not unique, but a linear combination of amplitude numerators.

3.1 Introduction

Recently, in Ref. [22] it was conjectured that after only fixing the number and form of possible singularities, gauge invariance uniquely determines the Yang-Mills and gravity scattering amplitudes. It was also stated that the same is true for scalar theories like the non-linear sigma model (NLSM) and Dirac-Born-Infeld (DBI), when gauge invariance is replaced by vanishing in the single soft limit. Crucially, in all of these cases locality and unitarity are never assumed, and so arise automatically as a consequence of uniqueness. Here by locality we mean that poles correspond to propagators of cubic diagrams (for YM and GR), and quartic diagrams (for NLSM and DBI), while by unitarity we mean factorization on those poles. In some sense then we see the emergence of spacetime and local quantum interactions purely from gauge invariance. A similar result was presented in Ref. [42], that locality and vanishing under large BCFW shifts are also sufficient to completely fix the Yang-Mills amplitude. Beyond their conceptual implications, these uniqueness results have a very practical application: if a given expression can be verified to be gauge invariant and contain the correct singularity structure, it is now guaranteed to match the corresponding amplitude. This has many implications for a wide variety of recently developed formalisms, like BCFW recursion relations [2], the BCJ duality [7], or the CHY scattering equations [43]-[44], among others.

The proof used in this article also demonstrates a new powerful application of soft limits, as well as novel derivations of the well-known leading theorems [9][11]. Leading and subleading soft theorems have already proven very useful in a number of very surprising ways. Originally, they showed that charge conservation or the equivalence principle can be derived from S-matrix arguments [9]. More recently, the theorems were interpreted as consequences of new symmetries [45, 46], with further implications

for black-hole information [12]. They were also used for recursion relations for effective field theories [47].

The goal of this paper is three-fold. First, we present the full details of the argument used in Ref. [22] to prove the uniqueness claims for Yang-Mills, gravity, NLSM, and DBI, when locality is assumed. Second, we extend the argument to prove the conjecture that uniqueness still holds without assuming locality. And third, we make an even larger conjecture, that gauge invariance alone, with no assumptions on the presence of any singularities, is sufficient to imply both locality and unitarity.[1]

3.1.1   Assumptions and results

In all four theories, Yang-Mills, gravity, NLSM, and DBI, we start with an ansatz $B_n(p^k)$, based on our assumptions of the singularity structure and mass dimension. In general, this ansatz contains functions of the form:

$$B_n(p^k) \equiv \sum_i \frac{N_i(p^k)}{P_i} \,, \tag{3.1.1}$$

where the numerators $N(p^k)$ are general polynomials with k powers of momenta, and linear in some number of polarization vectors/tensors (for YM/GR). The denominators P_i can be any polynomial of $p_i.p_j$ factors. If we assume locality, it means we must restrict each P_i to be a product of simple poles, which can be associated to propagators of cubic (for YM and GR) or quartic (for NLSM and DBI) diagrams. That is, a local ansatz has a form:

$$B_n(p^k) \equiv \sum_{\text{diags. } i} \frac{N_i(p^k)}{\prod_{\alpha_i} P_{\alpha_i}^2} \,, \tag{3.1.2}$$

---

[1]This is similar to the results of [48, 49], where it was found that imposing full gauge invariance, while also allowing extra kinematical invariants as coefficients, leads to $(n-3)!$ independent solutions, forming the BCJ basis of Yang-Mills amplitudes.

with α_i corresponding to the channels of each diagram. As discussed in [7], it is always possible to put amplitudes in this cubic diagram form, by adding artificial propagators to the higher point vertex interactions. For YM and NLSM the diagrams are ordered, while for gravity and DBI they are not. We can relax locality by dropping the underlying diagram structure, allowing each term to have some number s of any singularities $P_{\mathcal{S}}^2 = (\sum_i p_i)^2$, with the p_i consecutive for YM and NLSM:

$$B_n(p^k) \equiv \sum_i \frac{N_i(p^k)}{P_{\mathcal{S}_1}^2 \dots P_{\mathcal{S}_s}^2} , \qquad (3.1.3)$$

Then the claim is that for the smallest s and k which admit solutions, the ansatz (3.1.3) is uniquely fixed by gauge invariance/vanishing in the single soft limit in $n-1$ particles. Concretely, these smallest values for s and k are:

- Yang-Mills: $s = n - 3$, $k = n - 2$

- Gravity: $s = n - 3$, $k = 2n - 4$

- NLSM: $s = n/2 - 2$, $k = n - 2$

- DBI: $s = n/2 - 2$, $k = 2n - 4$

In this article, we prove the following results:

1. Local Singularities + Gauge Invariance $\Rightarrow$ Locality + Unitarity

2. Locality + Adler zero $\Rightarrow$ Unitarity

The stronger version of claim number 2 for NLSM and DBI (that Local Singularities + Adler zero $\Rightarrow$ Locality + Unitarity) is less susceptible to the argument we use in this article, but a more direct approach was already presented in [22]. We also prove a stronger result for Yang-Mills, by allowing non-local singularities $(\sum_i a_i p_i)^2$, with some mild restrictions. Further, we conjecture that completely ignoring the

35

singularity structure, gauge invariance alone forces general polynomials to be linear combinations of amplitude numerators.

Surprisingly, imposing gauge invariance/vanishing in the soft limit for the nth particle is not required, and is automatic once the other $n-1$ constraints have been imposed. Without loss of generality we can take particle 3 to be this n^{th} particle, and we will always impose momentum conservation by expressing p_3 in terms of the other momenta. Tying the unneeded n^{th} constraint to momentum conservation ensures that we always avoid checks of the form $e_3 \rightarrow p_3 = -p_1 - p_2 - p_4 - \ldots$, which would complicate the analysis.

To begin, the above statements can be easily tested explicitly for a small number of particles. For Yang-Mills, at four points, we can only have terms with one pole, either $(p_1 + p_2)^2$ or $(p_1 + p_4)^2$. Then the most general term we can write down is a linear combination of 60 terms, of the form

$$M_4(p^2) = a_1 \frac{e_1.e_2 \, e_3.e_4 \, p_1.p_4}{p_1.p_2} + a_2 \frac{e_1.p_2 \, e_3.p_2 \, e_2.e_4}{p_1.p_2} + \ldots . \tag{3.1.4}$$

Imposing gauge invariance in particles 1, 2, and 4 forces the coefficients a_i to satisfy some linear equations with a unique solution, which turns out to be precisely the scattering amplitude of four gluons.

At five points, the most general non-local ansatz, where we only assume two cyclic singularities per term, contains some 7500 terms, and it can be checked that gauge invariance in four particles leads to the five point amplitude. It is indeed quite remarkable that gauge invariance is so constraining to produce a unique solution. Actually, it is even more remarkable that any solution exists at all! The amplitudes are the result of a striking conspiracy between the propagator structure and momentum conservation.

It is easy to make a gauge invariant in n particles by taking different contractions of $\prod_{i=1}^{n}(e_i^{\mu_i}p_i^{\nu_i} - e_i^{\nu_i}p_i^{\mu_i})$, but this requires a mass dimension of $[n]$. No single diagram has enough momenta in the numerator to accommodate this product, so different diagrams must cancel each other. But less obviously, without momentum conservation, such contractions will always contain terms with at least n factors of $e_i.p_j$. This is impossible to achieve even with several diagrams, since each diagram numerator can have at most $n-2$ such factors per term, while the denominators only contains $p_i.p_j$ factors. But with momentum conservation, together with a cubic propagator structure, it turns out that $n-2$ factors are sufficient, creating an object which satisfies more constraints than expected by simple counting. In fact, we will show that this structure is indeed very special. There is no non-trivial way of deforming or adding things to produce different solutions. Identical facts hold for the other theories as well. For example, the NLSM requires vanishing under $n-1$ particles, which naively would require $A \propto \mathcal{O}(p_i)$ for $n-1$ particles. Again, however, only at $k = n-2$, with momentum conservation and quartic diagrams we find an exception, which is the amplitude itself. The absence of solutions below this critical mass dimension is at the heart of the proof.

The basic strategy for the proof is the following. We start with an appropriate ansatz (local, non-local, etc), and we take single/double soft expansions. Then order by order we show that gauge invariance/the Adler zero condition uniquely fix the corresponding amplitudes. In this process we do not assume any form of the soft theorems, but we end up re-deriving the well known leading terms [9, 11]. Our approach does not directly provide the subleading terms (which for gravity have a particularly nice form [10]), but only proves their uniqueness.[2] The whole proof rests on showing that after the first non-vanishing order is fixed, none of the higher orders can produce independent solutions. The reason for this is that the subleading

---

[2]See also [32] for a very illuminating discussion on fixing the subleading terms.

37

orders must have a growing number of soft momenta in the numerator, leaving fewer momenta to satisfy the necessary requirements.

3.1.2   Organization of the article

In section 3.2, we begin by exploring the notion of constrained gauge invariance. We find a very simple proof that functions with at most $k < n - 2$ factors of momenta in the numerator can be gauge invariant in at most k particles, while the same is true for tensors with $k \leq n - 2$.

In section 3.3, we first prove a weaker version of our statement for YM and GR, by assuming locality. In section 3.4 the same argument is applied to the NLSM and DBI amplitudes, with gauge invariance replaced by the Adler zero condition.

In section 3.5, relaxing our assumptions on the underlying cubic diagrams, we instead consider a more general singularity structure. We only keep the requirement of (local) singularities of the form $(\sum_i p_i)^2$, and recover the unique amplitude, as long as the number of such singularities per term is $s = n - 3$. For fewer singularities there are no solutions, while for more the answer can always be factorized as $(\sum \text{poles}) \times$ (amplitude). This proves the conjecture originally made in Ref [22].

Finally, in section 3.6, we investigate the extent to which more general singularities can be used to fix the Yang-Mills amplitude. Completely non-local singularities of the form $(\sum_i a_i p_i)^2$, with some minor restrictions, are also shown to provide a unique solution. Trying to find a less arbitrary ansatz, we are lead to consider polynomials again, with no singularities at all. We conjecture that yet an even stronger statement can be made, namely that the smallest mass dimension polynomial that admits a solution is fixed to a linear combination of amplitude numerators, when gauge invariance in *all n* particles is imposed. The usual argument can be used to provide leading order evidence for this fact.

3.2 Constrained gauge invariance

3.2.1 Polynomials

Let $B(k)$ be a polynomial linear in polarization vectors, with at most k factors of dot products of the type $e_i.p_j$ in any given term. Let g be the total number of gauge invariance requirements, and $\Delta = g - k$ be the "excess" of gauge invariance requirements compared to the maximum number of $e.p$ factors. When appropriate, we will use the notation $\bar{e}_i$ to distinguish polarization vectors which are not used for gauge invariance. We wish to prove that, without momentum conservation, $B(k)$ can be gauge invariant in at most k particles, ie. satisfy at most $\Delta = 0$ constraints. Then we will prove that with momentum conservation the statement is still true, but only for $k < n - 2$.

No momentum conservation

Having no momentum conservation implies gauge invariants in particle i must be proportional to $G_i^{\mu\nu} = e_i^\mu p_i^\nu - e_i^\nu p_i^\nu$. Therefore the only way to obtain gauge invariants in k particles is with linear combinations of different contractions of products $\prod_i G_i^{\mu_i \nu_i}$. We will show that such expressions always contain at least one term with k factors of $e.p$.

By assumption there are always more e's than $e.p$'s, so at least one of the polarization vectors needed for gauge invariance will be in a factor $e.e$ or $e.\bar{e}$. Consider first as an example the following term in a polynomial with $k = 2$:

$$e_1.e_2 \ e_3.p \ \bar{e}_4.p \ p.p \, . \tag{3.2.1}$$

We wish to show that such a term cannot be gauge invariant in three particles, say particles 1, 2 and 3. We start with a polarization vector sitting in a $e.e$ factor, for

39

example e_1. To make a gauge invariant in particle 1, at least one of the $p's$ above must be a p_1, and a pair term must exist with e_1 and a p_1 switched. We can use either the p in the $e_3.p$ (or $\bar{e}_4$) factor, or one in the $p.p$ factor, to make the gauge invariants:

$$G_1 = (e_1.e_2)\,(e_3.p_1)\,\bar{e}_4.p\ p.p - (p_1.e_2)\,(e_3.e_1)\,\bar{e}_4.p\ p.p\,, \qquad (3.2.2)$$

or

$$G_1' = (e_1.e_2)\,e_3.p\ \bar{e}_4.p\ (p_1.p) - (p_1.e_2)\,e_3.p\ \bar{e}_4.p\ (e_1.p)\,. \qquad (3.2.3)$$

However, the second option leads to a term with four $e.p$ factors, contradicting our claim that just two factors are sufficient. The first option is fine, and so we can only use p's in $e.p$ factors for this mark-and-switch procedure.

Now consider the second term in G_1 above, and note that we ended up with another $e.e$ factor, namely $e_3.e_1$. Applying the same reasoning for gauge invariance in 3 forces us to fix the factor $\bar{e}_4.p$ to $\bar{e}_4.p_3$. Therefore the second piece of G_1 can form a gauge invariant in particle 3 in the pair:

$$G_3 = (p_1.e_2)\,(e_3.e_1)\,\bar{e}_4.p_3\ p.p - p_1.e_2\ p_3.e_1\ \bar{e}_4.e_3\ p.p\,. \qquad (3.2.4)$$

Now we do not need gauge invariance in 4, so this chain $1 \to 3 \to \bar{4}$ is finished. Note we do not care about making the first piece of G_1 gauge invariant in particle 3, we are only interested in finding some minimal constraints. Instead, we go back to (3.2.1) to check gauge invariance in the remaining particle 2. But the choices we made so far by imposing gauge invariance in 1 and 3 fixed both $e.p$ factors in this initial term to

$$e_1.e_2\ e_3.p_1\ \bar{e}_4.p_3\ p.p\,, \qquad (3.2.5)$$

so now there is no way to make it gauge invariant in 2, as all the allowed p's have been used up. Therefore the term (3.2.1) is not compatible with gauge invariance in $\{1,2,3\}$.

The general strategy is the same. New chains always start in the original term from $e.\bar{e}$ or $e.e$ factors, which are aways present by assumption. Next, for each jump we fix a p in an $e.p$ or $\bar{e}.p$ factor, which becomes unavailable for other gauge invariants. The chain ends when reaching an $\bar{e}.p$ factor, and a new chain is started from the original term, and so on. The process ends when all the chains have ended on $\bar{e}.p$ factors, which means gauge invariance is compatible with this counting argument, or when all the $e.p$'s have been marked and a chain is unable to continue, which means the term is ruled out.

In general, we said we need gauge invariance in $k + 1 = \#[\bar{e}.e] + 2\#[e.e] + \#[e.p]$ particles, but have only $k = \#[\bar{e}.p] + \#[e.p]$ factors. This means the difference between how many chains must start and how many can end is: $(\#[\bar{e}.e] + 2\#[e.e]) - \#[\bar{e}.p] = 1$. Therefore there is always at least one chain which cannot end, so all possible starting terms are ruled out. Then there is no way to make a polynomial $B(k)$ gauge invariant in $k + 1$ particles.

With momentum conservation

Now we consider the same type of polynomials from above, but on the support of momentum conservation, $B_n(k) \equiv B(k)\delta_n$. To impose momentum conservation explicitly, we can choose three particles (for example 2, 3, and 4), and use the following

relations:

$$p_3 = -\sum_{i \neq 3} p_i \,, \tag{3.2.6}$$

$$e_3.p_4 = -\sum_{i \neq 3,4} e_3.p_i \,, \tag{3.2.7}$$

$$p_2.p_4 = -\sum_{i,j \neq 3} p_i.p_j \equiv P_{24} \,. \tag{3.2.8}$$

This allows other ways of forming gauge invariance in particles 2, 3 and 4. For example, for particle 2 we can now have a new gauge invariant of the form

$$G_2^\mu = e_2.p_4 p_2^\mu - P_{24} e_2^\mu \,. \tag{3.2.9}$$

Such an expression avoids our previous argument: now both particles 2 and 4 can share the same $e_2.p_4$ factor above. When checking gauge invariance in at most $n-3$ particles this is not a problem, as we can always impose momentum conservation in such a way as to avoid these three special particles. However, for more than $n-3$ particles, at least one particle has to be affected by momentum conservation.

The worst case that we will need to prove is that $B_n(n-3)$ cannot be gauge invariant in $n-2$ particles. Without loss of generality, since we can change how we impose momentum conservation, we can leave out 3 and 4, and assume that these $n-2$ particles are $\{1, 2, 5, 6, \ldots, n\}$. Then $n-3$ of the particles can only form gauge invariants of the form $e_i^\mu p_i^\nu - e_i^\nu p_i^\nu$, while particle 2 allows gauge invariants of the form (3.2.9). Since we only have $k = n-3$ factors of $e.p$, the first $n-3$ constraints already fix all such factors. However, we are not checking gauge invariance in particle 4, so none of the factors will be fixed to $e.p_4$, more specifically to the $e_2.p_4$ needed in eq. (3.2.9). Therefore there is still no room to form the gauge invariant for particle 2.

This proves that a polynomial $B_n(k)$ with $k < n-2$ can be gauge invariant in at most k particles. It is easy to see that the last step of the argument fails for $k = n-2$.

42

In that case the counting allows for gauge invariance in not just $n-1$ particles, but all the way to n particles. This is of course what we should expect, since a polynomial $B_n(n-2)$ corresponds to the full amplitude numerator, and is gauge invariant in n particles.

3.2.2 Functions and tensors with singularities

The previous results for polynomials can be extended to functions with poles, such as those we initially introduced in eq. (3.1.1):

$$B_n(p^k) \equiv \sum_i \frac{N_i(p^k)}{P_i} \, .$$
(3.2.10)

Because the P_i are only functions of $p.p$ factors, a function with at most k momenta in the numerators can be expressed in terms of a polynomial with at most k factors of $e.p$:

$$B_n(p^k) = \frac{B_n(k)}{\prod_i P_i} \, .$$
(3.2.11)

This implies that a function $B_n(p^k)$ cannot be gauge invariant in $k+1$ particles, if $k < n-2$. This statement can be generalized to tensors $B_n^{\mu\nu}(p^k)$. We can write out the components of such a tensor:

$$B_n^{\mu\nu}(p^k) = \sum_{i,j} p_i^\mu p_j^\nu B_{ij}(p^{k-2}) + \sum_{i,j} p_i^\mu e_j^\nu C_{ij}(p^{k-1}) + \sum_{i,j} e_i^\mu p_j^\nu C'_{ij}(p^{k-1}) + \sum_{i \neq j} e_i^\mu e_j^\nu D_{ij}(p^k) \, ,$$
(3.2.12)

and determine what constraints each of the functions above must satisfy in order for $B_n^{\mu\nu}(p^k)$ to be gauge invariant in $k+1$ particles. We can treat each different type of function in order:

- $p_i^\mu p_j^\nu B_{ij}(p^{k-2})$: if we check gauge invariance in some particle m, with $m \neq i, j$, the prefactor remains unique, and none of the other terms in (3.2.12) may cancel against B_{ij}. This implies $B_{ij}(p^{k-2})$ itself must be gauge invariant in at least $k + 1 - |\{i, j\}| = k - 1$ particles so is ruled out, if $k - 2 < n - 2$.

- $e_i^\mu p_i^\nu C_{ii}(p^{k-1})$, or $\overline{e}_i^\mu p_j^\nu C_{ij}(p^{k-1})$: the same logic as before implies $C(p^{k-1})$ must be gauge invariant in k particles, so is also ruled out if $k - 1 < n - 2$.

- $e_i^\mu p_j^\nu C_{ij}$: can only form a gauge invariant in i together with a term $p_i^\mu p_j^\nu B_{ij}$, which was ruled out.

- $e_i^\mu e_j^\nu D_{ij}(p^k)$, or $e_i^\mu \overline{e}_j^\nu D_{ij}(p^k)$: under $e_i \to p_i$, D_{ij} can only form a gauge invariant with C_{ij}, which vanished, so this case is ruled out.

- $\overline{e}_i^\mu \overline{e}_j^\nu D_{ij}(p^k)$: is only ruled out for $k < n - 2$

To summarize, we obtain just three types of cases:

$$B_n(p^{k-2}), \text{ gauge invariant in } k - 1, \tag{3.2.13}$$

$$C_n(p^{k-1}), \text{ gauge invariant in } k, \tag{3.2.14}$$

$$D_n(p^k), \text{ gauge invariant in } k + 1 \tag{3.2.15}$$

all of which vanish for $k < n - 2$. For $k = n - 2$, the first two also vanish, but for the third we have an apparent contradiction, since we know that functions $D_n(p^{n-2})$ have sufficient momenta to satisfy gauge invariance in $n - 1$ particles. However, when $k = n - 2$, case (3.2.15) does not exist. The only tensor $B_n^{\mu\nu}(p^{n-2})$ that will show up in the actual proof has $\overline{G} = \{e_3\}$, so it does not contain a component $\overline{e}_i \overline{e}_j D_{ij}$.

Therefore, the tensors we are interested in cannot be gauge invariant in $k + 1$ particles for $k \leq n - 2$. In conclusion, so far we have shown that:

- functions $B_n(p^k)$ cannot satisfy $\Delta = 1$ constraints for $k < n - 2$

44

- tensors $B_n^{\mu\nu}(p^k)$ cannot satisfy $\Delta = 1$ constraints for $k \leq n - 2$

It turns out that these results can be generalized even further: we can take linear combinations of the above functions/tensors, with factors of $p_i.p_j$ as coefficients, and still the above statements hold. These results will be necessary for the following sections.

3.3 Unitarity from locality and gauge invariance

In this section we will consider local functions, as in eq. (3.1.2):

$$B_n(p^k) = \sum_{\text{diags. } i} \frac{N_i(p^k)}{\prod_{\alpha_i} P_{\alpha_i}^2} \, . \tag{3.3.1}$$

In the above notation the actual gluon amplitude $A_n(p^{n-2})$ is a subset of $B_n(p^{n-2})$, with $G = \{1, 2, 4, \ldots, n\}$ and $\overline{G} = \{3\}$, so $g = n-1$ and $\Delta = 1$. Now we wish to prove that A_{n+1} is uniquely fixed by gauge invariance in n particles, under the assumption that A_n is fixed by gauge invariance in $n - 1$ particles. Consider the most general $(n + 1)$-point local function M_{n+1}, and let $p_{n+1} = zq$. The Taylor series expansion around $z = 0$ is:

$$
\begin{aligned}
M_{n+1}\delta_{n+1} =& (z^{-1}M_{n+1}^{-1} + z^0 M_{n+1}^0 + \ldots)(\delta_n + zq.\delta_n' + \ldots) \\
=& z^{-1}M_{n+1}^{-1}\delta_n + z^0 \left(M_{n+1}^{-1}q.\delta_n' + M_{n+1}^0\delta_n \right) + \ldots \\
=& z^{-1}\mathcal{M}_{n+1}^{-1} + z^0 \mathcal{M}_{n+1}^0 + \ldots \, . \tag{3.3.2}
\end{aligned}
$$

First we must investigate the pole structure of a local function in this limit. There are two types of poles that can show up. First, there are q-poles which are singular. These correspond to diagrams of the type:
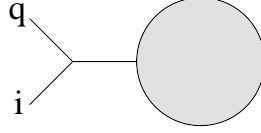
and can be written as:

$$\frac{N}{q.p_i \mathcal{P}_n(q)} = D_{n+1} \qquad (3.3.3)$$

with $i = 1$ or $i = n$ for Yang-Mills, because of ordering. In the limit $q \to 0$, we can factor out the $q.p_i$ pole and incorporate the remaining propagator structure into the lower point local function:

$$D_{n+1} \to \frac{1}{q.p_i} \frac{N}{\mathcal{P}_n(0)} = \frac{1}{q.p_i} B_n . \qquad (3.3.4)$$

Then there are non-singular poles, which appear when two propagators become equal in the $q \to 0$ limit:
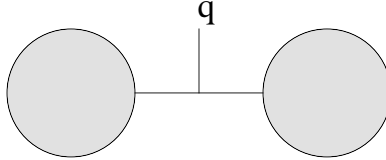
and can be written as:

$$D_{n+1} = \frac{N}{P_L(q)(p_1 + p_2 + \ldots + p_i)^2 (q + p_1 + p_2 + \ldots + p_i)^2 P_R(q)} , \qquad (3.3.5)$$

where for Yang-Mills $P_i^2 = (p_1 + p_2 + \ldots + p_i)^2$ contains only consecutive momenta up to particle i, $i = \overline{2, n-2}$. We will factor out one of the P_i^2's, and incorporate the other into the lower point local function B_n:

$$D_{n+1} \to \frac{N}{(p_1 + p_2 + \ldots + p_i)^2 \mathcal{P}_n(q)} = \frac{1}{(p_1 + p_2 + \ldots + p_i)^2} B_n . \qquad (3.3.6)$$

46

The argument by induction can be used precisely because of this factorization into the lower point local functions.

3.3.1 Yang-Mills

Leading order   The leading z^{-1} piece of the soft limit (3.3.2) can only come from q-pole terms. Using linearity in $e_{n+1} = e$, it can be written as:

$$M_{n+1}^{-1}(p^{n-1}) = \frac{e^{\mu} B_n^{\mu}(p^{n-1})}{q.p_1} + \frac{e^{\mu} C_n^{\mu}(p^{n-1})}{q.p_n}, \tag{3.3.7}$$

where B_n^{μ} and C_n^{μ} are local (vectors) at n-points. Next, gauge invariance in q requires $B_n^{\mu} = p_1^{\mu} B_n$, and $C_n^{\mu} = -p_n^{\mu} B_n$, where B_n is a local function at n points. Then the leading piece is:

$$M_{n+1}^{-1}(p^{n-1}) = \left(\frac{e.p_1}{q.p_1} - \frac{e.p_n}{q.p_n} \right) B_n(p^{n-2}). \tag{3.3.8}$$

By assumption, gauge invariance in the remaining $(n-1)$ particles uniquely fixes $B_n = A_n$, reproducing the well known Weinberg soft factor [9]. Note that unlike other methods of obtaining the soft term, we have not used any information on factorization, but only gauge invariance.

Now that the leading order is fixed, consider instead the function $B_{n+1} = M_{n+1} - A_{n+1}$. B_{n+1} is also local, and must be gauge invariant in n particles, but has a vanishing leading order. We will show that all higher orders in the soft expansion of B_{n+1} also vanish, implying that $M_{n+1} = A_{n+1}$. This procedure will be identical for gravity, NLSM and DBI.

Sub-leading order Because the leading order vanishes, the sub-leading piece is given only by:

$$B_{n+1}^0(p^{n-1}) = \sum_{i=1;n} \frac{e^\mu q^\nu B_{n;i}^{\mu\nu}(p^{n-2})}{q.p_i} + \sum_{i=2}^{n-2} \frac{e^\mu C_{n;i}^\mu(p^{n-1})}{P_i^2}, \qquad (3.3.9)$$

which includes both singular and non-singular pole parts. The non-singular pole terms are ruled out by gauge invariance in q, while the q-pole terms must be proportional to $e^{[\mu}q^{\nu]}$. Bringing everything under a common denominator we can write

$$B_{n+1}^0(p^{n-1}) \propto e^{[\mu}q^{\nu]} \left(q.p_n B_{n;1}^{\mu\nu}(p^{n-2}) + q.p_1 B_{n;n}^{\mu\nu}(p^{n-2}) \right) \equiv e^{[\mu}q^{\nu]} B'_n{}^{\mu\nu}(p^{n-2}), \quad (3.3.10)$$

where $B'_n{}^{\mu\nu}(p^{n-2})$ is a linear combination of tensors with $k = n - 2$, so is ruled out by requiring gauge invariance in $n - 1$ particles. Therefore $B_{n+1}^0 = 0$.

Sub-sub-leading order The sub-sub-leading piece is given by:

$$B_{n+1}^1(p^{n-1}) = e^{[\mu}q^{\nu]} \left(\sum_{i=1,n} \frac{q^\rho B_n^{\mu\nu\rho}(p^{n-3})}{q.p_i} + \sum_{i=2}^{n-2} \frac{C_n^{\mu\nu}(p^{n-2})}{P_i^2} \right). \qquad (3.3.11)$$

This time the non-singular pole terms are not ruled out just by gauge invariance in q. We can still write:

$$B_{n+1}^1(p^{n-1}) \propto e^{[\mu}q^{\nu]} B'_n{}^{\mu\nu}(p^{n-2}). \qquad (3.3.12)$$

We obtain similar constraints as in the subleading case, which imply $B_{n+1}^1 = 0$.

Sub$^{s\geq3}$-leading order At arbitrary order $s \geq 3$ we can write:

$$B_{n+1}^{s-1}(p^{n-1}) \propto e^{[\mu}q^{\nu]} q^{\rho_1} \dots q^{\rho_{s-2}} B_n^{\mu\nu\rho_1\dots\rho_{s-2}}(p^{n-s}). \qquad (3.3.13)$$

48

And all constraints will have $k \leq n - 3$, with $\Delta = s - 1 \geq 1$, so $B_{n+1}^{s-1} = 0$ to all orders up to $s = n$, where the soft expansion terminates. Therefore $M_{n+1} = A_{n+1}$, proving uniqueness.

3.3.2 Gravity

For gravity, we can simply write polarization tensors in terms of polarization vectors as $e_i^{\mu\nu} = e_i^{\mu} f_i^{\nu}$. Then gauge invariance in one graviton becomes equivalent to gauge invariance in two "gluons". The polynomial statement from section 2 still applies, so ignoring momentum conservation, no polynomial with at most k $e.p$ factors can be gauge invariant in $k + 1$ "gluons". With momentum conservation, in the case of gravity this is true for $k < 2n - 4$. This implies that a tensor $B_n^{\mu\nu}$ with k powers of momenta in the numerator cannot be gauge invariant in $k+1$ particles for $k \leq 2n - 4$.

One other difference is that for gravity we are no longer restricted only to cyclic poles, since there is no ordering. In the end, the proof is almost identical to that for Yang-Mills. We assume that $A_n(p^{2n-4})$ is unique, and prove the same is true for $A_{n+1}(p^{2n-2})$.

Leading order The leading piece has a form:

$$M_{n+1}^{-1} = \sum_i \frac{e^{\mu} f^{\nu} B_{n;i}^{\mu\nu}}{q.p_i} . \tag{3.3.14}$$

Gauge invariance in e and f can only be satisfied on the support of momentum conservation, by $B_{n;i}^{\mu\nu} = p_i^{\mu} p_i^{\nu} B_n$, where B_n is a local function at n points. Then the leading piece is:

$$M_{n+1}^{-1}(p^{2n-2}) = \sum_i \frac{e.p_i f.p_i}{q.p_i} B_n(p^{2n-4}) , \tag{3.3.15}$$

49

and now by assumption gauge invariance in the other particles fixes $B_n = A_n$. Using the same trick as before, we consider instead the function $B_{n+1} = M_{n+1} - A_{n+1}$.

Sub-leading order   The subleading piece is given by:

$$B_{n+1}^0 = \sum_i \frac{e^\mu f^\nu q^\rho B_{n;i}^{\mu\nu\rho}}{q \cdot p_i} + \sum_i \frac{e^\mu f^\nu C_{n;i}^{\mu\nu}}{P_i^2} \,. \tag{3.3.16}$$

Gauge invariance in e and f rules out the non-singular pole contributions, and fixes the first term to:

$$\begin{aligned}
B_{n+1}^0(p^{2n-2}) &= \sum_i \frac{e^{[\mu} q^{\nu]} f \cdot p_i B_{n;i}^{\mu\nu}(p^{2n-4})}{q \cdot p_i} \\
&= e^{[\mu} q^{\nu]} B_n^{\mu\nu}(p^{2n-4}) \,, \tag{3.3.17}
\end{aligned}$$

which is ruled out by gauge invariance in the remaining particles.

Sub-sub-leading order   For higher orders, which go up to $s = 2n - 1$, the same argument rules out any other solutions, so A_{n+1} is uniquely fixed by gauge invariance.

3.4   Unitarity from locality and the Adler zero

For the NLSM and DBI we will also deal with local functions $B_n(p^k)$, with k powers of momenta in the numerator. However, the poles are now associated to propagators of quartic diagrams, ordered for the NLSM, un-ordered for DBI. The Adler zero condition [17] states that the amplitude A_n must vanish when a particle is taken soft. Exactly how rapidly it must vanish sets the difference between the NLSM and DBI [18]-[19]. The limit $p_i \to 0$ is taken as $p_i = w_i p_i$, $w_i \to 0$. Then for the NLSM we require the amplitude to vanish as $\mathcal{O}(w_i)$, while for DBI we require $\mathcal{O}(w_i^2)$, $\forall i \neq 3$.

As for gauge invariance, it will be useful to quantify the difference between available momenta and total constraints. In general, if we require a function $B_n(p^k)$ to vanish as $\mathcal{O}(w_i^{g_i})$ for some particle i, let the corresponding constraint be g_i. Then $g = \sum_i g_i$ will the total constraints $B_n(p^k)$ must satisfy, and define $\Delta = g - k$ as before. This time we wish to show that the NLSM amplitude $A_n(p^{n-2})$ is the unique object satisfying $\Delta = 1$ constraints, while the DBI amplitude $A_n(p^{2n-4})$ uniquely satisfies $\Delta = 2$ constraints.

We will also show this by counting possible solutions, order by order in the double soft expansion $q = zq$, $\tilde{q} = z\tilde{q}$, $z \to 0$. The double soft expansion is chosen now, since the functions simply vanish in the single soft limit. The proof will again be almost identical with the ones for YM and GR, with one important difference. In the first two cases, the simple polynomial statement of section 2 significantly streamlined the argument. Remember that in the soft limit, we encountered tensors $B^{\mu\nu}(p^k)$, with $k \leq n-2$, which were immediately ruled out. The key in that case was that we could always associate a polynomial with k $e.p$ factors to a function with k momenta in the numerator. For the NLSM, there is no such (simple) distinction to be made, since all we have are $p_i.p_j$ factors, both in the numerator and denominator.

Therefore, for scalars we do not have a direct proof for the following fact: a function $B_n(p^k)$ cannot satisfy $k+1$ constraints, if $k < n-2$. Instead, this statement must be proven by induction. We will write the proof only for the uniqueness statement, ie $k = n - 2$, under the assumption that the non-existence statement is true. The proof for the latter case is identical, only with $k < n - 2$.

The Taylor series expansion is identical to eq. (3.3.2). We have the singular q-pole terms:

51
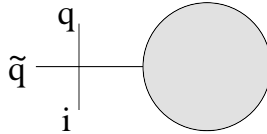
of the form:

$$D_{n+2} = \frac{N}{(q + \tilde{q} + p_i)^2 \mathcal{P}_n(q, \tilde{q})} \qquad (3.4.1)$$

where $i = 1$ or $i = n$ for NLSM due to ordering. In the soft limit we can also write this in terms of the lower point local function:

$$D_{n+2} \to \frac{1}{(q + \tilde{q}).p_i} \frac{N}{\mathcal{P}_n(0, 0)} = \frac{1}{(q + \tilde{q}).p_i} B_n . \qquad (3.4.2)$$

Next there are non-singular poles, which are more varied than in the cubic diagram case. There is still the equivalent of the double poles from before:
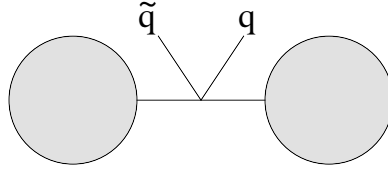
which we can write as:

$$D_{n+2} = \frac{N}{P_L(q, \tilde{q})(p_1 + p_2 + \ldots + p_i)^2 (q + \tilde{q} + p_1 + p_2 + \ldots + p_i)^2 P_R(q, \tilde{q})} , \qquad (3.4.3)$$

In the soft limit this becomes:

$$D_{n+2} \to \frac{1}{(p_1 + p_2 + \ldots + p_i)^2} B_n . \qquad (3.4.4)$$

There are also more complicated non-singular poles, when the q and $\tilde{q}$ legs are separated. However, even in such cases it is easy to write the terms in a form:

$$D_{n+2} \to \frac{1}{P_i^2} B_n . \qquad (3.4.5)$$

3.4.1 NLSM

Leading order   The leading $1/z$ term can only come from q-pole terms:

$$M_{n+2}^{-1} = \frac{N_1(0,0)}{p_1 \cdot (q + \tilde{q})} + \frac{N_2(0,0)}{p_n \cdot (q + \tilde{q})} \,, \tag{3.4.6}$$

imposing vanishing under $\tilde{q} \to 0$ implies:

$$\left(\frac{N_1}{q \cdot p_1} + \frac{N_2}{q \cdot p_n} \right) = 0 \,, \tag{3.4.7}$$

so $N_1 = N_2 = 0$, and $M_{n+2}^{-1} = 0$.

Sub-leading order   At this level both types of poles can contribute. The q-pole piece is:

$$M_{n+2}^{0} = \frac{q^\mu B_n^\mu + \tilde{q}^\mu C_n^\mu}{(q + \tilde{q}) \cdot p_1} + \frac{q^\mu D_n^\mu + \tilde{q}^\mu E_n^\mu}{(q + \tilde{q}) \cdot p_n} \,. \tag{3.4.8}$$

Vanishing under $\tilde{q} \to 0$ implies $B^\mu = p_1^\mu B_n$, and $D_n^\mu = -p_n^\mu B_n$, and similarly $q \to 0$ leads to $C_n^\mu = p_1^\mu C_n$ and $E_n^\mu = -p_n^\mu C_n$. The subleading term becomes:

$$M_{n+2}^{0}(p^n) = \left(\frac{q \cdot p_1}{(q + \tilde{q}) \cdot p_1} - \frac{q \cdot p_n}{(q + \tilde{q}) \cdot p_n} \right) (B_n - C_n) \,. \tag{3.4.9}$$

Now $(B_n - C_n) \equiv B_n(p^{n-2})$ is also a general local function at n-points, so by assumption vanishing in the other soft limits fixes $B_n = A_n$.

Terms with non-singular poles have a form:

$$\sum_i \frac{N_i}{P_i} \,, \tag{3.4.10}$$

but are quickly ruled out by requiring vanishing under q or $\tilde{q}$.

Sub-sub-leading order   The most general q-pole sub-sub-leading term is:

$$M_{n+2}^1 = \frac{1}{(q+\tilde{q}).p_1}(q^\mu q^\nu B^{\mu\nu} + q^\mu \tilde{q}^\nu C^{\mu\nu} + \tilde{q}^\mu \tilde{q}^\nu D^{\mu\nu} + q.\tilde{q}E)$$
$$- \frac{1}{(q+\tilde{q}).p_n}(q^\mu q^\nu F^{\mu\nu} + q^\mu \tilde{q}^\nu G^{\mu\nu} + \tilde{q}^\mu \tilde{q}^\nu H^{\mu\nu} + q.\tilde{q}I). \qquad (3.4.11)$$

Now we expand the remaining $p_i = w_i p_i$ and require $M_{n+2}^1 \propto \mathcal{O}(w_i^1)$ for each of the $(n-1)$ particles left. All of the above functions can be treated as independent because of their unique prefactors. Taking the p's in the denominator into account, we obtain the following constraints for all functions:

$$B^{\mu\nu}, C^{\mu\nu}\, D^{\mu\nu}, E \propto \mathcal{O}(w_1^2), \mathcal{O}(w_{i\neq 1}^1), \qquad (3.4.12)$$

$$F^{\mu\nu}, G^{\mu\nu}\, H^{\mu\nu}, I \propto \mathcal{O}(w_n^2), \mathcal{O}(w_{i\neq n}^1), \qquad (3.4.13)$$

while component-wise there will be two types of constraints. First:

$$B(p^{n-4}), C(p^{n-4}), D(p^{n-4}), G(p^{n-4}) \propto n-2. \qquad (3.4.14)$$

These are $\Delta = 2$ with $k < n-2$ so are ruled out. The other constraints are:

$$E(p^{n-2}),\ I(p^{n-2}) \propto n. \qquad (3.4.15)$$

First, we can use $n-1$ of the usual $\mathcal{O}(w_i^1)$ constraints to fix $E = I = A_n$. But then A_n cannot satisfy the extra requirement of $\mathcal{O}(w_1^2)$ or $\mathcal{O}(w_n^2)$, so it must mean that $I = E = 0$.

Non-singular poles are still ruled out by vanishing under q and $\tilde{q}$, and so $M_{n+2}^1 = 0$.

Sub$^{s\geq 3}$-leading   With each extra q^μ or $\tilde{q}^\mu$ being added, k decreases by 1, so Δ can only increase by at least 1, leading to $\Delta_s = s \geq 2$ constraints. Therefore any sub^s-

leading order vanishes, and so $M_{n+2} = 0$, proving our statement, with the caveat below.

Neutral poles At the sub³-leading order some special combinations of non-singular pole terms are not directly ruled out. Consider for example the two diagrams, which we take to have equal numerators:
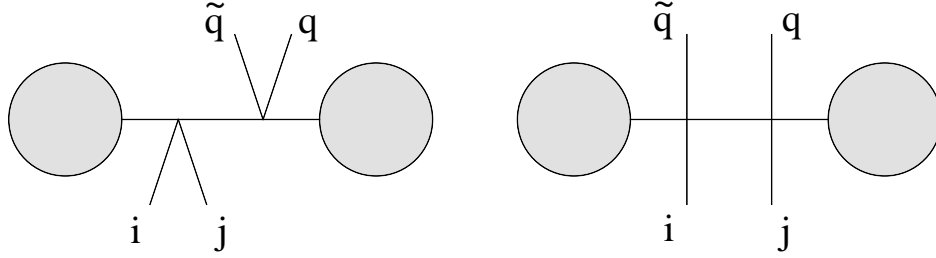
given by:

$$
D_{n+2}(p^n) = \frac{N(p^n)}{P_L^2(P_L + i + j)^2(P_L + i + j + q + \tilde{q})^2 P_R^2(q, \tilde{q})}
$$
$$
- \frac{N(p^n)}{P_L^2(P_L + i + \tilde{q})^2(P_L + i + j + q + \tilde{q})^2 P_R^2(q, \tilde{q})} . \tag{3.4.16}
$$

At sub$^{s \geq 3}$-leading order their contribution is:

$$
D_{n+2}^2(p^n) = q^\mu \tilde{q}^\nu N^{\mu\nu}(p^{n-2}) \left(\frac{1}{P_L^2(P_L + p_i + p_j)^2 P_R^2} - \frac{1}{P_L^2(P_L + p_i)^2 P_R^2} \right) . \tag{3.4.17}
$$

Now $N^{\mu\nu}(p^{n-2})$ has enough momenta to trivially satisfy $n - 2$ of the remaining $n - 1$ constraints. But vanishing in particle j is automatic because the two denominators in (3.4.17) become equal when $p_j \to 0$. Therefore D_{n+2}^2 is not ruled out by our usual argument. Instead, such terms can be ruled out by taking different soft limits. Specifically, it must be soft limits which lead to soft-singularities in P_L^2 or P_R^2. This ensures that D_{n+2}^2 above avoids non-singular pole terms in the new soft limit.

3.4.2 Dirac-Born-Infeld

For DBI we can use the same notation from the previous section. In this case, the Adler zero condition is stronger, as we require $A_n \propto \mathcal{O}(w_i^2)$ under $p_i = w_i p_i \to 0$. The proof is identical to the one for the NLSM, with the minor difference that now non-cyclic poles are allowed. Also, in all cases non-singular poles can be ruled out easily - the issue appearing in the NLSM is not present, since the vanishing under $p_j \to 0$ ensured by eq. (3.4.17) can not provide the full $\mathcal{O}(w_i^2)$ needed. Instead, there is a different issue appearing at the same order, which can be resolved by demanding permutation invariance.

Leading order   The leading piece is given by:

$$B_{n+2}^{-1} = \sum_{i=1}^{n} \frac{N_i}{(q+\tilde{q}).p_i}, \qquad (3.4.18)$$

but is ruled out by requiring $B_{n+2}^{-1} \propto \mathcal{O}(w^1)$ in q or $\tilde{q}$.

Sub-leading order   Regular q-pole terms have the form:

$$\sum_{i=1}^{n} \frac{1}{(q+\tilde{q}).p_i}(q^\mu B_i^\mu + \tilde{q}^\mu C_i^\mu), \qquad (3.4.19)$$

but are ruled out by requiring $\mathcal{O}(w^2)$ under q and $\tilde{q}$.

Sub-sub-leading order   Have the form:

$$M_{n+2}^1 = \sum_i \frac{1}{(q+\tilde{q}).p_i}(q^\mu q^\nu B_i^{\mu\nu} + \tilde{q}^\mu \tilde{q}^\nu C_i^{\mu\nu} + q^\mu \tilde{q}^\nu D_i^{\mu\nu} + q.\tilde{q} E_i). \qquad (3.4.20)$$

Requiring $\mathcal{O}(w^2)$ in $q, \tilde{q}$ we end up with:

$$M_{n+2}^1(p^{2n}) = \sum_i \frac{1}{(q+\tilde{q}).p_i}((q.p_i)^2 B + (\tilde{q}.p_i)^2 C + q.p_i\tilde{q}.p_i D)$$

$$= \sum_i \frac{q.p_i\tilde{q}.p_i}{(q+\tilde{q}).p_i}(-B-C+D). \tag{3.4.21}$$

Now $(-B-C+D) \equiv B_n(p^{2n-4})$ is a general local function at n-points, so imposing the remaining $2n-2$ constraints fixes $B_n = A_n$ by assumption.

Sub³-leading order Like for the NLSM, at this order extra care is required. The usual arguments rule out all terms except:

$$D_{n+2}^2(p^{2n}) = q.\tilde{q} \sum_i \frac{q.p_i B_{n;i} + \tilde{q}.p_i C_{n;i}}{(q+\tilde{q}).p_i}, \tag{3.4.22}$$

under the condition that $\sum_i B_{n;i} = \sum_i C_{n;i} = 0$. The functions $B_i(p^{2n-4})$ and $C_i(p^{2n-4})$ must satisfy $\Delta = 2$ constraints and are fixed (up to some coefficient) to A_n by assumption. Then the extra conditions become $\sum_i B_{n;i} = \sum_i b_i A_n = 0$, and similarly $\sum C_i = \sum_i c_i A_n = 0$. The sub³-leading term becomes:

$$D_{n+2}^2 = q.\tilde{q} A_n \sum_i \frac{b_i\, q.p_i + c_i\, \tilde{q}.p_i}{(q+\tilde{q}).p_i}. \tag{3.4.23}$$

But now if we require symmetry in $q \leftrightarrow \tilde{q}$, then $b_i = c_i$, so $D_{n+2}^2 = 0$, and this order vanishes.

Sub$^{s>3}$leading All such terms are ruled out, so $M_{n+2} = 0$ to all orders, and A_{n+2} is unique.

3.5 Locality and unitarity from singularities and gauge invariance

The general argument we used in the previous sections can be easily extended when we relax our cubic graph assumptions, and instead consider a more general singularity structure, as long as the singularities themselves have a form $\left(\sum_i p_i\right)^2$, with consecutive momenta in the case of Yang-Mills. This means we allow double poles as well as overlaps. There are three cases to consider depending on how many singularities s we allow. We will show that:

- for $s < n - 3$ there is no solution

- for $s = n - 3$ there is a unique solution, A_n

- for $s > n - 3$ solutions can be factorized in a form $\left(\sum \text{poles}\right) \times A_n$

We prove these three results for five points Yang-Mills. It is easy to extend the proof for general n, including for gravity.

In the following, we will call a function with s singularities $B_{n;s}$, and to simplify notation, let:

$$S_0 = \frac{e.p_1}{q.p_4} - \frac{e.p_4}{q.p_4}\,. \tag{3.5.1}$$

3.5.1 Case 1. $s < n - 3$

This case is easy to prove by induction. Assume that $B_{4;0}(p^2)$ is ruled out by gauge invariance. Then the five point function with just one singularity has a leading order:

$$M_{5;1}^{-1}(p^3) = S_0 B_{4;0}(p^2)\,, \tag{3.5.2}$$

which by assumption is ruled out. Higher order terms are again ruled out as usual, so there are no solutions for $s < n - 3$.

3.5.2 Case 2. $s = n - 3$

At five points, in this case we have two (cyclic) poles per term, and now we also allow double poles and overlaps.

Order z^{-2} The lowest order is now z^{-2}, coming from three possible terms, which were not present before:

$$M_{5;2}^{-2}(p^3) = \frac{N_a}{(q.p_1)^2} + \frac{N_b}{(q.p_1)(q.p_4)} + \frac{N_c}{(q.p_4)^2} \,. \tag{3.5.3}$$

Gauge invariance in q requires the forms:

$$M_{5;2}^{-2}(p^3) = \frac{1}{q.p_1} S_0 B_{4;0}(p^2) + \frac{1}{q.p_4} S_0 C_{4;0}(p^2) \,, \tag{3.5.4}$$

so both $B_{4;0}$ and $C_{4;0}$ vanish by the previous argument.

Order z^{-1} At this order we have the usual leading piece, but also terms with the non-local poles from above:

$$M_5^{-1} = S_0 B_4(p^2) + e^{[\mu} q^{\nu]} \left(\frac{N_a^{\mu\nu}}{(q.p_1)^2} + \frac{N_b^{\mu\nu}}{(q.p_1)(q.p_4)} + \frac{N_c^{\mu\nu}}{(q.p_4)^2} \right) \,.$$

For the second piece we need tensors $N_4^{\mu\nu}(p^2)$, gauge invariant in three particles, which is not possible. Therefore the leading piece is just

$$M_5^{-1} = S_0 A_4 \,, \tag{3.5.5}$$

and so far we get the same answer as usual. However, we must deal with a subtle issue that was not present before. We have shown that at the leading order, all possible functions must map onto the unique expression (3.5.5). But when we allow a non-local singularity structure, it is possible for two different $n + 1$ point functions to have an identical leading order piece. Consider for example the actual amplitude, which contains a local term such as:

$$A_5 = \ldots \frac{e.p_1 N}{q.p_1 (q + p_1 + p_2)^2} + \ldots, \tag{3.5.6}$$

and a similar function M_5, but where we replace the term from above with a non-local one:

$$M_5 = \ldots \frac{e.p_1 N}{q.p_1 (p_1 + p_2)^2} + \ldots. \tag{3.5.7}$$

In the soft limit $q \to 0$ both functions are equal at the leading order, so apparently we have two different solutions, contradicting our statement. The issue can still be resolved by considering all orders of the soft expansion of $B_5 = M_5 - A_5$. The subleading order is now different than the usual (3.3.9), because B_5 now has a contribution originating from the Taylor series expansion of the denominator in eq. (3.5.6), which was absent before. We obtain:

$$B_5^0 = \frac{e^\mu q^\nu B_4^{\mu\nu}}{q.p_i} + e^\mu C_4^\mu - \frac{e.p_i q.(p_1 + p_2) N}{q.p_i (p_1.p_2)^2} + \ldots \tag{3.5.8}$$

where the third term is new. But using our previous arguments B_5^0 is still ruled out by gauge invariance. Higher order terms can be treated in a similar manner, so $B_5 = 0$ to all orders. Therefore the five point Yang-Mills amplitude is completely fixed even if we start with these non-local assumptions.

3.5.3   Case 3. $s > n - 3$

For this case, where we are not expecting to obtain a unique answer, but the same soft limit argument can be used to count the maximum total number of independent solutions, order by order. First, at four points it is easy to check that with $s = 2$ poles, there are two solutions:

$$M_{4;2} = \left(\frac{a_1}{p_1 \cdot p_2} + \frac{a_2}{p_1 \cdot p_4} \right) A_4 \,. \tag{3.5.9}$$

Now at five points, with three poles, we want to show there are five solutions, corresponding to the five different cyclic poles:

$$M_{5;3} = \left(\frac{a_1}{p_1 \cdot p_2} + \frac{a_2}{p_2 \cdot p_3} + \ldots + \frac{a_5}{p_5 \cdot p_1} \right) A_5 \,. \tag{3.5.10}$$

Again taking a soft limit, and imposing gauge invariance in $p_5 = q$, we obtain:

Order $\mathcal{O}(z^{-3})$

$$M_{5;3}^{-3} = \frac{1}{(q \cdot p_1)^2} S_0 B_{4;0} + \frac{1}{(q \cdot p_4)^2} S_0 C_{4;0} \,, \tag{3.5.11}$$

which was shown to vanish, so no solutions at this level.

Order $\mathcal{O}(z^{-2})$

$$M_{5;3}^{-2} = \frac{1}{q \cdot p_1} S_0 B_{4;1} + \frac{1}{q \cdot p_4} S_0 C_{4;1} \,, \tag{3.5.12}$$

which is fixed by gauge invariance to

$$M_{5;3}^{-2} = \left(\frac{a_5}{q \cdot p_1} + \frac{a_4}{q \cdot p_4} \right) S_0 A_4 \,. \tag{3.5.13}$$

Therefore from this order we obtain two possible solutions.

Order $\mathcal{O}(z^{-1})$ Because we are only counting *independent* solutions, we can simply ignore the contributions from the lower order above. Therefore we are only interested in the term:

$$M_{5;3}^{-2} = S_0 B_{4;2} \,. \tag{3.5.14}$$

By assumption this gives two independent solutions corresponding to the poles $p_1.p_2$ and $p_1.p_4$, but starting from five points there are three poles which map onto these two in the soft limit:

$$p_1.p_2 \to p_1.p_2 \,, \tag{3.5.15}$$

$$p_2.p_3 = (p_4 + p_5 + p_1)^2 \to p_1.p_4 \,, \tag{3.5.16}$$

$$p_3.p_4 = (p_5 + p_1 + p_2)^2 \to p_1.p_2 \,. \tag{3.5.17}$$

And so we obtain three independent solutions at this order. For higher orders, the usual arguments rule out other solutions, and so we end up with at most five possible solutions. We have not derived what these must be, but since we can just write down the five terms of Eq. (3.5.10), this must be all of them. The result is easy to generalize to an arbitrary number of extra poles.

The argument can also easily be extended to general n-point amplitudes, as well as gravity. Once it is shown that functions with $s < n - 3$ singularities are ruled out, for $s = n - 3$ the only non-vanishing contribution will be the Weinberg term at order $1/z$, which by the usual argument implies uniqueness. We suspect the same type of argument can be used for NLSM and DBI, although some extra complications might appear at the sub³-leading orders, which were already troublesome. Regardless, a

more direct argument ruling out the non-local terms was already presented in Ref. [22] for these theories.

3.6 Generalizing singularities

3.6.1 Non-local singularities

In the previous sections we have assumed that the denominators are always products of singularities $P_{\mathcal{S}}^2 = (\sum_i p_i)^2$. An obvious next step is to relax even this assumption, and allow completely non-local singularities of the form $(\sum_i a_i p_i)^2$. In full generality, this doesn't work out. Even at four points, allowing a singularity of the form $a \, p_1.p_2 + b \, p_1.p_4$ no longer provides a unique local solution. We can write the four point numerator as $N_4 = (t N_s + s N_t) = (t, s) \cdot (N_s, N_t)$, with $A_4 = N_4/(st)$. Now we can do any 2D rotation to obtain $N_4 = (t', s') \cdot (N_s', N_t')$, where $s' = s \cos\theta - t \sin\theta$ and $t' = t \cos\theta + s \sin\theta$. But now diving by $(s't')$ we obtain a (non-unique) gauge invariant with the non-local poles s' and t', so our claim is invalidated if we allow such poles.

However, there exists a special set of non-local "cyclic" poles from which full locality can still be derived, if we are careful about momentum conservation. To obtain this set, we must start from a local cyclic pole $P_{jk}^2 = (\sum_{i=j}^{k} p_i)^2$. Now only *after* using momentum conservation $p_3 = -\sum p_i$, we can add arbitrary coefficients $(\sum_i p_i)^2 \to (\sum_i a_i p_i)^2$. For example, from a six point local pole like $(p_1 + p_2 + p_3)^2$, we can obtain $(a \, p_4 + b \, p_5 + c \, p_6)^2$. Note how this rule doesn't allow the four point pole $a \, p_1.p_2 + b \, p_1.p_4$ from above. It can only come from the pole $p_1.p_3 = p_1.p_2 + p_1.p_4$, which is not cyclic. At five points, the most general set of singularities that can be

used is:

$$(p_1 + p_2)^2 = p_1.p_2$$

$$(p_2 + p_3)^2 = (p_1 + p_4 + p_5)^2 \to a_1\, p_1.p_4 + a_2\, p_1.p_5 + a_3\, p_4.p_5$$

$$(p_3 + p_4)^2 = (p_1 + p_2 + p_5)^2 \to a_4\, p_1.p_2 + a_5\, p_1.p_5 + a_6\, p_2.p_5$$

$$(p_4 + p_5)^2 = p_4.p_5$$

$$(p_5 + p_1)^2 = p_1.p_5 \qquad (3.6.1)$$

For an n point amplitude, $n - 2$ of the singularities keep the form $p_i.p_{i+1}$, while the others are promoted to carry these extra coefficients. Now, the usual proof by induction will work, as long as we avoid taking soft the particles adjacent to 3, which is of course always possible from four points and higher. This procedure ensures that the soft-singularities $q.p_i$, critical for the leading term, are not affected in any way. Then the leading term is as usual

$$B_{n+1}^{\text{non-local}} \to \left(\frac{e.p_1}{q.p_1} - \frac{e.p_n}{q.p_n} \right) B_n^{\text{non-local}} , \qquad (3.6.2)$$

where now $B_n^{\text{non-local}}$ also contains the non-local singularities described above. If by assumption even this non-local B_n is uniquely fixed by gauge invariance, ultimately so will B_{n+1}. The claim is in fact trivial at four points, where none of the poles may be modified, so $B_4^{\text{non-local}} = B_4^{\text{local}}$.

With a few extra restrictions, a similar result can be shown for gravity as well, though the procedure is somewhat more complicated because for gravity soft-singularities involving p_3 are not so easily avoided. The solution is to require several extra poles to keep their usual local form, in such a way to ensure that even after taking multiple soft limits, there always exists a particle which forms no soft-singularities with p_3.

3.6.2 No singularities

So far, we have mostly looked at functions with singularities of the form $(\sum_i p_i)^2$, and in some cases we showed that singularities of the type $(\sum_i a_i p_i)^2$ also lead to uniqueness. But what about allowing the denominators to be polynomials of some degree s^2, instead of s products of singularities? In general, this is a very difficult question to systematically analyze, and given the four point counter-example from the previous section, it might simply be an ill-posed question. But instead of trying to understand all such completely general poles, there is an even more general alternative to pursue. We can completely disregard singularities, and investigate gauge invariance directly at the level of the total numerator, by considering general polynomials instead of functions with poles. Clearly, given sufficient mass dimension, a general polynomial can always be thought of as originating from the most general singularity structure possible. We can start with the minimal polynomial which admits any solution, which has $n-2$ $e.p$ factors, and $(n-2)^2$ total mass dimension, the same as an actual amplitude numerator. It turns out that imposing our usual $n-1$ gauge invariance constraints does not provide a meaningful solution, but imposing the full n constraints does: we obtain a linear combination of amplitude numerators! The nth extra constraint essentially is required to replace the information we lost by not considering denominators which are products of singularities. From this perspective, the singularities do no play any crucial or physical role, but only provide a useful method of organizing terms in the polynomial. While we do not have a proof for this fact for $n > 4$, it is easily testable at five points. There we obtain six solutions, which are linear combinations of five point amplitude numerators, corresponding to different orderings. Below we provide leading order evidence for this fact.

We can again use our usual soft argument to count the solutions at leading order. First, it is easy to check that imposing all four gauge invariance conditions on the

four point polynomial $N_4((e.p)^2, p^4)$ gives a unique solution. This corresponds to the fact that all four point amplitudes have the same numerator. That is, any amplitude can be obtained by dividing the same numerator by the desired propagator structure:

$$A(1, 2, 3, 4) = \frac{N}{p_1.p_2 \, p_1.p_4} \, , \tag{3.6.3}$$

$$A(1, 3, 2, 4) = \frac{N}{p_1.p_3 \, p_1.p_4} \, . \tag{3.6.4}$$

At five points, the leading piece of the general polynomial must have a form:

$$N_5((e.p)^3, p^9) = e^{[\mu} q^{\nu]} N^{\mu\nu}(p^8) \, . \tag{3.6.5}$$

After imposing the other four constraints all possible components are ruled out, except the following:

$$N_5((e.p)^3, p^9) = S_{12} N_a + S_{14} N_b + S_{24} N_c \, , \tag{3.6.6}$$

where $S_{ij} = e.p_i q.p_j - e.p_j q.p_i$. Now the $N_i((e.p)^2, p^6)$ must also satisfy the four constraints. First, we can rewrite $N((e.p)^2, p^6) = N((e.p)^2, p^4) \sum_{i,j} a_{ij} p_i.p_j$, after a reshuffling of the coefficients. Then, the constraints imposed on $N((e.p)^2, p^6)$ instead act on $N((e.p)^2, p^4)$, which by assumption is fixed uniquely to the four point numerator. Finally, there are two independent $p_i.p_j$ factors at four points. Therefore we obtain

$$N((e.p)^3, p^9) = (a_1 p_1.p_2 + a_2 p_1.p_4)(b_1 S_{12} + b_2 S_{14} + b_3 S_{24}) N_4 \, , \tag{3.6.7}$$

ie. six independent solutions, which are related to the leading pieces of amplitude numerators. Unfortunately, the subleading order is not ruled out so quickly. The N still have enough momenta to provide gauge invariant contributions even at this

order:

$$N((e.p)^3, p^9) = (a_1.q.p_1 + a_2 q.p_2 + a_3 q.p_4)(b_1 S_{12} + b_2 S_{14} + b_3 S_{24})N_4 \,, \qquad (3.6.8)$$

and so the usual argument fails in its simplest form. However, considering all orders, eventually these extra solutions become tied to the original six, and in the end just six solutions are left. The argument becomes even less well suited for higher points, so clearly a better strategy is required.

3.7   Summary of the results and future directions

In this note we have presented the full proofs for some of the uniqueness claims originally made in [22]. We summarize these results below. Let s be the number of poles of the form $(\sum_i p_i)^2$, and k the mass dimension of the numerators.

Yang-Mills and General Relativity:

- Unique solution for $s = n - 3$, with $k_{\text{YM}} = n - 2$, $k_{\text{GR}} = 2n - 4$

- No solutions for s or k smaller than above

- Factorized solutions $(\sum \text{poles}) \times A_n$ for s larger than above

NLSM and DBI:

- Uniqueness assuming quartic diagrams, with $k_{\text{NLSM}} = n - 2$, $k_{\text{DBI}} = 2n - 4$

- No solutions for k smaller than above

For Yang-Mills, we also proved that uniqueness holds when allowing specific types of non-local singularities $(\sum_i a_i p_i)^2$. Finally, we conjectured that general polynomials

of minimal mass dimension lead to linear combinations of amplitude numerators, and so to both locality and unitarity.

The next step is understanding how to approach such polynomials with absolutely no singularity structure. It would be very interesting to see if the soft limit argument can be extended even further, or if an even more powerful argument is required. Meanwhile, for NLSM and DBI, it is not even clear what the equivalent claim should be, if it exists. For YM, the number of $e.p$ factors always helped distinguish what p's come from numerators and which come from propagators. For scalar theories, there is no distinction to be made: all the p's are equal. We should note that an equivalent claim for gravity does not exist. It is trivial to obtain many different solutions by gluing together Yang-Mills amplitudes, while there is a unique gravity numerator. Nevertheless, even if the numerator statement is less fundamental than the other results, it is a very useful exercise. After all, thinking about polynomials lead to the crucial results of section 3.2, so perhaps there is more to be learned from this perspective.

A more important issue to be understood is that of the gram determinant relations. When working in some fixed dimension D, at most $D - 1$ vectors can be linearly independent (-1 because of momentum conservation). For example, if we restrict to 4D, starting at six points, we can express p_6 in terms of the other four independent momenta:

$$p_6 = a\,p_1 + b\,p_2 + c\,p_4 + d\,p_5 \tag{3.7.1}$$

This could allow for different solutions to our requirements. The linear dependence (3.7.1) can be viewed as another form of momentum conservation:

$$p_3 = -(p_1 + p_2 + p_4 + p_5 + p_6) \tag{3.7.2}$$

We already saw that adding momentum conservation limited the applicability of our initial polynomial argument to $k < n - 2$: at $k = n - 2$ momentum conservation allowed for some "free" gauge invariants to be formed. Luckily, this was still sufficiently constraining for our purposes. It is not inconceivable, though would be very surprising, that the gram determinant relations could allow such free gauge invariants starting at $k = n - 3$ for example.

Ultimately, these results strongly suggest that scattering amplitudes might have a different definition, perhaps geometric, in line with the amplituhedron program [5]. A formulation where both this minimal singularity structure and gauge invariance/vanishing in the soft limit are manifest could potentially uncover yet more unknown features of these theories.

Acknowledgments

Chapter 4

Uniqueness from locality and BCFW

4.1 Introduction

The traditional formulation of Quantum Field Theory is based on Feynman diagrams, which ensure that locality and unitarity are manifest at all times. But to accomplish this, Feynman diagrams introduce a large amount of unphysical redundancy, which hides the ultimate simplicity of scattering amplitudes in many theories. Nowhere is this more striking than in Yang-Mills and General Relativity, where gauge and diffeomorphism invariance lead to very complicated Feynman diagram expansions, containing thousands of terms even for five particle scattering. But quite surprisingly, with the right variables, such expressions can ultimately be collapsed into remarkably simple answers, such as the Parke-Taylor formula for Yang-Mills [1], or the Kawai-Lewellen-Tye relations for gravity [6].

The S-matrix program on the other hand aims to replace the Lagrangian formulation by directly imposing physical principles on scattering amplitudes. Motivated by this hidden simplicity and dealing away with gauge invariance, the on-shell perspec-

tive recently lead to many computational and conceptual advances, chiefly through the use of recursion relations [50, 2]. Other recent developments have revealed more and more previously unknown facets of scattering amplitudes: the twistor string picture [15], the BCJ duality [7], scattering equations [14], and many others. However, in this approach the physical nature of scattering amplitudes - that is, particles scattering off of each other in local quantum interactions - is completely lost in favor of more abstract properties and symmetries. The amplituhedron [5] is a prime example of this newer perspective. There, scattering amplitudes can be understood as volumes of certain polytopes, with locality and unitarity emerging from the geometry itself.

The on-shell and off-shell techniques therefore capture very different aspects of scattering amplitudes, and the goal of this paper is to explore the interface of these two otherwise orthogonal perspectives.

In section 4.2, we bridge the gap between on-shell recursions and Feynman amplitudes, by introducing a BCFW shift compatible with arbitrary polarization vectors. This shift is in fact the manifestly covariant form of the shift used in Ref. [51], and can be used in the usual way to recursively build the full amplitude, valid in any dimension and for all helicity configurations.

In section 4.3, we show that the Yang-Mills amplitude is completely fixed by imposing locality (singularity structure given by propagators associated to cubic diagrams) and constructability (vanishing of poles at infinity under BCFW shifts). Unitarity (factorization) is never used, but instead emerges as a consequence of uniqueness. This result is very similar to (and was motivated by), the recent results in [22], which showed that gauge invariance uniquely fixes the Yang-Mills and gravity amplitudes. It is worth noting that the needed asymptotic behaviors which will show up are more general than those required in the usual BCFW recursion, and in fact we will not use the Cauchy theorem at any point. Instead of building the scattering amplitude directly via on-shell recursions (which would mean assuming unitarity), we

71

simply prove that there is a unique local object satisfying the vanishing conditions for $z \to \infty$.

The strategy is almost identical to the one in [22]: we show uniqueness order by order in the soft expansion by using induction. However, checking BCFW behavior is more complicated than checking gauge invariance. Instead of simply imposing vanishing under $e_i \to p_i$ for $n-1$ particles, now we must require some specific $\mathcal{O}(z^m)$ behavior under $n(n-1)$ BCFW shifts $[i, j\rangle$ which involve e_i, p_i, e_j, and p_j. This makes finding a precise inductive argument more difficult. Even fixing only the leading term, which was immediate with gauge invariance, is a lot more involved. This time, we cannot even impose any shift involving the soft particle, as it interferes with momentum conservation. In fact, in this case the process is reversed: the lower point amplitude is fixed first, and the soft factor is fixed last. For this reason, we limit our discussion to the leading order, and conjecture that the same argument can be used for the subleading orders as well. Nevertheless, explicit checks of the all-order statement have been made up to five points.

4.2 BCFW with polarization vectors

We define our $[i, j\rangle$ shift in the following way:

$$
\begin{aligned}
e_i &\to \hat{e}_i \,, \\
e_j &\to \hat{e}_j + z p_i \frac{\hat{e}_i . \hat{e}_j}{p_i . p_j} \,, \\
p_i &\to p_i + z \hat{e}_i \,, \\
p_j &\to p_j - z \hat{e}_i \,,
\end{aligned}
\tag{4.2.1}
$$

where $\hat{e}_i = e_i - p_i \frac{e_i . p_j}{p_i . p_j}$, and similarly for $\hat{e}_j$. The motivation for this peculiar shift, which generalizes the shift in Ref. [51], is that it maintains the on shell conditions $e_i . p_i = 0$ and $e_j . p_j = 0$. Alternativey, a simpler version of the shifts may used, by

72

dropping the gauge shift on the polarization vectors, $\hat{e} \to e$, but manually imposing $e_i.p_j = e_j.p_i = 0$ after performing the shift. The shifts are then equivalent, and in the rest of this paper the second shorter version will be used. It is also worth noting that the shifts are gauge invariant in i and j, but non-local due to the extra poles.

It can be checked, though we do not prove, that any gluon amplitude satisfies the usual BCFW behavior under this shift. That is, under any shift $[i, j\rangle$:

$$A_n \propto \mathcal{O}(z^{-1}) \text{ for } i \text{ and } j \text{ adjacent} ,$$
$$A_n \propto \mathcal{O}(z^{-2}) \text{ for } i \text{ and } j \text{ non-adjacent} . \tag{4.2.2}$$

In this case there is no "bad shift", common to the on-shell method, as that is merely a by-product of an asymmetry imposed on the λ's and $\tilde{\lambda}$'s. This shift can be used in the usual way to build general gluon amplitudes. For example, with a $[1, n\rangle$ shift:

$$A_n = \sum \frac{A_{i+1}^L(\hat{1}, 2, ..., i, P) A_{n-i+1}^R(-P, i+1, ..., n-1, \hat{n})}{P_i^2} , \tag{4.2.3}$$

where $z_i = \frac{P_i^2}{\hat{e}_i.P_i}$, and the usual summing over internal polarizations can be done using $\sum e^\mu e^\nu = \eta^{\mu\nu}$. Starting from the three point amplitude (which is just the three point Feynman vertex):

$$A_3(1, 2, 3) = V_3(1, 2, 3) = e_1.e_2 e_3.p_1 + e_2.e_3 e_1.p_2 - e_3.e_1 e_2.p_1 , \tag{4.2.4}$$

the four point amplitude can be obtained from a $[1, 4\rangle$ shift as:

$$A_4(1, 2, 3, 4) = \frac{A_3(\hat{1}, 2, P) A_3(-P, 3, \hat{4})}{p_1.p_2} , \tag{4.2.5}$$

with $z = p_1.p_2/\hat{e}_1.p_2$. It is easy to verify this is equal to the known amplitude, which in terms of Feynman diagrams is given by:

$$A_4(1,2,3,4) = \frac{V_3(1,2,P)V_3(-P,3,4)}{p_1.p_2} + \frac{V_3(1,4,P)V_3(-P,3,2)}{p_1.p_4} + V_4(1,2,3,4)\,,$$

(4.2.6)

where $V_4 = e_1.e_3\, e_2.e_4$. Comparing eqs. (4.2.5) and (4.2.6) makes clear the purpose of the non-local pole contained in the shifts: it generates the $p_1.p_4$ pole of the other channel. The computational advantage of this approach comes from the fact that fewer BCFW terms have to be considered compared to Feynman diagrams. In general, to compute an n point amplitude there will be just $n-3$ BCFW terms to write down, compared to the factorially growing number of Feynman diagrams.

4.3   Uniqueness from BCFW and locality

Besides the usual application of this shift to recursion relations, we conjecture that in fact the Yang-Mills scattering amplitude is the unique local object of mass dimension $[4-n]$ compatible with the usual BCFW behavior (4.2.2). As in Ref. [22], we start with an ansatz of local functions:

$$M_n(p^{n-2}) = \sum_i \frac{N_i(p^{n-2})}{\prod_{\alpha_i} P_{\alpha_i}^2}\,,$$

(4.3.1)

where the sum is taken over all ordered cubic diagrams i, and α_i correspond to the channels of diagram i. The numerators N_i are general polynomials of mass dimension $[n-2]$, and are linear in n polarization vectors, but carry no information of factorization. Then, by requiring vanishing at infinity in a sufficient number of shifts, we obtain a unique solution, the gluon amplitude A_n. Empirically, it turns out that some shifts can be ignored completely, and the amplitude is still fixed. For example,

at four points three shifts (for example, $[1,2\rangle$, $[2,1\rangle$, and $[2,3\rangle$) are enough to fix the answer, while at five points five shifts are needed. Furthermore, the required behavior in some shifts can be relaxed, and still the amplitude is fixed.

For the purposes of this proof, we will impose the maximal number of shifts, that is for all pairs i and j from 1 to n, but with one crucial modification. For some shifts we will impose weaker constraints: under all the shifts involving some particle h, we will demand only $\mathcal{O}(z^0)$ for adjacent, and $\mathcal{O}(z^{-1})$ for non-adjacent shifts. This modification will be necessary for the inductive argument, which is carried out precisely by taking the special particle h soft. We leave to future work the issue of finding the minimal set of shifts which fixes the amplitude.

4.3.1 Overview of the proof

It will be useful to introduce the following notation. Let E_n be the set of all polarization vectors at n points, and call $\mathcal{G}_n^h(E_n)$ the constraints (4.2.2), relaxed for particle h. Then we would like to prove that if $\mathcal{G}_n^h(E_n)$ for all $h = \overline{1,n}$ uniquely fixes $A_n(E_n, p^{n-2})$, the equivalent higher point set $\mathcal{G}_{n+1}^{h'}(E_{n+1})$ uniquely fixes $A_{n+1}(E_{n+1}, p^{n-1})$, for all $h' = \overline{1, n+1}$. We will prove this statement for the choice $h' = n+1$, under the assumption that A_n is fixed by both $\mathcal{G}_n^1(E_n)$ and $\mathcal{G}_n^n(E_n)$. All other choices for h' can be treated in an identical manner, by taking h' soft.

The basic logic of the argument is identical to that in Ref. [22]. We consider a general local object at $n+1$ points, $M_{n+1}(p^{n-1})\delta_{n+1}$, and show that imposing our constraints forces $M_{n+1} = A_{n+1}$, order by order in the soft expansion. Let $e_{n+1} = e$, $p_{n+1} = z\,q$, and expand $M_{n+1}\delta_{n+1}$ around $z = 0$. Using momentum conservation to express p_3 in terms of the other momenta, the leading $1/z$ term has the general form:

$$M_{n+1}^{-1} = \frac{\sum_i e.e_i B_{n;1}^i + \sum_{i\neq 3} e.p_i C_{n;1}^i}{q.p_1} + \frac{\sum_i e.e_i B_{n;n}^i + \sum_{i\neq 3} e.p_i C_{n;n}^i}{q.p_n}, \qquad (4.3.2)$$

where $B_{n;h}^i = B_{n;h}^i(\{e_1, e_2, \ldots e_n\} \setminus e_i, p^{n-1})$ and $C_{n;h}^i = C_{n;h}^i(\{e_1, e_2, \ldots e_n\}, p^{n-2})$, with $h = 1, n$, are local functions at n-points. We will show that imposing the BCFW constraints (4.2.2), relaxed for particle $n + 1$, uniquely fixes:

$$M_{n+1}^{-1} = \left(\frac{e.p_1}{q.p_1} - \frac{e.p_n}{q.p_n} \right) A_n , \tag{4.3.3}$$

which is the known leading piece of the Yang-Mills scattering amplitude [9]. In principle, the argument would then continue to show that all subleading order terms of the object $M_{n+1}' \equiv M_{n+1} - A_{n+1}$ vanish, impying that $M_{n+1} = A_{n+1}$, completing the induction.

Formula (4.3.2) also reveals why we must carry around this extra modification due to h. Shifts involving h also shift the $q.p_h$ poles, producing one power of z in the denominator. For example, if $M_{n+1} \propto z^{-1}$ under a shift $[1, 2\rangle$, we should expect $B_{n,1} \propto z^0$ and $C_{n,1} \propto z^0$ (ignoring the prefactors). But then in order for the inductive argument to close, the $n + 1$ point constraints must also include such modified behaviors under particular shifts, which should not carry over to the lower point functions. Otherwise such relaxations would keep accumulating whenever we take a soft limit. This is precisely why particle h is the one taken soft. After a soft limit $p_{n+1} \to 0$ we cannot not impose any shifts involving particle $n + 1$, as it would not be consistent with momentum conservation. This is the only way to ensure that all functions always have just one particle with relaxed constraints.

The proof will have four steps:

1. We show by induction that all functions $B_{n;h}^i$ in (4.3.2) are ruled out.

2. We show that the $C_{n;h}^i$ are fixed by the uniqueness assumption at n-points, such that $C_{n;1}^i = a_i A_n$, and $C_{n;n}^i = b_i A_n$.

3. Using shifts involving particle 3, which was chosen to impose momentum conservation, we show that $a_i = 0$ for $i \neq 1$, and $b_i = 0$ for $i \neq n$.

4. Finally, we use the $[1, n\rangle$ shift to fix $a_1 = -b_n$. This shift is special because it is adjacent in A_n, but non-adjacent in A_{n+1}.

The first step is the most laborious, and is carried out in section (4.3.2). The last three steps are completed in section (4.3.3).

4.3.2 Ruling out $B(E_n^a)$ functions

First, to prove that the $B_{n;h}^i(E_n^{n-1}, p^{n-1})$ functions are ruled out, we will have to consider the whole set of functions:

$$B_{n;h}(a) \equiv B_{n;h}(E_n^a, p^{2n-2-a})\delta_n, \ a = \overline{0, n-1}, \tag{4.3.4}$$

linear in just a polarization vectors, which form the set E_n^a. The second lower index designates precisely the special particle h mentioned above. In this case we can have $h = 1$ or $h = n$.

Now that E_n^a does not contain all polarization vectors at n points, we would like to find general constraints $G_n^h(a)$ which rule out a function $B_{n;h}(a)$, and that also induct correctly. That is, G_{n+1} constraints imposed on B_{n+1} should imply G_n constraints imposed on B_n. Furthermore, for $a = n - 1$ the constraints should become the set we obtain by imposing $G_{n+1}(n)$ on M_{n+1} in eq (4.3.2).

The expected BCFW behavior will change based on what polarization vectors are missing. Let $\bar{i}$ designate particle i if $B_{n;h}(a)$ is not a function of e_i, ie $e_i \notin E_n^a$. We will show that $B_{n;h}(a)$ functions cannot satisfy the following general $\mathcal{G}_n^h(a)$ constraints:

- i and j adjacent:

    - $[i, j\rangle, [j, i\rangle \propto z^{-1}$, $\tag{4.3.5a}$

    - $[i, \bar{j}\rangle \propto z^{-1}$, $\tag{4.3.5b}$

    - $[\bar{i}, j\rangle \propto z^1$, $\tag{4.3.5c}$

77

$$- \; [\bar{i}, \bar{j}\rangle \propto z^0 \text{ and } [\bar{j}, \bar{i}\rangle \propto z^1 \; , \tag{4.3.5d}$$

- i and j non-adjacent: a z^m from above becomes z^{m-1} , $\qquad$ (4.3.6)

- shifts containing particle h: a z^m from above becomes z^{m+1} . $\qquad$ (4.3.7)

The cases (4.3.5a)-(4.3.5d) pertain to the polarization structure, while the modifications (4.3.6) and (4.3.7) are related to the pole structure.

It can be verified explicitly that at four points these constraints rule out all functions:

$$
\begin{aligned}
a = 0 &: B_{4,h}(p^6) \,, \\
a = 1 &: B_{4;h}(E_4^1, p^5) \,, \\
a = 2 &: B_{4;h}(E_4^2, p^4) \,, \\
a = 3 &: B_{4;h}(E_4^3, p^3) \,,
\end{aligned}
\tag{4.3.8}
$$

with $h = 1$ and $h = 4$. To be specific, $a = 1$ functions include $B(e_1, p^5)$, $B(e_2, p^5)$ and so on, while functions with $a = 2$ include $B(e_1, e_2, p^4)$, $B(e_1, e_3, p^4)$, and so on. Now we move to the inductive step, and assume that B_n functions (4.3.4) are indeed ruled out by the n-point constraints. Then we must show this implies the higher point B_{n+1} functions are ruled out by the $(n+1)$-point constraints. The $(n+1)$-point versions of the functions (4.3.4) have the form:

$$B_{n+1}(E_{n+1}^a, p^{2n-a})\delta_{n+1}, \; a = \overline{0, n} \,. \tag{4.3.9}$$

However, a function B_{n+1} is not necessarily a function of e_{n+1}, just like not all functions (4.3.8) contain e_4. The absence of e_{n+1} changes the form of the soft limit, so we must treat each case separately.

Functions with $e_{n+1} \in E^a_{n+1}$

If $e_{n+1} \in E^a_{n+1}$, the functions (4.3.9) can be written as:

$$B_{n+1}(a) \equiv B_{n+1}(\{E^a_n, e_{n+1}\}, p^{2n-a-1})\delta_{n+1}, \; a = \overline{0, n-1}. \tag{4.3.10}$$

Again let $e_{n+1} = e$. The soft expansion of a function (4.3.10) has the same form of (4.3.2):

$$B_{n+1}(a) \rightarrow \sum_{h=1,n} \frac{1}{q.p_h} \left(\sum_r e.e_r B^r_h(a-1) + \sum_r e.p_r C^r_h(a) \right). \tag{4.3.11}$$

Therefore a function $B_{n+1}(a)$ vanishes only if all functions $B_h(a)$ and $B_h(a-1)$ also vanish, explaining why we needed to consider the whole tower of functions in (4.3.4).

It is easy to see that, because of the different denominators, functions in one pole do not mix with functions in the other pole under any shifts except $[1, n\rangle$ and $[n, 1\rangle$, so we can treat the numerators as independent. The $[1, n\rangle$ and $[n, 1\rangle$ shifts will be discussed separately.

We want to show that $G^{n+1}_{n+1}(a)$ constraints on $B_{n+1}(a)$ imply $G^h_n(a)$ constraints on $B_{n;h}(a)$, or in other words that any shift $[i, j\rangle$ inducts appropriately. Consider a function B^k_n corresponding to the $q.p_n$ pole in eq. (4.3.11), with $i, j \neq k$. Then we can write out the numerator of $q.p_n$ pole term:

$$\sum_{r \neq i,j,k} e.e_r B^r_n + e.e_i B^i_n + e.e_j B^j_n + e.e_k B^k_n$$

$$+ \sum_{r \neq i,j,k} e.p_r C^r_n + e.p_i C^i_n + e.p_j C^j_n + e.p_k C^k_n. \tag{4.3.12}$$

Now assume that both B_{n+1} and B^k_n are functions of e_i and e_j. Then this shift belongs to case (4.3.5a) for both functions. We must show that if $B_{n+1} \propto z^{-1}$ under this shift,

this implies that B_n^k must have the same behavior. We can express this condition as:

$$[i,j\rangle[B_{n+1}] \propto z^{-1} \Rightarrow [i,j\rangle[B_n^k] \propto z^{-1}. \tag{4.3.13}$$

To see this is the case, we apply the shift to eq. (4.3.12):

$$z^{-1} \propto \sum_{r \neq i,j,k} e.e_r B_n^r + e.e_i B_n^i + (e.e_j + ze.p_i \frac{e_i.p_j}{p_i.p_j}) B_n^j + e.e_k B_n^k$$

$$+ \sum_{r \neq i,j,k} e.p_r C_n^r + (e.p_i + ze.e_i) C_n^i + (e.p_j - ze.e_i) C_n^j + e.p_k C_n^k. \tag{4.3.14}$$

The prefactor $e.e_k$ remains unique so B_n^k cannot cancel against any of the other functions, and so must carry the same z^{-1} behavior as B_{n+1}.

Similar reasoning can be applied for all the other cases. For some shifts however, such as $[k,i\rangle$, which is case (4.3.5a) for B_{n+1}, but becomes a $[\bar{k},i\rangle$ case (4.3.5c) for B_n^k, the argument will not be so simple: the shift mixes several functions together. The constraint can luckily be disentangled, and we can still obtain a (weaker, but necessary) constraint for B_n^k. This case is treated in appendix A.1, and all the others can be derived using identical arguments.

Next, we have to show that the modifications due to the pole structure also induct correctly. This is easy to see, since shifts involving 1 or n also shift the $q.p_1$ and $q.p_n$ poles in eq. (4.3.11), contributing one power of z in the denominator. Therefore these shifts weaken any constraints found above by one power of z. Finally, we have to show that the $[1,n\rangle$ shift also transforms accordingly. This is covered in appendix A.2.

Therefore all $B_{n;h}$ functions in (4.3.11) vanish, and then the same reasoning can be applied for the C_n^k functions, which will also vanish. This proves that all functions of the type $B_{n+1}(E_{n+1}^a)$ with $e_{n+1} \in E_{n+1}^a$ vanish.

Functions with $e_{n+1} \notin E^a_{n+1}$

In this case $E^a_{n+1} = E^a_n$ and we must consider all functions of the type:

$$B_{n+1}(a) = B_{n+1}(E^a_n, p^{2n-a}), \ a = \overline{0, n} \,. \tag{4.3.15}$$

In the absence of e_{n+1}, the soft limit is given by a simpler expression:

$$B_{n+1}(a) \to \sum_{h=1,n} \frac{1}{q.p_h} B_{n;h}(E^a_n, p^{2n-a}) \,, \tag{4.3.16}$$

but now the functions on the right side are not of the type (4.3.4): they have two extra powers of momenta. We can distinguish between functions with $a < n$ and $a = n$, which we denote:

$$B'(a) \equiv B_{n;h}(E^a_n, p^{2n-a}) \,, \tag{4.3.17}$$

$$B' \equiv B_{n;h}(E^n_n, p^n) \,. \tag{4.3.18}$$

For $a < n$, the functions can be written in terms of the previous functions (4.3.4), as $B'(a) = B(a) \sum a_{ij} p_i p_j$, since they have more momenta than polarization vectors. Then it is easy to show that if $B(a)$ functions are ruled out by $G^h_n(a)$ constraints, so must the $B'(a)$ functions. The details of this proof are given in appendix A.3.

Finally, the functions with $a = n$, B', cannot be expressed in terms of $B_n(E^n_n, p^{n-2})$. However, the higher point version of such functions, $B_{n+1}(E^{n+1}_{n+1}, p^{n+1})$, is always a function of e_{n+1}, so if we use the soft limit:

$$B_{n+1}(E^{n+1}_{n+1}, p^{n+1}) \to \sum_{h=1,n} \frac{1}{q.p_h} \left(\sum e.e_r B^r_h(E^{n-1}_n, p^{n+1}) + e.p_r C^r_h(E^n_n, p^n) \right) \,, \tag{4.3.19}$$

we are guaranteed to land only on functions which were already shown to vanish. The B functions are of type (4.3.17) with $a = n-1$ and were shown to vanish, while the C

81

functions are just the original n-point functions (4.3.18), and vanish by assumption. Therefore B_{n+1} functions (4.3.15) with $e_{n+1} \notin E_{n+1}^a$ also vanish (at the leading level) under the corresponding constraints.

In conclusion, all possible types of B_{n+1} functions vanish under G_{n+1}, concluding the inductive proof that all functions $B_{n;h}(E_n^a, p^{2n-a-2})$ vanish under $\mathcal{G}_n^h(E_n^a)$ constraints, including those in our original eq. (4.3.2).

4.3.3 Fixing $C(E_n^n)$ functions

Once the B functions in (4.3.2) have been shown to vanish, using the same arguments as before it is easy to see that the C functions must satisfy $\mathcal{G}_n^h(E_n^n)$ constraints. But by assumption functions of the type $C_{n;h}(E_n^n, p^{n-2})$ are uniquely fixed by these constraints, up to some numerical coefficient. Therefore we obtain $C_1^i = a_i A_n$ and $C_n^i = b_i A_n$, and eq. (4.3.2) becomes:

$$M_{n+1}^{-1} = \sum_{i \neq 3} \left(a_i \frac{e.p_i}{q.p_1} + b_i \frac{e.p_i}{q.p_n} \right) A_n \, . \tag{4.3.20}$$

Now we exploit the fact that due to our choice of imposing momentum conservation, p_3 is not present in the sums above. Consider the $q.p_1$ pole term first. Under a shift $[i,3\rangle$, $i \neq 1$, only the prefactor $e.p_i$ is affected, $e.p_i \to e.p_i - ze.e_i$. Both M_{n+1} and A_n have the same scaling $\mathcal{O}(z^m)$ (with $m = -1$ for adjacent and $m = -2$ for non-adjacent), and we obtain:

$$\mathcal{O}(z^m) \propto M_{n+1}^{-1} = \frac{\sum_j a_j e.p_j}{q.p_1} \times A_n \propto \frac{\sum_j a_j e.p_j - za_i e.e_i}{q.p_1} \times \mathcal{O}(z^m) = \mathcal{O}(z^m) + a_i \mathcal{O}(z^{m+1}) \, ,$$
$$\tag{4.3.21}$$

which implies $a_i = 0$, for $i \neq 1$. The coefficient a_1 is not ruled out by the same trick, since under $[1, 3\rangle$ the pole is also shifted:

$$\mathcal{O}(z^m) \propto M_{n+1}^{-1} = a_1 \frac{e.p_1}{q.p_1} \times A_n \propto a_i \frac{e.p_1 + ze.e_1}{q.p_1 + zq.e_1} \times \mathcal{O}(z^m) \propto a_1 \mathcal{O}(z^m). \qquad (4.3.22)$$

Similarly we obtain that $b_i = 0$ for $i \neq n$, and we are left with:

$$M_{n+1}^{-1} = \left(a_1 \frac{e.p_1}{q.p_1} + b_n \frac{e.p_n}{q.p_n} \right) A_n . \qquad (4.3.23)$$

Finally, under the special $[1, n\rangle$ shift, which crucially is adjacent in A_n but non-adjacent in M_{n+1}:

$$\mathcal{O}(z^{-2}) \propto M_{n+1}^{-1} = \left(a_1 \frac{e.p_1}{q.p_1} + b_n \frac{e.p_n}{q.p_n} \right) \times A_n \propto \left((a_1 + b_n) \frac{e.e_1}{q.e_1} + \mathcal{O}(z^{-1}) \right) \times \mathcal{O}(z^{-1})$$

$$\propto (a_1 + b_n) \mathcal{O}(z^{-1}) + \mathcal{O}(z^{-2}), \quad (4.3.24)$$

so $a_1 = -b_n$, and we obtain Eq. (4.3.3):

$$M_{n+1}^{-1} = \left(\frac{e.p_1}{q.p_1} - \frac{e.p_n}{q.p_n} \right) A_n , \qquad (4.3.25)$$

completing the leading order proof.

4.4 Future directions

In this article we have provided the leading order step in the proof that locality and correct behavior under BCFW shifts uniquely fix the Yang-Mills amplitude. It is very likely that the subleading terms can be treated in the same way, but finding a more direct proof would be far more rewarding. The most direct approach would be to show that large z BCFW shifts are somehow related to gauge invariance. Then

the proof in Ref. [22] would immediately apply. But this connection would be very surprising in its own right. For instance, it might help explain why Yang-Mills and gravity have this surprising behavior in the first place. One option in this direction would be to use the Cauchy theorem to build the amplitude from different shifts (but without assuming unitarity). It is very peculiar that these shifts are gauge invariant, but non-local, so the object they construct is guaranteed to unfortunately inherit both properties, and thus avoid the locality+gauge invariance argument. Yet somehow, constructing the object from many different shifts must eliminate the non-local terms. Ultimately, this result suggests the notion of "constructability" [52]-[53] might play a more fundamental role, beyond recursion relations.

Another obvious direction is determining whether an equivalent statement holds for gravity. In this case it is likely that demanding the stronger $\mathcal{O}(z^{-2})$ behavior will be required. We suspect this to be the case due to the fact that even though $\mathcal{O}(z^{-1})$ is sufficient for recursion relations to exist, it was discovered that the so called "bonus" behavior of gravity is actually required for full on-shell consistency [52]. In Ref. [26] it was shown that this bonus behavior automatically emerges from Bose-symmetry, but it was unclear whether the logic can be reversed: could $\mathcal{O}(z^{-2})$ behavior imply Bose-symmetry? If the claim of this present article can indeed be extended to gravity, then clearly the answer is yes. And not only would the Bose-symmetry emerge, but (assuming locality) the whole amplitude emerges.

Scalar theories like the non-linear sigma model or Dirac-Born-Infeld are another obvious target. Recently, it was shown that recursion relations can be applied to such theories as well, when the soft behavior is taken into account [54]. On the other hand, in Ref. [22] it was shown that locality and the soft behavior completely fix the amplitude. It would be very interesting to fully work out the interplay between locality, vanishing in the soft limit, and BCFW shifts, in the context of this article.

As a by-product of this investigation, we have also uncovered a new method for computing the full Feynman amplitude via BCFW recursion relations, a method which to our knowledge has not yet been explored. It would be interesting to see what applications might be derived from this new approach.

Acknowledgments

Chapter 5

Consistency conditions on massless S-matrices

5.1 Introduction

Pioneering work by Weinberg showed that simultaneously imposing Lorentz-invariance and unitarity, while coupling a hard scattering process to photons, necessitates *both* charge conservation and the Maxwell equations [55]. Similarly, he showed the same holds for gravity: imposing Lorentz-invariance and unitarity on hard scattering processes coupled to gravitons implies both the equivalence principle and the Einstein equations [56]. Weinberg's theorems were then extended to fermions in [57], where it was shown that spin-3/2 particles lead to supersymmetry. In the case of higher spin theories [58], which are closely related to string theory [59], it shown that unitarity and locality impose severe restrictions, and many no-go theorems were established [60][61], more recently including in CFT's [62][63].

The goal of this paper is to provide a basis for systematic analyses of the leading-order interactions between any set of massless states in four dimensions, within the context of the *on-shell* perturbative S-matrix. In short, our results are: (1) a new

classification of three-particle amplitudes in constructible massless S-matrices, (2) ruling out all S-matrices built from three-point amplitudes with $\sum_{i=1}^{3} h_i = 0$ (other than ϕ^3-theory), (3) a new on-shell proof of the uniqueness of interacting gravitons and gluons, (4) development of a new test on four-particle S-matrices, and (5) showing how supersymmetry naturally emerges from consistency constraints on certain four-particle amplitudes which include spin-3/2 particles.

Massless vectors (gluons and photons) and tensors (gravitons) are naturally described via on-shell methods [20]-[64]. On- and off-shell descriptions of these massless higher-spin states are qualitatively different: on-shell they have only transverse polarization states, while off-shell all polarization states may be accessed. Local field theory descriptions necessarily introduce these unphysical, longitudinal, polarization states. They must be removed through introducing extra constraints which "gauge them away". Understanding consistency conditions on the interactions of massless gravitons and gluons/photons, should therefore stand to benefit from moving more-and-more on-shell, where gauge-invariance is automatic.

"Gauge anomalies" provide a recent example [65][66]. What is called a "gauge-anomaly" in off-shell formulations, on-shell is simply a tension between parity-violation and locality. Rational terms in parity-violating loop amplitudes either do not have local descriptions, or require the Green-Schwarz two-form to restore unitarity to the S-matrix [67].

Along these lines, there are recent, beautiful, papers by Benincasa and Cachazo [20], and by Schuster and Toro [21] putting these consistency conditions more on-shell. Ref. [20] explored the constraints imposed on a four-particle S-matrix through demanding consistent on-shell BCFW-factorization in various channels on the coupling constants in a given theory [2].[1] Four-particle tests based on BCFW have been used to show the inconsistency of some higher spin interactions in Refs. [69][70],

---

[1]This consistency condition is further explored, for instance, in ref. [68].

where it was suggested that non-local objects must be included in order to provide consistent theories (see also [71]).

The analysis of Ref. [20] hinged upon the existence of a "valid" BCF-shift,

$$A_n(z) = A_n(\{p_i + qz, p_j - qz\}) \text{ , with } p_i \cdot q = p_j \cdot q = 0 \text{ and } q \cdot q = 0,$$

of the amplitude that does not have a pole at infinity. If $A(z)$ does not have a large-z pole, then its physical value, $A(0)$, is a sum over its residues at the finite-z poles. These finite-z poles are factorization channels; their residues are themselves products of lower-point *on-shell* amplitudes: an on-shell construction of the whole S-matrix [50][72][73]. However, then extant existence proofs for such shifts resorted to local field theory methods [2][39][74].

Hence, ref. [21] relaxes this assumption, through imposing a generalized notion of unitarity, which they refer to as "complex factorization". The consequences of these consistency conditions are powerful. For example, they uniquely fix (1) the equivalence of gravitational couplings to all matter, (2) decoupling of multiple species of gravitons within S-matrix elements, and (3) the Lie Algebraic structure of spin-1 interactions.

Our paper is organized as follows. We begin in section 5.2 by developing a useful classification of all on-shell massless three-point amplitudes. We here pause to make contact with standard terminology for "relevant", "marginal" and "irrelevant" operators in off-shell formulations of Field Theory, and to review basic tools of the on-shell S-matrix.

This is applied in section 5.3, first, to constructible four-particle amplitudes in these theories. Locality and unitarity sharply constrain the analytic structure of scattering amplitudes. Specifically, four-point amplitudes cannot have more than three poles. Simple pole-counting, using the classification system in section 5.2, rules out

all lower-spin theories, save ϕ^3-theory, (S)YM, and GR/SUGRA–and one pathological example, containing the interaction vertex $A_3(\frac{1}{2}, -\frac{1}{2}, 0)$. In section 5.4, we show that the gluons can only consistently interact via YM, GR and the higher-spin amplitude $A_3(1, 1, 1)$. Similarly, gravitons can only interact via GR and the higher-spin amplitude $A_3(2, 2, 2)$. Gravitons and gluons are unique, and cannot couple to higher-spin states. These two sections strongly constrain the list of possible interacting high-spin theories, in accordance with existing no-go theorems.

Utilizing the information in sections 5.3 and 5.4, in section 5.5 we derive and apply a systematic four-particle test, originally discussed in Ref. [75]. This test independently demonstrates classic results known about S-matrices of massless states, such as the equivalence principle, the impossibility of coupling gravitons to massless states with $s > 2$ [76], decoupling of multiple spin-2 species, and the Lie Algebraic structure of vector self-interactions.

Knowing the equivalence principle, in section 5.6 we then study the consistency conditions of S-matrices involving massless spin-3/2 states. From our experience with supersymmetry, we should expect that conserved fermionic currents correspond to massless spin-3/2 states. In other words, we expect that a theory which interacts with massless spin-3/2 particles should be supersymmetric.

Supersymmetry manifests itself through requiring all poles within four-point amplitudes have consistent interpretations. The number of poles is fixed, mandated by locality and unitarity and the mass-dimension of the leading-order interactions. Invariably, for S-matrices involving external massless spin-3/2 particles, at least one of these poles begs for inclusion of a new particle into the spectrum, as a propagating internal state on the associated factorization channel. For these S-matrices to be consistent, they *require* both gravitons to be present in the spectrum [77] and supersymmetry. We close with future directions in section 5.7.

5.2   Basics of on-shell methods in four-dimensions

In this section, we briefly review three major facets of modern treatments of massless S-matrices: the spinor-helicity formalism, kinematic structure of three-point amplitudes in these theories, and the notion of constructibility. The main message is three-fold:

- The kinematic dependence of three-point on-shell amplitudes is uniquely fixed by Poincare invariance (and is best described with the spinor-helicity formalism).

- On-shell construction methods, such as BCFW-recursion, allow one to recursively build up the entire S-matrix from these on-shell three-point building blocks. Amplitudes constructed this way are trivially "gauge-invariant". There are no gauges.

- Any pole in a local and unitary scattering amplitude must both (a) be a simple pole in a kinematical invariant, e.g. $1/K^2$, and (b) have a corresponding residue with a direct interpretation as a factorization channel of the amplitude into two sub-amplitudes.

5.2.1   Massless asymptotic states and the spinor-helicity formalism

In a given theory, scattering amplitudes can only be functions of the asymptotic scattering states. Relatively few pieces of information are needed to fully characterize an asymptotic state: momentum, spin, and charge/species information. Spinor-helicity variables automatically and fully encode both momentum and spin information for massless states in four-dimensions.

Four-dimensional Lorentz vectors map uniquely into bi-spinors, and vice versa (the mapping is bijective): $p_{\alpha\dot\alpha} = p_\mu \sigma^\mu_{\alpha\dot\alpha}$. Determinants of on-shell momentum bi-spinors are proportional to m^2. Bi-spinors of massless particles thus have rank-1, and must factorize into a product of a left-handed and a right-handed Weyl spinor: $p^2 = 0 \Rightarrow p^{\alpha\dot\alpha} = \lambda^\alpha \tilde\lambda^{\dot\alpha}$.

These two Weyl spinors λ and $\tilde\lambda$ are the spinor-helicity variables, and are uniquely fixed by their corresponding null-momentum, p, up to rescalings by the complex parameter z: $(\lambda, \tilde\lambda) \to (z\lambda, \tilde\lambda/z)$. Further, they transform in the $(1/2,0)$ and $(0,1/2)$ representations of the Lorentz group. Dot products of null momenta have the simple form, $p_i \cdot p_j = \langle ij \rangle [ji]$, where the inner-product of the (complex) LH-spinor-helicity variables, is $\langle AB \rangle \equiv \lambda^A_\alpha \lambda^B_\beta \epsilon^{\alpha\beta}$, and the contraction of the RH- Weyl spinors is $[AB] \equiv \tilde\lambda^A_{\dot\alpha} \tilde\lambda^B_{\dot\beta} \epsilon^{\dot\alpha\dot\beta}$.

A good deal of the power of the spinor-helicity formalism derives from the dissociation between the left-handed and right-handed degrees of freedom. Real null-momenta are defined by the relation,

$$\tilde\lambda = \bar\lambda \tag{5.2.1}$$

between the two Weyl-spinors. Complex momenta are not similarly bound: the left-handed and right-handed Weyl-spinors need not be related for complex momentum. For this reason, they can be independently deformed by complex parameters; this efficiently probes the analytic properties of on-shell amplitudes that depend on these variables. From here on out, we refer to the left-handed Weyl-spinors, i.e. the λs, as *holomorphic* variables; right-handed Weyl-spinors, i.e. the $\tilde\lambda$s are referred to as *anti-holomorphic* variables. Similarly, *holomorphic* spinor-brackets and *anti-holomorphic* spinor-brackets refer to $\langle \lambda, \chi \rangle$- and $[\tilde\lambda, \tilde\chi]$-contractions.

Identifying the ambiguity $(\lambda, \tilde{\lambda}) \rightarrow (z\lambda, \tilde{\lambda}/z)$ with little-group (i.e. helicity) rotations, $(\lambda, \bar{\lambda}) \rightarrow (e^{-i\theta/2}\lambda, e^{i\theta/2}\bar{\lambda})$, allows one to use the spinor-helicity variables to express not only the momenta of external states in a scattering process, but also their spin (helicity). In other words, the spinor-helicity variables encode all of the data needed to characterize massless asymptotic states, save species information.

5.2.2 Three-point amplitudes

Scattering processes involving three massless on-shell states have no non-trivial kinematical invariants. At higher-points, complicated functions of kinematical invariants exist that allow rich perturbative structure at loop level. These invariants are absent at three-points. Poincare invariance, up to coupling constants, thus uniquely and totally fixes the kinematical structure of all three-point amplitudes for on-shell massless states.

The standard approach to solving for the three-point amplitudes (see for example Ref. [20]) involves first writing a general amplitude as:

$$A_3 = A_3^{(\lambda)}(\langle 12 \rangle, \langle 23 \rangle, \langle 31 \rangle) + A_3^{(\tilde{\lambda})}([12], [23], [31]) \qquad (5.2.2)$$

where (λ) denotes exclusive dependence on holomorphic spinors, and $(\tilde{\lambda})$ denotes the same for anti-holomorphic spinors. Imposing momentum conservation forces $[12] = [23] = [31] = 0$, and/or $\langle 12 \rangle = \langle 23 \rangle = \langle 31 \rangle = 0$. Typically, only one of the two functions in Eq. (5.2.2) is smooth in this limit, and is thus selected as the physical one, while the other is discarded.

Explicitly, in these cases, the amplitudes become:

$$A_3(1_a^{h_1}, 2_b^{h_2}, 3_c^{h_3}) = g_{abc}^- \langle 12 \rangle^{h_3 - h_1 - h_2} \langle 23 \rangle^{h_1 - h_2 - h_3} \langle 31 \rangle^{h_2 - h_3 - h_1} \, , \text{ for } \sum_{i=1}^{3} h_i < 0 \, ,$$

$$A_3(1_a^{h_1}, 2_b^{h_2}, 3_c^{h_3}) = g_{abc}^+ [12]^{h_1 + h_2 - h_3} [23]^{h_2 + h_3 - h_1} [31]^{h_3 + h_1 - h_2} \quad , \text{ for } \sum_{i=1}^{3} h_i > 0 \, , \quad (5.2.3)$$

where $g_{abc}^{\pm}$ is the species dependent coupling constant.

However, this approach leads to ambiguities in the $\sum_{i=1}^{3} h_i = 0$ case. Consider for example a three-point interaction between two opposite-helicity fermions and a scalar. Equation (5.2.2) reads in this case:

$$A_3 \left(1^0, 2^{-\frac{1}{2}}, 3^{\frac{1}{2}} \right) = g^- \frac{\langle 12 \rangle}{\langle 13 \rangle} + g^+ \frac{[13]}{[12]} \quad (5.2.4)$$

Imposing momentum conservation, for example by setting $\langle 12 \rangle = \langle 23 \rangle = \langle 31 \rangle = 0$, is clearly ill-defined[2]. Because of this ambiguity, $\sum_{i=1}^{3} h_i = 0$ amplitudes have generally been ignored in most of the on-shell literature. However, the ambiguity is only superficial.

The inconsistencies arise because we first find the most general eigenfunction of the helicity operator, ie. Eq. (5.2.2), and only after that do we impose momentum conservation. However, this order of operations is arbitrary. Since we always only deal with on-shell amplitudes, we can simply first fix for example $\langle 12 \rangle = \langle 23 \rangle = \langle 31 \rangle = 0$, and then look for solutions which are functions only of $\tilde{\lambda}$s. In this case, the amplitudes

---

[2]Attempting to impose momentum conservation by a well-defined limit leads to other inconsistencies as well, for example with the helicity operator.

are perfectly well defined as:

$$A_3 = g^-_{abc} f^-(\lambda_i), \quad \text{when } [12] = [23] = [31] = 0 \tag{5.2.5}$$

and

$$A_3 = g^+_{abc} f^+(\tilde{\lambda}_i), \quad \text{when } \langle 12 \rangle = \langle 23 \rangle = \langle 31 \rangle = 0 \tag{5.2.6}$$

Ultimately, it will in fact turn out that none of these amplitudes are consistent with locality and unitarity, but this approach clears any ambiguities related to $\sum_{i=1}^3 h_i = 0$ amplitudes.

Before moving on, we pause to consider the role of parity in the on-shell formalism. Parity conjugation swaps the left-handed and right-handed $SU(2)$s that define the (double-cover) of the four-dimensional Lorentz-group. As such, parity swaps the left-handed Weyl-spinors with the right-handed Weyl-spinors, $(1/2, 0) \leftrightarrow (0, 1/2)$. Therefore, within the spinor-helicity formalism, in the context of Eq. (5.2.3),

$$g^-_{abc} = g^+_{abc} \Longleftrightarrow \text{Parity} - \text{conserving interactions , and} \tag{5.2.7}$$

$$g^-_{abc} = -g^+_{abc} \Longleftrightarrow \text{Parity} - \text{violating interactions .}$$

Further, as we associate the right-handed Weyl-spinors, i.e. the λs, with holomorphic degrees of freedom and the left-handed Weyl-spinors, i.e. the $\tilde{\lambda}$s, with anti-holomorphic degrees of freedom, we see that parity-conjugation swaps the holomorphic and anti-holomorphic variables. In other words, parity- and complex- conjugation are one-and-the-same. The conjugate of a given three-point amplitude is the same amplitude with all helicities flipped: the "conjugate" of $A_3(1^{+h_1}, 2^{+h_2}, 3^{+h_3})$ is $A_3(1^{-h_1}, 2^{-h_2}, 3^{-h_3})$.

We will find it useful to classify all such three-particle amplitudes by two numbers:

$$A = \left| \sum_{i=1}^{3} h_i \right|, \ H = \max\left\{ |h_1|, |h_2|, |h_3| \right\}. \tag{5.2.8}$$

Comparing the relevant operator in ϕ^3-theory to its corresponding primitive three-point amplitude, we infer that three-point amplitudes with $A = 0$ correspond to *relevant* operators. Similarly, QCD's $A = 1$ three-point amplitude corresponds to *marginal* operators; GR has $A = 2$, and interacts via irrelevant, $1/M_{pl}$ suppressed, operators.

5.2.3 Four points and higher: Unitarity, Locality, and Constructibility

There are several, complimentary, ways to build up the full S-matrix of a theory, given its fundamental interactions. Conventionally, this is through Feynman diagrams, the work-horse of any perturbative analysis of a given field theory. However, this description of massless vector- (and higher-spin-) scattering via local interaction Lagrangians necessarily introduces unphysical, longitudinal, modes into intermediate expressions [78][75]. To project out these unphysical degrees of freedom, one must impose the gauge conditions.

On the other hand, recent developments have elucidated methods to obtain the full S-matrix, while keeping *all* states involved on-shell (and physical) throughout the calculation [20][21][64][2][72][79]. We refer to these methods, loosely speaking, as "constructive". Crucially, because all states are on-shell, all degrees of freedom are manifest, thus: *amplitudes that are directly constructed through on-shell methods are automatically gauge-invariant.* This simple fact dramatically increases both (a) the computational simplicity of calculations of scattering amplitudes, and (b) the physical transparency of the final results.

The cost is that amplitudes sewn together from on-shell, delocalized, asymptotic states do not appear to be manifestly local. Specifically, at the level of the amplitude, locality is reflected in the pole-structure of the amplitude. Scattering amplitudes in local theories have exclusively propagator-like, $\sim 1/K^2$, poles ($K = \sum_i p_i$ is a sum of external null momenta). Non-local poles correspond to higher-order poles, i.e. $1/(K^2)^4$, and/or poles of the form, $1/\langle i|K|j]$, where K is a sum of external momenta.[3] An on-shell S-matrix is local if its only kinematical poles are of the form $1/(\sum_i p_i)^2$.

Unitarity, as well, has a slightly different incarnation in the on-shell S-matrix. In its simplest guise, unitarity is simply the dual requirement that (a) the residue on each and every pole in an amplitude *must* have an interpretation as a physical factorization channel,

$$
A^{(n)} \to \frac{1}{K^2} A_L^{(n-m+1)} \times A_R^{(m+1)} , \tag{5.2.9}
$$

and (b) that any individual factorization channel, if it is a legitimate bridge between known lower-point amplitudes in the theory, must be a residue of a fully legitimate amplitude with the same external states in the theory. For example, given a factorization channel of the form $A_3(1^{-2}, 2^{-2}, P_{12}^{+2})\frac{1}{s_{12}}A_3(P_{12}^{-2}, 3^{+2}, 4^{+2})$, within a theory *constructed* from the three-point amplitude $A_3(+2, -2, +2)$ and its parity-conjugate, *then* this must be a factorization channel of the four-point amplitude $A_4(1^{-2}, 2^{-2}, 3^{+2}, 4^{+2})$.

Poincare invariance uniquely fixes the three-particle S-matrix in a theory, up to coupling constants. Constructive methods, such as the BCFW recursion relations, use these fixed forms for the three-point amplitudes as input to build up the entire S-matrix, without making reference to Feynman diagrams [2][72]. Basic symmetry con-

---

[3]Indeed, individual terms within gluon amplitudes generated by, for instance, BCFW-recursion[50][2] contain "non-local" poles, specifically of this second type, $\sim 1/\langle i|K|j]$. These non-local poles, however, always cancel in the total sum, and the final expression is manifestly local [78][72].

siderations, residue theorems, and judicious application of tree-level/single-particle unitarity, fix the entire S-matrix![4]

Before closing, we motivate the most famous on-shell construction of massless scattering amplitudes: BCFW-recursion. In it, two null external momenta, p_1^μ and p_2^μ, are deformed by a complex null-momentum, $z \times q^\mu$. The shift is such that (a) the shifted momenta $p_1(z) = p_1 + qz$ and $p_2(z) = p_2 - qz$ remain on-shell (possible, as momentum q^μ is complex), and (b) the total sum of external momenta remains zero.

As tree amplitudes are rational functions of their external kinematical invariants with, at most, simple poles, this deformation allows one to probe the analytic pole structure of the deformed amplitude, $A^{\text{tree}}(z)$:

$$A^{\text{tree}}(z) \equiv A^{\text{tree}}(p_1^{h_1}(z), p_2^{h_2}(z), ...p_n^{h_n}). \qquad (5.2.10)$$

Kinematical poles in $A^{\text{tree}}(z)$ are either un-shifted, or scale as $1/K^2 \to 1/(2z(q \cdot K) + K^2)$, if K includes only one of $\hat{p}_1$ or $\hat{p}_2$. Cauchy's theorem then gives a simple expression for the physical amplitude, $A^{\text{tree}}(z = 0)$,

$$A^{\text{tree}}(z = 0) = \sum_{z_P} \text{Res}\left\{ \frac{A_4(z)}{z} \right\}\bigg|_{z_P = -\frac{K^2}{2q \cdot K}} + \left(\text{Pole at } z \to \infty \right). \qquad (5.2.11)$$

Existence of such a BCFW-shift, in both Yang-Mills/QCD and in General Relativity, that dies off at least as quickly as $1/z$ for large-z can be elegantly shown through imposing complex factorization[21], and allows the entire on-shell S-matrix to be built up from three-point amplitudes. Existence of valid BCFW-shifts were originally shown within local formulations of field theory [2][80][39]. In section 5.5 we develop a shift at four-points which is guaranteed to die off for large-z by simple dimensional analysis.

---

[4]Invocations of "unitarity" in this paper do not refer to the standard two-particle unitarity-cuts.

5.3 Ruling out constructible theories by pole-counting

Pedestrian counting of poles, mandated by constructibility in four-point amplitudes, strongly constrains on-shell theories. The number of poles in an amplitude *must* be less than or equal to the number of accessible, physical, factorization channels at four points. Tension arises, because the requisite number of poles in a four-point amplitude *increases* with the highest-spin particle in the theory, while the number of possible factorization channels is *bounded* from above by three, the number of Mandelstam variables.

This tension explicitly rules out the following theories as inconsistent with constructibility, locality, and unitarity: (1) all relevant interactions (A = 0), save ϕ^3 and an "exotic" Yukawa-like interaction, (2) all marginal interactions (A = 1) save those in YM, QCD, Yukawa theory, and scalar QED, and another "exotic" interaction between spin-3/2 particles and gluons, and (3) all first-order irrelevant interactions (A = 2) save those in GR. Further consistency conditions later rule out those two unknown, pathological, relevant ($A = 0$) and marginal ($A = 1$) interactions.

Further, incrementally more sophisticated pole-counting sharply constrains highly irrelevant ($A > 2$) higher-spin amplitudes. Specifically, save for two notable examples, they cannot consistently couple either to gravitational interactions or to more conventional Yang-Mills theories or "gauge"-theories. This is the subject of section 5.4. It is somewhat striking that simply counting poles in this way so powerfully constrains the palate of three-point amplitudes which may construct local and unitary S-matrices. The results of this pole-counting exercise are succinctly summarized in Fig. 5.1.

5.3.1 The basic consistency condition

Explicitly we find that four-particle S-matrices constructed from primitive three-particle amplitudes are inconsistent with locality and unitarity if there are more than three poles in any given term in an amplitude. More specifically, the number of poles in the simplest amplitudes has to be *at least* $N_p = 2H + 1 - A$. Thus, a theory is necessarily inconsistent if

$$2H + 1 - A = N_p > 3 \iff \text{Number of poles} > \text{cardinality of } \{s, t, u\}. \qquad (5.3.1)$$

Recall that, in accordance with Eq. (5.2.8), $A = |h_1 + h_2 + h_3|$ and $H = \max\{|h_1|, |h_2|, |h_3|\}$.

We prove constraint (5.3.1) below.[5] First, we note there are A total spinor-brackets in three-point amplitudes of the type in Eq. (5.2.3):

$$A_3(1^{h_1}, 2^{h_2}, 3^{h_3}) = \kappa_A [12]^c [13]^b [23]^a \Rightarrow a + b + c = \sum_{i=1}^{3} h_i = A > 0 . \qquad (5.3.2)$$

Thus, on a factorization channel of a four-point amplitude, A_4, *constructed* from a given three-point amplitude multiplied by its parity conjugate amplitude, $A_3 \times \bar{A}_3$, there will be A net holomorphic spinor-brackets and A net anti-holomorphic spinor-brackets: $A \langle \, \rangle$s and A [ ]s. Therefore, generically on such a factorization channel, the mass-squared dimension of the amplitude is:

$$A_4 \to \frac{\kappa_A^2}{s_{\alpha\beta}} A_3 \times \bar{A}_3 \Rightarrow \left[\frac{A_4}{\kappa_A^2} \right] = (K^2)^{A-1} . \qquad (5.3.3)$$

By locality, an amplitude may only have $1/K^2$-type poles. Therefore the helicity information, captured by the non-zero little-group weight of the spinor-products, can only be present in an overall *numerator* factor multiplying the amplitude. Four-point

---

[5] For expediency, we defer discussion of one technical point, proof of Eq. (5.3.5), to appendix A.4.

amplitudes thus naturally split into three parts: a numerator, N, which encodes helicities of the states, a denominator, $F(s,t,u)$, which encodes the pole-structure, and the coupling constants, κ_A^2, which encode the species-dependent characters of the interactions (discussed in section 5.5):

$$A_4 = \kappa_A^2 \frac{N}{F(s,t,u)} \Rightarrow \left[\frac{N}{F(s,t,u)}\right] = (K^2)^{A-1} \,, \tag{5.3.4}$$

where the last equality is inferred from Eq. (5.3.3). We prove in Appendix A.4, that *minimal* numerators N which accomplish this goal are comprised of exactly $2H$ holomorphic and $2H$ anti-holomorphic spinor-brackets, *none* of which can cancel against any pole in $F(s,t,u)$:

$$N \sim \langle\,\rangle_{(1)}...\langle\,\rangle_{(2H)} \;[\,]_{(1)}...[\,]_{(2H)} \Rightarrow [N] = (K^2)^{2H} \,. \tag{5.3.5}$$

Thus, by (5.3.3), (5.3.4), and (5.3.5), we see

$$\left[\frac{A_4}{\kappa_A^2}\right] = \left[\frac{N}{F(s,t,u)}\right] = (K^2)^{A-1} \,, \text{ and } [N] = (K^2)^{2H}$$

$$\Rightarrow [F(s,t,u)] = (K^2)^{2H+1-A}$$

$$\Rightarrow N_p = 2H + 1 - A \,. \tag{5.3.6}$$

Constraint (5.3.1) naturally falls out from Eq. (5.3.6), after observing that there can be at most three legitimate, distinct, factorization channels in any four-point tree amplitude. This specific constraint, and others arising from pole-counting from minimal numerators, is extremely powerful. The catalogue of theories they together rule out are succinctly listed in Fig. 5.1. We explore the consequences of this constraint below.
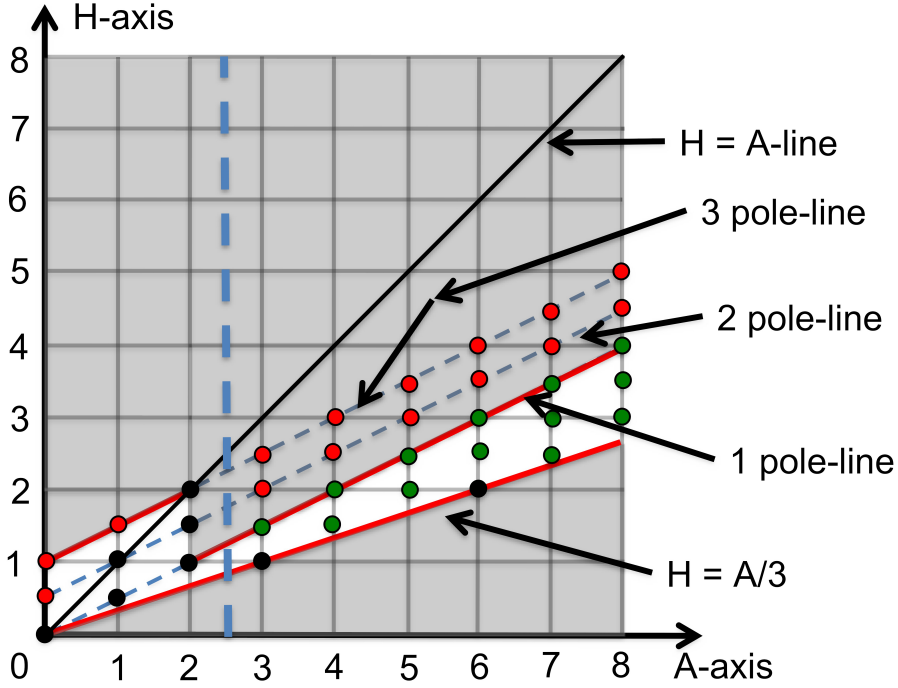
Figure 5.1: Summary of pole-counting results. Recall $N_p = 2H + 1 - A$, where $A = |\sum_{i=1}^{3} h_i|$ and $H = \max\{|h_i|\}$. (Color online.) In short: black-dots represent sets of three-point amplitudes that define self-consistent S-matrices that can couple to gravity; green-dots represent sets of three-point amplitudes which—save for two exceptions explicitly delineated in Eq. (5.4.6)—define S-matrices that *cannot* couple (in the sense defined in section 5.4) to any S-matrix defined by the black-dots; red-dots represent sets of three-point amplitudes that cannot ever form consistent S-matrices. Straightforward application of constraint (5.3.1), in subsection 5.3.2, rules out all A_3s with (H, A)-*above* the $N_p = 3$-line. More careful pole-counting, in subsection 5.3.3 and appendix A.5, rules out all interactions above the $N_p = 1$ line, save for those with $(H, A) = (1/2, 0)$, $(1, 1)$, $(3/2, 2)$, and $(2, 2)$. Further, in section 5.4, a modified pole-counting rules out interaction between the $(H, A) = (2, 2)$ gravity theory and any other theory with a spin-2 particle, save the unique $(H, A) = (2, 6)$-theory. Similar results hold for gluon self-interactions: vectors present in any higher-spin amplitude with $A > 3$, save the unique $(H, A) = (1, 3)$-theory, cannot couple to the vectors interacting via leading $(H, A) = (1, 1)$ interactions. Section 5.5.3 rules out the $(H, A) = (1/2, 0)$-interaction. Amplitudes in the grey-shaded regions can never be consistent with locality and unitarity. Higher-spin, $A > 3$, amplitudes between the $H = A/2$ and $A = A/3$ lines may be consistent. However, they *cannot* be coupled either to GR or YM, save for $(H, A) = (1, 3)$ or $(2, 6)$. In section 5.6, we show inclusion of leading-order interactions between massless spin-3/2 states, at $A = 2$, promotes gravity to supergravity. Supergravity cannot couple to even these two $A > 2$ interactions, as seen in appendix A.7.

101

5.3.2 Relevant, marginal, and (first-order) irrelevant theories ($A \leq 2$): constraints

To begin with, note that constraint (5.3.1) immediately rules out all theories with $N_p > 3$. Beginning with relevant, $A = 0$, interactions, we see that $N_p = 2H + 1 \leq 3 \Rightarrow H \leq 1$. Already this rules out relevant interactions between massless spin-3/2 and spin-2 states.

Next, argument by contradiction rules out relevant amplitudes involving massless vectors, i.e. the $(H, A) = (1, 0)$-theories wh. Consider such a relevant amplitude, for example $A_3(+1, -1/2, -1/2)$. It constructs a putative four-point amplitude with external vectors,

$$A_4\left(-1, -\frac{1}{2}, \frac{1}{2}, 1\right). \tag{5.3.7}$$

This amplitude must have $2 + 1 - 0 = 3$ poles, each of which must have an interpretation as a *valid* factorization channel of the amplitude. So it must have valid s-, t-, and u-factorization channels, with *relevant* ($A = 0$) three-point amplitudes on either side. However, on the $s \to 0$ pole, A_4 factorizes as,

$$A_4\left(-1, -\frac{1}{2}, \frac{1}{2}, 1\right)\bigg|_{s\to 0} = \frac{1}{s}\bar{A}_3\left(-1, -\frac{1}{2}, h\right) A_3\left(1, \frac{1}{2}, -h\right). \tag{5.3.8}$$

where h must be 3/2 to make the interaction relevant. Thus, to be consistent with locality and unitarity, relevant vector couplings must also include spin-3/2 particles. But, as mentioned above, including these particles in the spectrum, and then taking them as external state invariably leads to too many poles. An identical argument shows that the remaining relevant vertex $A_3(+1, -1, 0)$ requires spin 2 particles, again leading to an inconsistency. Thus all $(H, A) = (1, 0)$ interactions are also ruled out.

Thus the only admissible relevant three-point amplitudes are

$$A_3(0,0,0) \;,\; \text{and}\; A_3\left(0, \frac{1}{2}, -\frac{1}{2}\right)\;. \tag{5.3.9}$$

The first amplitude is the familiar one from ϕ^3-theory. We rule out the second amplitude in section 5.5.3.

Further, we see that marginal interactions cannot contain particles with helicities larger than $3/2$. Directly, requiring $2H + 1 - A \leq 3$ for $A = 1$ forces $H \leq 3/2$. $(H, 1)$-type three-point amplitudes cannot build S-matrices consistent with locality and unitarity for $H > 3/2$.

We rule out *marginal* $(H, A) = (3/2, 1)$ amplitudes, i.e. marginal coupling to massless spin-$3/2$ states, using the same logic as above. This time, marginal amplitudes with external $3/2$ particles require *all* three poles. Two factorization channels have consistent interpretations within the theory; however, the "third" channel does not. It necessitates exchange of a spin-2 state between the three-point amplitudes. But this violates constraint (5.3.1): marginal amplitudes with spin-2 states lead to amplitudes with four kinematic poles. Thus, the only admissible marginal three-point amplitudes are,

$$A_3\left(1, 1, -1\right)\;,\; A_3\left(1, \frac{1}{2}, -\frac{1}{2}\right)\;,\; A_3\left(1, 0, 0\right)\;,\; \text{and}\; A_3\left(0, \frac{1}{2}, \frac{1}{2}\right)\;, \tag{5.3.10}$$

and their conjugate three-point amplitudes. We refer to this set of three-point amplitudes, loosely, as "the $\mathcal{N} = 4$ SYM interactions."

Finally, constraint (5.3.1) rules out leading-order gravitational coupling to particles of spin-$H > 2$. Such three-point amplitudes, of the form $A_3(H, -H, \pm 2)$, have $A = 2$ and $H > 2$, and yield four-point amplitudes with $2H - 1 > 3$ poles; this cannot

be both unitary and local for $H > 2$. Admissible $A = 2$ amplitudes are restricted to:

$$A_3\left(2, 2, -2\right) , \; A_3\left(2, \frac{3}{2}, -\frac{3}{2}\right) , \; A_3\left(2, 1, -1\right) , \; A_3\left(2, \frac{1}{2}, -\frac{1}{2}\right) , \; A_3\left(2, \frac{1}{2}, -\frac{1}{2}\right) ,$$

(5.3.11)

$$A_3\left(\frac{3}{2}, \frac{3}{2}, -1\right) , \; A_3\left(\frac{3}{2}, 1, -\frac{1}{2}\right) , \; A_3\left(\frac{3}{2}, \frac{1}{2}, 0\right) , \; A_3\left(1, \frac{1}{2}, \frac{1}{2}\right) , \; \text{and } A_3\left(1, 1, 0\right) ,$$

(5.3.12)

and their conjugate three-point amplitudes. We refer to the amplitudes in (5.3.11) as "gravitational interactions." More generally, we refer to this full set of three-point amplitudes, loosely, as "the $\mathcal{N} = 8$ SUGRA interactions."

5.3.3   Killing $N_p = 3$ and $N_p = 2$ theories for $A \geq 3$

It is relatively simple to show that any theory constructed from A_3s with $N_p = 3$ poles, beyond $A = 2$, cannot be consistent with unitarity and locality. To begin, we note that

$$\left\{N_p = 3 \iff 2H + 1 - A = 3\right\} \Rightarrow H = A/2 + 1 .$$

(5.3.13)

We label the helicities in the three-point amplitudes with $N_p = 3$, as $A_3(H, g, f)$. Without loss of generality, we order them as $f \leq g \leq H = A/2 + 1$. As $A > 2$, then $g + f$ must be positive: at a minimum $g > 0$.

Now construct the four-point amplitude $A_4(H, -H, f, -f)$ from this three-point amplitude and its parity-conjugate. By assumption, this amplitude must have three poles, each of which must have an interpretation as a legitimate factorization channel within the theory constructed from $A_3(A/2 + 1, g, f)$ (or some mild extension of the theory/spectrum).

104

However, in order for the t-channel pole in the amplitude $A_4(H, -H, g, -g)$ to have a viable interpretation as a factorization channel, it requires a state with spin *greater* than $A/2 + 1 = H$. Specifically, on this t-pole,

$$A_4(H, -H, g, -g)\Big|_{t \to 0} = \frac{1}{K_{14}^2} A_3\left(\frac{A+2}{2}, -g, \frac{A+2}{2} + g\right) \bar{A}_3\left(-\frac{A+2}{2}, g, -\frac{A+2}{2} - g\right)$$

(5.3.14)

By assumption, $g > 0$: the intermediate state must have helicity $\tilde{H} = A/2 + 1 + g$. Clearly this new state has helicity larger than $H = A/2 + 1$. A priori, there is no problem: new particles mandated by consistency conditions may be included into the spectrum of a theory without necessarily introducing inconsistencies. However, if *these* particles of spin $\tilde{H} > H = A/2 + 1$ are put as external states of the new three-point amplitudes in the modified theory, then these new four-point amplitudes will necessarily have $3 + 2g > 3$ poles, in violation of constraint (5.3.1).

Hence all theories constructed from A_3s with $N_p \geq 3$ and $A > 2$ are inconsistent. Similar arguments show that theories with $N_p = 2$ cannot be consistent for $A > 2$; they are however slightly more detailed, and involve several specific cases at low-A values. Proof of this extended claim is relegated to Appendix A.5.

5.4 There is no GR (YM) but the true GR (YM)

In this section, we investigate further constraints imposed by coupling $A \geq 3$ theories to GR (YM) interactions. This is done by considering four-point amplitudes which factorize as $A_4 \to A_{GR} \times A_3$ and $A_4 \to A_{YM} \times A_3$, where A_3 is the vertex of some other theory. Note however that the arguments in this section apply only to three-point interactions which contain either a spin-2 or a spin-1 state. Other higher spin theories are not constrained in any way by this reasoning.

First, we find that all higher-spin theories are inconsistent if coupled to gravity. This is in addition to the previous section, where spin $s > 2$ theories with $A > 2$ were allowed if $N_p \leq 1$. Further, we show that massless spin-2 states participating in $A > 2$ three-point amplitudes must be identified with *the* graviton which appears in the usual $A = 2$ $A_3(+2, -2, \pm 2)$ three-point amplitudes defining the S-matrix of General Relativity. Pure pole-counting shows that *no* massless spin-2 state in *any* three-point amplitude with $A > 2$ can couple to GR, unless they are within the unique $(H, A) = (2, 6)$ three-point amplitudes, $A_3(2, 2, 2)$ and its complex conjugate. Similar results hold for gluons.[6]

To rule out higher-spin theories interacting with gravity, we show that amplitudes with factorization channels of the type,

$$A_4(1^{+2}, 2^{-2}, 3^{-H}, 4^{-h}) \rightarrow \frac{1}{K^2} A_3(2, -2, +2) \times A_3(-2, -H, -h) \tag{5.4.1}$$

cannot be consistent with unitarity and locality, unless $|H| \leq 2$ and $|h| \leq 2$.

It is relatively easy to see this, especially in light of the constraints from sections 5.3.2 and 5.3.3, which fix $H \leq A/2$ for $A \geq 3$. Note that, in order to even couple to GR's defining three-graviton amplitude, the three-point amplitude in question must have a spin-2 state. These two conditions admit only three possible three-point amplitudes, for a given A:

$$A_3(A/2 - 1, A/2 - 1, 2) \Rightarrow A_4(1^{+2}, 2^{-2}, 3^{-(A/2-1)}, 4^{-(A/2-1)}),$$
$$A_3(A/2 - 1/2, A/2 - 3/2, 2) \Rightarrow A_4(1^{+2}, 2^{-2}, 3^{-(A/2-1/2)}, 4^{-(A/2-3/2)}), \text{ and} \tag{5.4.2}$$
$$A_3(A/2, A/2 - 2, 2) \Rightarrow A_4(1^{+2}, 2^{-2}, 3^{-(A/2)}, 4^{-(A/2-2)}).$$

---

[6]We further show, in appendix A.6 that theories with spin-3/2 states are also unique in a similar manner.

The minimal numerator which encodes the spins of the extenal states in, for instance, the first amplitude, must be,

$$N \sim [1|P|2\rangle^4 \left(\langle 34 \rangle^2 \right)^{(A/2-1)} \Rightarrow [N] = (K^2)^{3+A/2} . \tag{5.4.3}$$

However, by power-counting, the kinematic-dependent part of the amplitude must have mass-dimension,

$$\left[\frac{N}{f(s,t,u)} \right] = \left[\frac{1}{K^2} A_{\text{Left}}^{(GR)} A_{\text{Right}}^{(A)} \right] = \frac{(K^2)^{2/2} (K^2)^{A/2}}{(K^2)} = (K^2)^{A/2} , \tag{5.4.4}$$

and thus the denominator, $f(s,t,u)$, must have mass-dimension,

$$[f(s,t,u)] = (K^2)^3 \Rightarrow f(s,t,u) = s\,t\,u\,! \tag{5.4.5}$$

Casual inspection shows us that the "third" factorization channel, to be sensible, requires an intermediary with spin $A/2 - 1$ to couple directly via the leading $A = 2$ gravitational interactions. This, and similar analysis for the other two classes of three-point amplitudes in Eq. (5.4.2), proves that the spin-2 particle associated with the graviton in the leading-order, $(H, A) = (2, 2)$, gravitational interactions can only participate in three higher-derivative three-point amplitudes, namely,

$$A_3 \left(+2, +1, +1\right) , \; A_3 \left(+2, +\frac{3}{2}, +\frac{3}{2}\right) , \; A_3 \left(+2, +2, +2\right) \tag{5.4.6}$$

In the special case of the three-point amplitude $A_3(+2, +2, +2)$, the third channel simply necessitates an intermediate spin-2 state, the "graviton." Thus GR can couple to itself, or amplitudes derived from $R^a{}_b R^b{}_c R^c{}_a$, its closely related higher-derivative cousin [56][27].[7]

---

[7]The minimal numerators for the other candidate amplitudes in this theory, Eq. (5.4.2), have the same number of spinor-brackets in their numerator as that in Eq. (5.4.3); thus have the same

Second, gluons. Specifically, we show that gluons, i.e. the massless spin-1 particles which couple to each-other at *leading* order via the $H = A = 1$ three-point amplitudes, can *not* consistently couple to any spin $s > 1$ within $A \geq 3$ amplitudes. This means that any constructible amplitude with factorization channels of the type,

$$A_4(1^{+1}, 2^{-1}, 3^{-H}, 4^{-h}) \to \frac{1}{K^2} A_3(1, -1, +1) \times A_3(-1, -H, -h) \qquad (5.4.7)$$

cannot be consistent with unitarity and locality, unless $|H| \leq 1$ and $|h| \leq 1$.

Again, in light of the constraints from sections 5.3.2 and 5.3.3, which fix $H \leq A/2$ for $A \geq 3$, it is relatively easy to see this. To even possibly couple to this three-gluon amplitude, the three-point amplitude in question must have a spin-1 state. These two conditions allow only two possible three-point amplitudes, for a given A:

$$A_3(A/2 - 1/2, A/2 - 1/2, 1) \Rightarrow A_4(1^{+1}, 2^{-1}, 3^{-(A/2-1/2)}, 4^{-(A/2-1/2)}), \text{ and}$$

$$A_3(A/2, A/2 - 1, 1) \Rightarrow A_4(1^{+1}, 2^{-1}, 3^{-(A/2)}, 4^{-(A/2-1)}). \qquad (5.4.8)$$

The minimal numerator which encodes the spins of the extenal states in, for instance, the first amplitude, must be,

$$N \sim [1|P|2\rangle^2 (\langle 34 \rangle^2)^{(A/2-1/2)} \Rightarrow [N] = (K^2)^{A/2+3/2}. \qquad (5.4.9)$$

However, by power-counting, the kinematic-dependent part of the amplitude must have mass-dimension,

$$\left[\frac{N}{f(s, t, u)} \right] = \left[\frac{1}{K^2} A_{\text{Left}}^{(YM)} A_{\text{Right}}^{(A)} \right] = \frac{(K^2)^{1/2} (K^2)^{A/2}}{(K^2)} = (K^2)^{A/2-1/2}, \qquad (5.4.10)$$

mass-dimensions. Therefore all amplitudes must identical number of poles, and as in Eq. (5.4.5), they have $f(s, t, u) = stu$.

and thus the denominator, $f(s, t, u)$ must have mass-dimension two:

$$[f(s, t, u)] = (K^2)^2 \Rightarrow 1/f(s, t, u) \text{ must have at least two poles.} \qquad (5.4.11)$$

Again, casual inspection shows that, while the one pole—that in Eq. (5.4.7)—indeed has a legitimate interpretation as a factorization channel within this theory, the "second" channel generically does not: it requires the gluon to *marginally* couple to spin $A/2 - 1/2 \geq 1$ states. As seen in section 5.3.2, this cannot happen—unless $A/2 - 1/2 = 1 \Leftrightarrow A = 3$.

For the second amplitude in Eq. (5.4.8) the argument is a bit more subtle when $A = 3$. In this case, the u-channel is prohibited, but the t-channel is valid:

$$A_4(1^{+1}, 2^{-1}, 3^{-3/2}, 4^{-1/2}) \to \frac{1}{K^2} A_3(1, -1/2, 1/2) \times A_3(-1/2, -3/2, -1) \qquad (5.4.12)$$

This interaction is ruled out through slightly more detailed arguments, involving the structure of the vector self-coupling constant in $A_3(1, -1, \pm 1)$—discussed in section 5.5. We pause to briefly describe how this is done, but will not revisit this particular, $A_3(1, -1/2, 1/2)$ interaction further (it is just a simple vector-fermion QED or QCD interaction). Simply, we note that $A_3(1, -1, \pm 1) \propto f_{abc}$, the structure-constant for a simple and compact Lie-Algebra; see Eq. (5.5.7). From here, it suffices to note that either by choosing the external vectors to be photons, or gluons of the same color, this amplitude vanishes, and then so does the original s-channel. Nothing is affected in Eq. (5.4.12) and so Eq. (5.4.11) cannot be fulfilled, implying that the coupling constant of $A_3(1, 1/2, 3/2)$ must vanish.

Thus at four-points YM can only couple to itself, gravity via the $A_3(\pm 2, 1, -1)$ three-point amplitude, or amplitudes derived from $F^a_{\;b} F^b_{\;c} F^c_{\;a}$, its closely related higher-derivative cousin.

5.5 Behavior near poles, and a possible shift

In this section we explain a new shift which is guaranteed to vanish at infinity. Using this shift, we re-derive classic results, such as (a) decoupling of multiple species of massless spin-2 particles [61], (b) spin-2 particles coupling to all particles (with $|H| \leq 2$, of course!) with identical strength, $\kappa = 1/M_{pl}$, (c) Lie Algebraic structure-constants for massless spin-1 self-interactions, and (d) arbitrary representations of Lie Algebra for interactions between massless vectors and massless particles of helicity $|H| \leq 1/2$.

Note that in section 5.3, we proved that a four point-amplitude, constructed from a given three-point amplitude and its parity conjugate, $A_3^{(H,A)}$ and $\bar{A}_3^{(H,A)}$, takes the generic form,

$$A_4 \sim \frac{(\langle \ \rangle [\ ])^{2H}}{(K^2)^{2H-A+1}} . \tag{5.5.1}$$

Consequently in the vicinity of, say, the s-pole, the four-point amplitude behaves as,

$$\lim_{s \to 0} A_4 = \frac{1}{s} \frac{N}{t^{2H-A}} , \text{ where } N \sim (\langle \ \rangle [\ ])^{2H} . \tag{5.5.2}$$

We exploit this scaling to identify a useful shift that allows us to analyze constraints on the coupling-constants, the "g_{abc}"-factor in three-point amplitudes [see Eq. (5.2.3)]. Complex deformation of the Mandelstam invariants, which we justify in appendix A.8, for arbitrary $\tilde{s}$ and $\tilde{t}$,

$$(s, t, u) \to (s + z\tilde{s}, t + z\tilde{t}, u + z\tilde{u}) , \tag{5.5.3}$$

grants access to the poles of $A_4(s,t,u)$ without deforming the numerator. Partitioning,

$$A_4(z=0) \sim \kappa_A^2 \frac{N}{f(s,t,u)} \rightarrow A_4(z) \sim \kappa_A^2 \frac{N}{f(s(z),t(z),u(z))}, \qquad (5.5.4)$$

accesses the poles in each term, while leaving the helicity-dependent numerator *unshifted*. Basic power-counting implies that four-point amplitudes, constructed from three-point amplitudes of the type $A_3^{(H,A)} \times \bar{A}_3^{(H,A)}$, die off as $z \rightarrow \infty$ for $2H - A = 1,2$ under this shift. Thus, four-point amplitudes are uniquely fixed by their finite-z residues under this deformation:

$$A_4(z=0) = \sum_{z_P} \mathrm{Res}\left(\frac{A_4(z)}{z}\right). \qquad (5.5.5)$$

Straightforward calculation of the residues on the s-, t-, and u-poles yields,

$$A(1_a, 2_b, 3_c, 4_d) = \left\{ \frac{(\tilde{s})^{2H-A}}{s} g^{abi} g^{icd} + \frac{(\tilde{t})^{2H-A}}{t} g^{adi} g^{ibc} + \frac{(\tilde{u})^{2H-A}}{u} g^{aci} g^{ibd} \right\} \frac{\mathrm{Num}}{(\tilde{s}t - \tilde{t}s)^{2H-A}}.$$
$$(5.5.6)$$

Notably, this closed-form expression for the amplitude contains a non-local, spurious, pole which depends explicitly on the shift parameters, $\tilde{s}$ and $\tilde{t}$ (note: $\tilde{u} = -\tilde{s} - \tilde{t}$). Requiring these spurious parameters to cancel out of the final expression in theories, of self-interacting spin-1 particles, forces the Lie Algebraic structure of Yang-Mills [20][21]. Similarly, for theories of interacting spin-2 particles, we recover the decoupling of multiple species of massless spin-2 particles[61], and the equal coupling of all spin $|H| < 2$ particles to a spin-2 state [56][20][21].

5.5.1   Constraints on vector coupling $(A = 1)$

Here we derive consistency conditions on Eq. (5.5.6) for scattering amplitudes with external vectors, interacting with matter via leading-order, $A = 1$, couplings; $2H - A = 1$. Now, if the amplitude is invariant under changes of $\tilde{s} \to \tilde{S}$, then it necessarily follows that the same holds for re-definitions $\tilde{t} \to \tilde{T}$, and thus that unphysical pole cancels out of the amplitude.

Therefore, if $\frac{\partial A}{\partial \tilde{s}} = 0$, then it indeed follows that the amplitude is invariant under redefinitions of the shift parameter, $\tilde{s}$, and the unphysical pole has trivial residue. Beginning with the all-gluon amplitude, where the three-point amplitudes are $A_3(1_a^{+1}, 2_b^{-1}, 3_c^{\pm 1}) \propto f_{abc}$, we see that the derivative is,

$$\frac{\partial A_4}{\partial \tilde{s}}\bigg|_{(H,A)=(1,1)} \propto f^{abi} f^{icd} + f^{aci} f^{ibd} + f^{adi} f^{ibc}. \tag{5.5.7}$$

Requiring this to vanish is equivalent to imposing the Jacobi identity on these f_{abc}s. Thus, requiring the amplitude to be physical forces the gluon self-interaction to be given by the adjoint representation of a Lie group [20][21][75].

Next, considering four-point amplitudes with two external gluons and two external fermions or scalars, we are forced to introduce a new type of coupling: $A_3(1_a^{\pm 1}, 2_b^{+h}, 3_c^{-h}) \propto (T_a)_{bc}$. Concretely, we wish to understand the invariance of $A_4(1_a^{+1}, 2_b^{-1}, 3_c^{-h}, 4_d^{+h})$, constructed from the shift (5.5.3), under redefinitions $\tilde{s} \to \tilde{S}$.

Factorization channels on the t- and u-poles are given by the products of two A_3s with one gluon and two spin-h particles, and thus are proportional to $(T_a)_{ci}(T_b)_{di}$ and $(T_a)_{di}(T_b)_{ci}$, respectively–while the s-channel is proportional to $f_{abi}(T_i)_{cd}$. So, $\frac{\partial A}{\partial \tilde{s}}$ is proportional to,

$$(T_a)_{ci}(T_b)_{id} - (T_a)_{di}(T_b)_{ic} + f_{abi}(T_i)_{cd}. \tag{5.5.8}$$

This is nothing other than the definition of the commutator of two matrices, T_a and T_b in an arbitrary representation of the Lie group "defined" by the gluons in Eq. (5.5.7) [20][21][75].

5.5.2   Graviton coupling

Four-point amplitudes with two external gravitons have $2H - A = 2$, and so Mandelstam deformation (5.5.3) yields,

$$A(1_a^{-2}, 2_b^{-h}, 3_c^{+2}, 4_d^{+h}) = \left\{ \frac{\tilde{s}^2}{s} \kappa_h^{abi} \kappa_h^{icd} + \frac{\tilde{t}^2}{t} \kappa_h^{adi} \kappa_h^{ibc} + \frac{\tilde{u}^2}{u} \kappa_{h=2}^{aci} \kappa_h^{ibd} \right\} \frac{(\langle 12 \rangle [34])^{4-2h} \langle 1 | 2 - 3 | 4]^{2h}}{(\tilde{s} t - \tilde{t} s)^2} ,$$

$$(5.5.9)$$

where κ_h^{abc} is the coupling constant in $A_3(1_a^{\pm 2}, 2_b^{-h}, 3_c^{+h})$. Demanding this amplitude be independent of redefinitions of $\tilde{s} \to \tilde{S}$, again reduces to the constraint that the partial derivative of Eq. (5.5.9) must vanish. Evaluating the derivative, we see,

$$\left. \frac{\partial A_4}{\partial \tilde{s}} \right|_{(H,A)=(2,2)} \propto \tilde{s} \left(\kappa_h^{abi} \kappa_h^{icd} - \kappa_{h=2}^{adi} \kappa_h^{ibc} \right) + \tilde{t} \left(\kappa_h^{aci} \kappa_h^{ibd} - \kappa_{h=2}^{adi} \kappa_h^{ibc} \right) , \qquad (5.5.10)$$

which vanishes only if,

$$\kappa_h^{abi} \kappa_h^{icd} = \kappa_{h=2}^{adi} \kappa_h^{ibc} , \text{ and } \kappa_h^{aci} \kappa_h^{ibd} = \kappa_{h=2}^{adi} \kappa_h^{ibc} . \qquad (5.5.11)$$

As noted in [20], for $h = 2$, this implies that the κ_h^{abc}s are a representation of a commutative, associative algebra. Such algebras can be reduced to self-interacting theories which decouple from each other. In other words, multiple gravitons, i.e. species of massless spin-2 particles interacting via the leading-order $(H, A) = (2, 2)$ three-point amplitudes, necessarily decouple from each-other. This is the perturbative casting of the Weinberg-Witten theorem [61]. As the multiple graviton species decouple, we refer to the *diagonal* graviton self-interaction coupling as, simply, κ.

113

Diagonal gravitational self-coupling powerfully restricts the class of solutions to Eq. (5.5.11) for $h < 2$. Directly, it implies that any individual graviton can only couple to a particle-antiparticle pair. In other words, $\kappa_h^{gab} = 0$, for different particle flavors a and b on the spin-$\pm h$ lines. Similar to the purely gravitational case, we write simply $\kappa_h^{gaa} = \kappa_h$. Furthermore, to solve Eq. (5.5.11) for $h \neq 2$ then it also must hold that $\kappa_h = \kappa_{h=2} = \kappa$. In other words, the graviton self-coupling constant κ is a simple constant; all particles which interact with a given unique graviton do so diagonally and with identical strengths. Thus multiple graviton species decouple into disparate sectors, and, within a given sector, gravitons couple to all massless states with identical strength, κ—the perturbative version of the equivalence principle [56][20][21].

5.5.3 Killing the relevant $A_3\left(0, \frac{1}{2}, -\frac{1}{2}\right)$-theory

This shift neatly kills the S-matrix constructed from the three-point amplitudes $A_3(\frac{1}{2}, -\frac{1}{2}, 0)$. Just as before, we will see that in order for $A_4(0, 0, \frac{1}{2}, -\frac{1}{2})$ to be constructible (via complex Mandelstam - deformations) and consistent, the coupling constant in the theory must vanish.

As in YM/QCD, in this theory $2H - A = 1$. Invariance of $A_4(0, 0, \frac{1}{2}, -\frac{1}{2})$ under deformation redefinitions $\tilde{s} \to \tilde{S}$ again boils down to a constraint akin to Eq. (5.5.7)—with one exception. Namely, there are only two possible factorization channels in this theory and not three: any putative s-channel pole would require a ϕ^3 interaction, not present in this minimal theory. And so invariance under redefinitions $\tilde{s} \to \tilde{S}$ reduces to,

$$\frac{\partial A_4}{\partial \tilde{s}}\bigg|_{(H,A)=(\frac{1}{2},0)} = f^{acp} f^{bdp} + f^{adp} f^{bcp}. \tag{5.5.12}$$

The only solution to this constraint is for $f^{acp}f^{bdp} = 0 = f^{adp}f^{bcp}$, i.e. for the coupling constant to be trivially zero.[8]

5.6 Interacting spin-$\frac{3}{2}$ states, GR, and supersymmetry

Supersymmetry automatically arises as a consistency condition on four-point amplitudes built from leading-order three-point amplitudes involving spin-3/2 states. In a sense, this should be more-or-less obvious from inspection of the leading-order spin-3/2 amplitudes in Eqs. (5.3.11), and (5.3.12). For convenience, they are,

$$A_3\left(\frac{3}{2}, \frac{1}{2}, 0\right), \quad A_3\left(\frac{3}{2}, 1, -\frac{1}{2}\right), \quad A_3\left(\frac{3}{2}, \frac{3}{2}, -1\right), \quad \text{and } A_3\left(\frac{3}{2}, 2, -\frac{3}{2}\right). \quad (5.6.1)$$

Clearly, every non-gravitational $A = 2$ amplitude with a spin-3/2 state involves one boson and one fermion, with helicity (magnitudes) that differ by exactly a half-unit. This should be unsurprising, as $A - 3/2 = 1/2$. Nonetheless, we should expect supersymmetry to be an emergent phenomena: throughout the previous examples, mandating a unitary interpretation of a factorization channel within novel four-point amplitudes in a theory forced introduction of new states with new helicities into the spectrum/theory. In a sense, the novelty of $A = 2$ amplitudes with external spin-3/2 states is that these new helicities do *not* lead to violations of locality and unitarity.

In amplitudes with external spin-3/2 states (and no external gravitons), each term in the amplitude must have $2H + 1 - A \to 3 + 1 - 2 = 2$ poles. Generically, one of these two poles will mandate inclusion of states with *new* helicities into the spectrum

---

[8]One may wonder why such an argument does not also rule out conventional well-known theories, such as spinor-QED or GR coupled to spin-1/2 fermions, as inconsistent. The resolution to this question is subtle, but boils down to the fact that amplitudes involving fermions in these $A > 0$ theories have extra, *antisymmetric* spinor-brackets in their numerators. These extra spinor-brackets introduce a relative-sign between the two terms, and in effect modify the condition (5.5.12) from $\{ff + ff = 0 \Rightarrow f = 0\}$ to $ff - ff = 0$, which is trivially satisfied.

of the theory. Fundamentally, we see that the *minimal* $A = 2$ theory with a single species of spin-3/2 state is given by the two three-point amplitudes (and and their parity-conjugates):

$$A_3\left(\frac{3}{2}, 2, -\frac{3}{2}\right) , \text{ and } A_3(2, 2, -2). \tag{5.6.2}$$

These interactions *define* pure $\mathcal{N} = 1$ SUGRA, and are indicative of all other theories which contain massless spin-3/2 states (at leading order). All non-minimal extensions of any theory containing spin-3/2 states necessarily contain the graviton. As we will make precise below, supersymmetry necessitates gravitational interactions— supersymmetry requires the graviton.

Minimally, consider a four-particle amplitude which ties together four spin-3/2 states, two with helicity $h = +3/2$, and two with helicity $h = -3/2$, via leading-oder $A = 2$ interactions: $A_4^{(A=2)}(1^{+\frac{3}{2}}, 2^{+\frac{3}{2}}, 3^{-\frac{3}{2}}, 4^{-\frac{3}{2}})$. As this is a minimal amplitude, we consider the case where the like-helicity spin-3/2 states are identical: there is only one flavor/species of a spin-3/2 state. How many poles would such an amplitude have? By Eq. (5.3.1), there must be

$$2H + 1 - A = N_p \longrightarrow N_p = 2 \tag{5.6.3}$$

poles in any four-point amplitude constructed from $A = 2$ three-point amplitudes which has spin-3/2 states as its highest-spin external state. The key point here is really only that $N_p > 0$: the amplitude must have a factorization channel. Because it has two poles, at least one of them must be mediated by graviton exchange. In this minimal theory, as (a) gravitons can only be produced through particle-antiparticle annihilation channels and (b) the like-helicity spin-3/2 states are identical, both channels occur via graviton exchange. See Fig. 5.2(a) for specifics.

$$A_4(1^{+3/2}, 2^{+3/2}, 3^{-3/2}, 4^{-3/2}): \qquad A_4(1^{+3/2}, 2^{+a}, 3^{-3/2}, 4^{-a}):$$

Figure 5.2: Factorization necessitates gravitation in theories with massless spin-3/2 states. Specifically, figure (a) represents the two factorization channels in the minimal four-point amplitude, $A_4(1^{+\frac{3}{2}}, 2^{+\frac{3}{2}}, 3^{-\frac{3}{2}}, 4^{-\frac{3}{2}})$ in an S-matrix involving massless spin-3/2 states. Further, figure (b) shows the two factorization channels present in the amplitude $A_4(3/2, -3/2, +a, -a)$.

Because this set of external states should always be present in *any* theory with leading-order interactions between any number of spin-3/2 states, S-matrices of these theories must *always* include the graviton.

This can be made even more explicit. Consider an S-matrix constructed, at least in part, from a three-point amplitude, $A_3(3/2, a, b)$, and its conjugate, $A_3(-3/2, -a, -b)$, where $H = 3/2$ and $A = 2$. These three-point amplitudes tie-together a spin-3/2 state with two other states which, collectively, have helicity-magnitudes $|H| \leq 3/2$. This theory necessarily contains the four-point amplitude,

$$A_4(1^{+\frac{3}{2}}, 2^{+a}, 3^{-\frac{3}{2}}, 4^{-a}). \tag{5.6.4}$$

As noted in Eq. (5.6.3), the denominator within this amplitude has two kinematic poles. Clearly the s-channel is has the spin-b state for an intermediary. However, as (a) the three-point amplitudes in the theory all have $A = 2$, and (b) the opposite-helicity spin-3/2 states (equivalently, the spin-a states) are antiparticles, the u-channel

factorization channel must be mediated by a massless spin-2 state: the graviton. This is depicted in Fig. 5.2.

Note that the t-channel is also possible, mediated by a helicity $a + 1/2$ particle. However, repeating the above reasoning for the new $A_3(3/2, a + 1/2, -a)$ amplitude will eventually lead to the necessity of introducing a graviton. This is because in each step the helicity of a is increased by $1/2$, and this process stops once a reaches $3/2$, when both the t and u-channels can only be mediated by a graviton. This pattern of adding particles with incrementally different spin will be investigated further in the following sections.

Before delving into details of the spectra in theories with multiple species of spin-3/2 states, we note one final feature of these theories. Analysis of their four-particle amplitudes, e.g. the amplitude in Eq. (5.6.4), via on-shell methods such as the Mandelstam deformation introduced in the previous section, straightforwardly shows that the coupling constants in this theory [the $g_{abc}^{\pm}$s in the language of Eq. (5.2.3)] are *equal* to $\kappa = 1/M_{pl}$, the graviton self-coupling constant. More generally, in any $A = 2$ theory with spin-3/2 states, *each and every* defining three-point amplitude, $A_3(1_a^{h_a}, 2_b^{h_b}, 3_c^{h_c}) = \kappa_{abc} M_{abc}(\langle, \rangle)$ has an identical coupling constant, $\kappa_{abc} = \kappa = 1/M_{pl}$, up to (SUSY preserving Kronecker) delta-functions in flavor-space.

It is important to emphasize here that, as the spin-3/2 gravitinos only interact via $A = 2$ three-point amplitudes, they cannot change the $A < 2$ properties of any state within the same amplitude. Concretely, a bosonic (fermionic) state which transforms under a given specific representation of a compact Lie Algebra, i.e. a particle which interacts with massless vectors (gluons) via leading order ($A = 1$) interactions, can only interact with a fermionic (bosonic) state which transforms under the *same* representation of the Lie Algebra when coupled to spin-3/2 states within $A = 2$ three-point amplitudes. From the point of view of the marginal interactions, only the spin of the states which interact with massless spin-3/2 "gravitino(s)" may

change. This is the on-shell version of the statement that all states within a given supermultiplet have the same quantum-numbers, but different spins.

We now consider the detailed structure of interactions between states of various different helicities which participate in S-matrices that couple to massless spin-3/2 states. Minimally, such theories include a single graviton and a single spin-3/2 state (and its antiparticle). Equipped with this, we can ask what the next-to-minimal theory might be. There are two ways one may enlarge the theory: (1) introducing a state with a *new* spin into the spectrum of the theory, or (2) introducing another species of massless spin-3/2 state. We pursue each in turn.

5.6.1 Minimal extensions of the $\mathcal{N}=1$ supergravity theory

First, we ask what the minimal enlargement of the $\mathcal{N}=1$ SUGRA theory is, if we require inclusion of a single spin-1 vector into the spectrum. In other words, what three-particle amplitudes must be added to,

$$\mathcal{N}=1\,\mathrm{SUGRA} \iff \left\{A_3(+2,\pm2,-2)\,,\,A_3(+3/2,\pm2,-3/2)\right\}, \qquad (5.6.5)$$

in order for all four-particle amplitudes to factorize properly on all possible poles, once vectors are introduced into the spectrum. Clearly, inclusion of a vector requires inclusion of,

$$A_3(+1,\pm2,-1)\,, \qquad (5.6.6)$$

into the theory. It is useful to consider the four-particle amplitude, $A_4(-3/2,+3/2,+1,-1)$; its external states are only those known from the minimal theory and this extension, i.e. a particle-antiparticle pair of the original spin-3/2 "gravitino" and a particle-antiparticle pair of the new massless spin-1 vector.

By Eq. (5.6.3), this amplitude must have two poles. The s-channel pole is clearly mediated by graviton-exchange, as the amplitude's external states are composed of two distinct pairs of antiparticles. The *second* pole brings about new states. There are two options for which channel the second pole is associated with: the u-channel pole, or the t-channel pole. On the t-channel, the amplitude must factorize as,

$$A_4(-3/2, +3/2, +1, -1)\Big|_{t \to 0} \to A_3(1^{-\frac{3}{2}}, 4^{-1}, P_{14}^{+\frac{1}{2}}) \frac{1}{(p_1 + p_4)^2} A_3(2^{+\frac{3}{2}}, 3^{+1}, -P_{14}^{-\frac{1}{2}}),$$

$$(5.6.7)$$

and we see that, by virtue of the fact that the three-point amplitudes must have $A = 2$, the new particle introduced into the spectrum is a spin-1/2 fermion. This option corresponds to the CPT-complete spectrum for $\mathcal{N} = 1$ SUGRA combined with the CPT complete spectrum for $\mathcal{N} = 1$ SYM. In this case, the full $A = 2$ sector of the theory would be,

$$\left\{ A_3(2, 2, -2), \ A_3\left(2, \frac{3}{2}, -\frac{3}{2}\right), \ A_3(2, 1, -1), \ A_3\left(2, \frac{1}{2}, -\frac{1}{2}\right), \ A_3\left(\frac{3}{2}, 1, -\frac{1}{2}\right) \right\},$$

$$(5.6.8)$$

Note that, as the spin-2 and spin-3/2 states interact gravitationally, one can add extra, leading-order $A = 1$ ("gauge") interactions between the spin-$H \leq 1$—but it is not necessary.

If the second pole is in the u-channel, then the amplitude must factorize as,

$$A_4(-3/2, +3/2, +1, -1)\Big|_{u \to 0} \to A_3(1^{-\frac{3}{2}}, 3^{+1}, P_{13}^{-\frac{3}{2}}) \frac{1}{(p_1 + p_3)^2} A_3(2^{+\frac{3}{2}}, 4^{-1}, -P_{13}^{+\frac{3}{2}}),$$

$$(5.6.9)$$

120

and we see that, by virtue of the fact that the three-point amplitudes must have $A = 2$, the new particle introduced into the spectrum must be *another* spin-3/2 gravitino. This option corresponds to the CPT-complete spectrum for $\mathcal{N} = 2$ SUGRA.

It is not immediately obvious that this internal spin-3/2 state must be distinguishable from the original spin-3/2 state. Distinguishability comes from the fact that the factorization channel $A_3(1_a^{3/2}, 2_b^{3/2}, P^{-1})A_3(3_{\bar{a}}^{-3/2}, 4_{\bar{b}}^{-3/2}, -P^{+1})/K_{12}^2$ must be part of a four-particle amplitude with both the "new" and the "old" spin-3/2 species as external states. This amplitude also has only two poles. As one of them is mediated by vector exchange, we see that there is only *one* graviton-exchange channel. Therefore the new and old spin-3/2 states cannot be identical.

Constructing theories in this way is instructive. As a consequence of requiring a unitary interpretation of all factorization channels in non-minimal S-matrices involving massless spin-3/2 states, we are forced to introduce a new fermion for every new boson and vice-versa. Further, we see that through allowing minimal extensions to this theory, we can either have extended supergravity theories, i.e. $\mathcal{N} = 2$ SUGRA theories, truncations of the full $\mathcal{N} = 8$ SUGRA multiplet, or supergravity theories and supersymmetric Yang-Mills theories in conjunction, i.e. $\mathcal{N} = 1$ SUGRA $\times$ $\mathcal{N} = 1$ SYM theories. The same lessons apply for more extended particle content. However, it is difficult to make such constructions systematic. Below, we discuss the second, more systematic, procedure which hinges upon the existence of $\mathcal{N}$ distinguishable species of spin-3/2 fermions.

5.6.2 Multiple spin-$\frac{3}{2}$ states and (super)multiplets

Another way to understand these constructions is as follows: specify the number $\mathcal{N}$ of distinguishable species of spin-3/2 states, and then specify what else (besides the graviton) must be included into the theory. This amounts to specifying the number of supersymmetries and the number and type of representations of the supersymmetry

algebra. In the above discussion, the two minimal extensions to the $\mathcal{N} = 1$ SUGRA theory were: (a) $\mathcal{N} = 1$ SUGRA $\times$ $\mathcal{N} = 1$ SYM, and (b) $\mathcal{N} = 2$ SUGRA, with *two* gravitinos and one vector.

To render this construction plan unique, we require that all spins added to the theory besides the graviton and the $\mathcal{N}$ gravitinos, i.e. all *extra* supermultiplets included in the theory, be the "top" helicity component of whatever comes later. So, again, the discussion in subsection 5.6.1 would cleanly fall into two pieces: (A) a *single* gravitino (in the graviton supermultiplet) together with a gluon and its descendants, and (B) *two* distinct gravitinos (in the graviton supermultiplet) and their descendants. Clearly this procedure can be easily extended (see subsection 5.6.3).

The general strategy is to look at amplitudes which tie together gravitinos and lower-spin descendants (ascendants) of the "top" ("bottom") helicities in the theory, of the type

$$A(1_x^{+3/2}, 2_y^{-3/2}, 3_u^{-s}, 4_v^{+s}) \, . \tag{5.6.10}$$

Here, the x and y labels describe the species information of the two gravitinos, and u and v describe the species information of the lower-spin particles in the amplitude. Note graviton-exchange can only happen in the s-channel. Unitary interpretation of one of the other channels generically forces the existence of a new spin-$s - \frac{1}{2}$ state into the spectrum. For pure SUGRA (i.e. no spin-1, 1/2, or 0 "matter" supermultipets), this works as follows,

1. One gravitino ($\{a\}$). The unique amplitude to consider, after the archetype in Eq. (5.6.10), is $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_a^{3/2}, 4_a^{-3/2}\big)$. Both s- and t-channels occur via graviton-exchange. The theory is self-complete: the other amplitude does not require any new state.

2. Two gravitinos ($\{a, b\}$). The unique amplitude to consider is $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_b^{3/2}, 4_b^{-3/2}\big)$. Here, the t-channel is disallowed; the u-channel needs a vector with gravitino-label $\{ab\}$. Inclusion of this state completes the theory.

3. Three gravitinos ($\{a, b, c\}$). Two classes amplitudes of the type in Eq. (5.6.10) to consider. First, $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_c^{3/2}, 4_c^{-3/2}\big)$ requires a vector with gravitino-label ac in its u-channel; as there are three amplitudes of this type, there are three distinguishable vectors: $\{ab, ac, bc\}$. Second, $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_{bc}^{+1}, 4_{bc}^{-1}\big)$ needs a fermion with gravitino-label $\{abc\}$ in the u-channel. No other amplitudes require any new states.

4. Four gravitinos ($\{a, b, c, d\}$). Here, the structure is slightly more complicated, but similarly hierarchical. Three classes of amplitudes, each following from its predecessor. First, there are $\binom{4}{2}$ distinct $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_c^{3/2}, 4_c^{-3/2}\big)$s. They require vectors with gravitino labels $\{ab, ac, ad, bc, bd, cd\}$. Second, there are $\binom{4}{3}$ distinct $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_{bc}^{+1}, 4_{bc}^{-1}\big)$s, which require spin-1/2 fermions with labels $\{abc, abd, acd, bcd\}$. Third and finally, we consider $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_{bcd}^{1/2}, 4_{bcd}^{-1/2}\big)$. On its u-channel, it requires a spin-0 state with gravitino-label $\{abcd\}$.

Crucially, we observe that all spins present in the graviton supermultiplet (the graviton and all of its descendants) with $\mathcal{N}$ gravitinos are still present in the graviton supermultiplet with $\mathcal{N} + 1$ gravitinos—but with higher multiplicities. These descendant states are explicitly labeled by the gravitino species from whence they came. Spin-s states in the graviton multiplet have $\binom{\mathcal{N}}{s}$ distinct gravitino/SUSY labels.

Importantly, if h is the unique lowest helicity descendant of the graviton with $\mathcal{N}$ gravitinos, then inclusion of an extra gravitino allows for $\mathcal{N}$ *new* helicity-h descendants of the graviton. Now, studying $A_4\big(1_{\mathcal{N}+1}^{3/2}, 2_{\mathcal{N}+1}^{-3/2}, 3_{ab...\mathcal{N}}^{+h}, 4_{ab...\mathcal{N}}^{-h}\big)$, we see that again the u-channel requires a single new descendant with helicity $h - 1/2$ and SUSY-label $\{ab...\mathcal{N}, \mathcal{N} + 1\}$.

This logic holds for the descendants of all "top" helicity states: isomorphic tests and constructions, for example, allow one to construct and count the descendants from the gluons of SYM theories. We see below that, by obeying the consistency conditions derived from pole-counting and summarized in Fig. 5.1, this places strong constraints on the number of distinct gravitinos in gravitational and mixed gravitational and $A < 2$-theories.

5.6.3 Supersymmetry, locality, and unitarity: tension and constraints

As we have seen, inclusion of $\mathcal{N}$ distinguishable species of massless spin-3/2 states into the spectrum of constructible theories forces particle helicities $\{H, H - 1/2, ..., H - \mathcal{N}/2\}$ into the spectrum. But, as we have seen in sections 5.3 and 5.4, the $A = 2$ gravitational interactions cannot consistently couple to helicities $|h| > 2$. And so, within the supersymmetric gravitational sector, we must have (a) $H = 2$, and (b) $H - \mathcal{N}/2 \geq -2$. Otherwise, we must couple a spin-5/2 > 2 to gravity–which is impossible. Locality and unitarity constrains $\mathcal{N} \leq 8$.

So there is tension between locality, unitarity, and supersymmetry. We now ask about the spectrum of next-to-minimal theories coupled to spin-3/2 states. There are two options for such next-to-minimal theories: either (relevant) self-interacting scalars or (marginal) self-interacting vectors coupled to $\mathcal{N}$ flavors of spin-3/2 particles. Immediately, we see that ϕ^3 cannot be consistently coupled to spin-3/2 states. Coupling the spin-0 lines in ϕ^3 to even one spin-3/2 state would force the existence of non-zero $A_3(1/2, -1/2, 0)$-type interactions. But these interactions, as discussed in section 5.5.3, are not consistent with unitarity and locality. So *relevant* interactions cannot be supersymmetrized in flat, four-dimensional, Minkowski space.

However, for $(A = 1)$ self-interacting gluons, the story is different. By the arguments above, unitarity and locality dictate that if $\mathcal{N}$ spin-3/2 particles are coupled

to gluons, then gluons must couple via marginal interactions, to spin-$\pm|1 - \mathcal{N}/2|$ states. Again basic pole-counting in section 5.3.2, $A = 1$ interactions are only valid for $|h| \leq 1$. And so, we must have $\mathcal{N} \leq 4$, if we would like to couple interacting vectors to multiple distinct spin-3/2 particles while also respecting locality and unitarity of the S-matrix.

5.7 Future directions and concluding remarks

Our results can be roughly separated into two categories. First, we classify and systematically analyze all possible three-point massless S-matrix elements in four-dimensions, via basic pole-counting. The results of this analysis are succinctly presented in Fig. 5.1. Second, we study the couplings and spectra of the few, special, self-interactions allowed by this first, broader-brush, analysis. In this portion of the paper, we reproduce standard results on the structures of massless S-matrices involving higher-spin particles, ranging from the classic Weinberg-Witten theorem and the Equivalence Principle to the existence of supersymmetry, as consequences of consistency conditions on various S-matrix elements. We recap the main results below.

Locality and constructibility fix the generic pole-structure of four-point tree-amplitudes constructed from fundamental higher-spin three-point massless amplitudes. Tension between the number of poles mandated by these two principles, and unitarity, which bounds the number of poles in an amplitude from above ($N_p \leq 3$), eliminates all but a small (yet infinite) sub-class of three-point amplitudes as leading to four-particle tree-level S-matrices that are inconsistent with locality and unitarity.

Already from this point of view we see that, for low $A = |\sum_{i=1}^{3} h_i|$, (Super-)Gravity, (Super-)Yang-Mills, and ϕ^3-theory are the unique, leading, interactions between particles of spin-$|h| \leq 2$. Further, we see that gravitational interactions cannot *directly* couple to particles of spin-$|H| > 2$. Similarly, massless vectors interacting

at leading-order ($A = 1$) cannot consistently couple to massless states with helicity-$|H| > 1$.

In light of these constraints, we study higher-A theories. The upper-bound on the number of poles in four-particle amplitudes, imposed by unitarity and locality, is even stronger for higher-spin *self-interactions* ($N_p \leq 1$ for $A > 2$). The set of consistent three-point amplitudes with $A > 2$ is further reduced to lie between the lines $H = A/2$ and $H = A/3$.

Exploiting this, we re-examine whether-or-not the primitive amplitudes which define the S-matrices of General Relativity and Yang-Mills can *indirectly* couple to higher-spin states in a consistent manner. As they cannot directly couple to higher-spin states, this coupling can only happen within non-primitive four-point (and higher) amplitudes, which factorize into GR/YM self-interaction amplitudes (with $H = A$), multiplied by an $A > 2$ three-point amplitude with an external tensor or vector. Again, simple pole-counting shows that amplitudes which couple the $A \leq 2$ to the $A > 2$ theories generically have poles whose unitary interpretation mandates existence of a particle with spin $\tilde{H} > A/2$.

Having such a high-spin particle contradicts the most basic constraint (5.3.1), and thus invalidates the interactions—save for two special examples. These examples are simply the higher-derivative amplitudes which also couple three like-helicity gravitons, $A_3(2, 2, 2)$, and/or like-helicity gluons, $A_3(1, 1, 1)$. There is a qualitative difference between massless spin-2(1) particles participating in lower-spin ($A \leq 2$) amplitudes, and massless spin-2(1) particles participating in higher-spin ($A > 3$) amplitudes. The graviton is unique. Gluons are also unique. They cannot be coupled to particles of spin-$|h| > 2$!

Equipped with the (now) *finite* list of leading interactions between spin-1, spin-2, and lower-spin states, we then analyze the structure of their interactions—i.e. their coupling constants. To do this, we set up, and show the validity of, the Mandel-

stam shift (5.5.3). Assuming parity-invariance, and thus $g_{ijk}^+ = +g_{ijk}^-$ [in the notation of Eq. (5.2.3)], we perform the Mandelstam shift on four-point amplitudes in these theories. Invariance with respect to redefinitions of the unphysical shift-parameter directly implies the Lie Algebraic structure of the marginal ($A = 1$) coupling to massless vectors; similarly massless tensors must couple (a) diagonally (in flavor space), and (b) with equal strength to all states.

Finally, we analyze consistency conditions on four-point amplitudes which couple to massless spin-3/2 states. The minimal theory/set of interacting states, at leading order, which include a single spin-3/2 state is the theory with a single graviton and a single spin-3/2 state, at $A = 2$. From this observation, we identify the spin-3/2 state with the gravitino. The gravitino also couples to matter with strength $\kappa = 1/M_{pl}$, but as it is not a boson, it does not couple "diagonally": coupling to non-graviton states within the leading-order $A = 2$ interactions automatically necessitates introduction of a fermion for every boson already present in the theory, and vice-versa. We recover the usual supersymmetry constraints, such as fermion-boson level matching, and the maximal amount of distinguishable gravitinos which may couple to gluons and/or gravitons; above these bounds, the theories becomes inconsistent with locality and unitarity.

We close with future directions. Clearly, it would be interesting to discuss on-shell consistency conditions for theories which have primitive amplitudes which begin at *four* points, rather than at three-points. Certain higher-derivative theories, such as the nonlinear sigma model[81], are examples of this type of theory: in the on-shell language, derivative interactions between scalars can only act to give factors of non-trivial kinematical invariants within the numerator of a given amplitude. All kinematical invariants are identically zero at three points. So the first non-zero S-matrix elements in derivatively-coupled scalar theories must be at four-points. Supported by the existence of semi- on-shell recursions in these theories[82], it is conceivable

that these theories are themselves constructible. Straightforwardly, this leads to the on-shell conclusion that all S-matrix elements in these theories have an even number of external legs. Much more could be said, and is left to future work.

Besides theories with derivative interactions, there is also a large of class of higher-spin theories *not* constrained by any of the arguments presented in this paper. These are $A \geq 3$, $N_p \leq 1$ theories which do not contain any spin-1 or spin-2 states, for example $A_3(3/2, 3/2, 0)$. It is not clear from this on-shell perspective whether such theories are completely compatible with locality and unitarity, or more sophisticated tests can still rule them out.

Indeed, an exhaustive proof of the spin-statistics theorem has yet to be produced through exclusively on-shell methods. Proof of this theorem usually occurs, within local formulations of field theory, through requiring no information propagation outside of the light-cone. In the manifestly on-shell formalism, all lines are on their respective light-cones; superluminal propagation, and (micro-)causality violations are naively inaccessible. Ideally some clever residue theorem, such as that in [83], should prove the spin-statistics theorem in one fell swoop. Further, it would be interesting to prove that parity-violation, with $g_{abc}^+ = -g_{abc}^-$, within three-point amplitudes only leads to consistent four-particle amplitudes for parity-violating gluon/photon-fermion amplitudes ($A = 1$).

Finally, one may reasonably ask what the corresponding analysis would yield for *massive* states in constructible theories. As is well known, massive vectors must be coupled to spinless bosons (such as the Higgs), to retain unitarity at $E_{\text{CM}} \sim s \gtrsim m_V$ [75]. It would be extremely interesting to see this consistency condition, and analogous consistency conditions for higher-spin massive particles, naturally fall out from manifestly on-shell analyses.

In conclusion, these results confirm the Coleman-Mandula and the Haag-Lopuszanski-Sohnius theorems for exclusively massless states in four-dimensions [84][85].

Through assuming a constructible, non-trivial, S-matrix that is compatible with locality and unitarity, we see that the maximal structure of non-gravitational interactions between low-spin particles is that of compact Lie groups. Only through coupling to gravitons and gravitinos can additional structure be given to the massless tree-level S-matrix (at four-points). This additional structure is simply supersymmetry; it relates scattering amplitudes with asymptotic states of different spin, within the same theory. Further, no gravitational, marginal, or relevant interaction may consistently couple to massless asymptotic states with spin greater than two.

Acknowledgements

Chapter 6

Bonus BCFW behavior from Bose symmetry

6.1 Introduction

Modern on-shell S-matrix methods may dramatically improve our understanding of perturbative quantum gravity, but current foundations of on-shell techniques for General Relativity still rely on off-shell Feynman diagram analysis. Here, we complete the fully on-shell proof of Ref. [21] that the recursion relations of Britto, Cachazo, Feng, and Witten (BCFW) apply to General Relativity tree amplitudes. We do so by showing that the surprising requirement of "bonus" z^{-2} scaling under a BCFW shift directly follows from Bose-symmetry. Moreover, we show that amplitudes in generic theories subjected to BCFW deformations of identical particles necessarily scale as z^{even}. When applied to the color ordered expansions of Yang-Mills, this directly implies the improved behavior under non-adjacent gluon shifts. Using the same analysis, three-dimensional gravity amplitudes scale as z^{-4}, compared to the z^{-1} behavior for conformal Chern-Simons matter theory.

Mysteries abound at the interface between General Relativity and Quantum Field Theory. Particularly, graviton scattering amplitudes in maximally supersymmetric $\mathcal{N} = 8$ Supergravity have surprisingly soft behavior in the deep ultraviolet (UV). To four loops, it has been shown that the critical dimension of supergravity is the same as $\mathcal{N} = 4$ Super Yang-Mills, a conformally invariant theory free of UV divergences [86]. This result was obtained through the peculiar BCJ duality between color and kinematics, which relates graviton amplitudes to the squares of gluon amplitudes [7][87]. Other arguments, based the non-linearly realized $E_{7(7)}$ symmetry of $\mathcal{N} = 8$ supergravity, predict UV finiteness to six-loops [88]. Yet others hint at a full finiteness (see e.g. [89]).

Standard perturbative techniques, i.e. Feynman diagrams, lead to incredibly complicated expressions, and obfuscate general features of the theory. Reframing the discussion in terms of the modern analytic S-matrix has so far proven incredibly useful for discussing Yang-Mills theory (for example, in Ref. [90]), and may provide crucial insights into quantum gravity as well. The on-shell program offers a different perspective on the principles of locality and unitarity, and their powerful consequences [21][91]. It also provides a computational powerhouse, the BCFW on-shell recursion relation [2].

Briefly, if two external momenta in the amplitude A_n are subjected to the on-shell BCFW shift:

$$p_1^\mu \to p_1^\mu + zq^\mu \qquad p_2^\mu \to p_2^\mu - zq^\mu \tag{6.1.1}$$

and $A_n(z) \to 0$ for large z, then $A_n(z = 0)$ can be recursively constructed from lower-point on-shell amplitudes:

$$A_n = \oint \frac{dz}{z} A_n(z) = \sum_{\{L\}} \frac{A_L(\hat{1}, \{L\}, \hat{P}) A_R(\hat{P}, \{R\}, \hat{n})}{P^2}. \tag{6.1.2}$$

Initial proofs required sophisticated Feynman diagram analyses, and found that gluon amplitudes have the minimum scaling of z^{-1}, but that graviton amplitudes have a "bonus", seemingly unnecessary, scaling of z^{-2} [2]-[78]. Surprisingly, Ref. [21] found that a fully on-shell proof of BCFW constructability actually requires this improved scaling for gravitons, in order for Eq. (6.1.2) to satisfy unitarity. The bonus scaling is not just a "bonus", but a critical property of General Relativity. This z^{-2} scaling, also present in the case of non-adjacent gluon shifts [92], implies new residue theorems:

$$0 = \oint A_n(z)dz = \sum_{\{L\}} z_P \frac{A_L(\hat{1}, \{L\}, \hat{P}) A_R(\hat{P}, \{R\}, \hat{n})}{P^2}, \qquad (6.1.3)$$

i.e., new relations between terms in Eq. (6.1.2): the bonus relations. The bonus scaling and the bonus relations have a number of important implications. In [93], it was shown that BCJ relations can be extracted from bonus relations. In the case of gravity, bonus relations have been used to simplify tree level calculations [26]. At loop level, the large z scaling of the BCFW shift corresponds to the high loop momenta limit; unsurprisingly improved scaling implies improved UV behavior [78][94][95].

In this paper, we prove that the inherent Bose-symmetry between gravitons directly implies this improved bonus scaling, completing the arguments of Ref. [21]. Bose-symmetry in General Relativity endows it with a purely on-shell description and constrains its UV divergences[1]. We further apply the same argument to gauge theories and gravity in various dimensions.

6.2 Completing on-shell constructability

Reference [21] first assumes n-point and lower amplitudes scale as z^{-1}—thereby ensuring Eq. (6.1.2) holds—and then checks if the BCFW expansion of the $(n+1)$-point

---

[1]The better than expected UV behavior was also at least partially understood from the "no-triangle" hypothesis of $\mathcal{N} = 8$ supergravity, as a consequence of crossing symmetry and the colorless nature of gravitons in Ref. [96]

amplitude factorizes correctly on all channels. Factorization on all channels is taken to define the amplitude. Correct factorization in most channels requires z^{-1} scaling of lower point amplitudes. However, some channels do not factor correctly without improved z^{-2} scaling, as well as a z^6 scaling on the "bad" shifts. In the following, we present a proof for both of these scalings.

Essentially, the argument rests on a very simple observation: any symmetric function $f(i,j)$, under deformations $i \to i + zk$, $j \to j - zk$, must scale as an even power of z. In particular, any function with a strictly better than $\mathcal{O}(1)$ large z behavior (no poles at infinity), is automatically guaranteed to decay at least as z^{-2}.

Although straightforward, this is not manifest when constructing the amplitude. BCFW terms typically scale as z^{-1}, but only specific pairs have canceling leading z^{-1} pieces. Similarly, the bad BCFW shift behavior of z^6 is only obtained when the leading z^7 pieces cancel in pairs.

Consider the five point amplitude in $\mathcal{N} = 8$ SUGRA, exposed by the $[1, 5\rangle$ BCFW shift, where $|1] \to |1] - z|5]$ and $|n\rangle \to |n\rangle + z|1\rangle$,

$$
\begin{aligned}
M_5 =& M(123P) \times M(P45)/P_{123}^2 + (4 \leftrightarrow 3) + (4 \leftrightarrow 2) \\
=& \frac{[23][45]}{\langle 12 \rangle \langle 13 \rangle \langle 23 \rangle \langle 24 \rangle \langle 34 \rangle \langle 45 \rangle \langle 15 \rangle^2} \\
& + \frac{[24][35]}{\langle 12 \rangle \langle 14 \rangle \langle 24 \rangle \langle 23 \rangle \langle 43 \rangle \langle 35 \rangle \langle 15 \rangle^2} \\
& + \frac{[43][25]}{\langle 14 \rangle \langle 13 \rangle \langle 43 \rangle \langle 42 \rangle \langle 32 \rangle \langle 25 \rangle \langle 15 \rangle^2},
\end{aligned}
\tag{6.2.1}
$$

with the SUSY-conserving delta-function stripped out.

Under a $[2, 3\rangle$ shift, the first term scales as z^{-2}, while the other two scale as z^{-1}. However, their sum (now symmetric in 2 and 3) scales as z^{-2}: the whole amplitude has the correct scaling. This pattern holds true in general. Where present, z^{-1}'s cancel between pairs of BCFW terms, $M_L(K_L, i, P) \times M_R(-P, j, K_R)$ and $M_L(K_L, j, P) \times M_R(-P, i, K_R)$. Further, terms without pairs over saturate the bonus scaling.

One such example is $M(1^-2^-3^+P^+)M(P^-4^-5^+6^+)$, appearing in M_6^{NMHV}. Under a $[4,3\rangle$ shift, it has no corresponding pair: $M(1^-2^-4^-P)M(P\,3^+5^+6^+)$ vanishes for all helicities h_P. Luckily, it turns out these types of terms have a surprisingly improved scaling of z^{-9}. Hence, they never spoil the scaling of the full amplitude.

In the next section we classify and prove the scalings of all possible BCFW terms. Following this, we demonstrate how leading z pieces cancel between BCFW terms.

6.3 BCFW terms under secondary z-shifts

Consider the $[1,n\rangle$ BCFW expansion of a n-point GR tree amplitude M_n (where $\tilde\lambda_1 \to \tilde\lambda_1 - w\tilde\lambda_n,\ \lambda_n \to \lambda_n + w\lambda_1$):

$$M_n = \sum_{L,R} \frac{M_L(\hat 1, \{L\}, \hat P)M_R(-\hat P, \{R\}, \hat n)}{P^2} \tag{6.3.1}$$

We would like to understand how BCFW terms in $\mathcal{M}_n$ scale under secondary $[i,j\rangle$ z-shifts

$$\tilde\lambda_i \to \tilde\lambda_i - z\tilde\lambda_j \qquad \lambda_j \to \lambda_j + z\lambda_i. \tag{6.3.2}$$

We recall two features of these terms as they appear in Eq. (6.3.1). First, the value of the primary deformation parameter $w = w_P$, which accesses a given term, is

$$w_P = \frac{P^2}{\langle 1|P|n]}, \tag{6.3.3}$$

and, on this pole, the intermediate propagator factorizes:

$$\hat P^{\alpha\dot\alpha} = \frac{\{[\tilde\lambda_n|P\}^\alpha\,\{\langle\lambda_1|P\}^{\dot\alpha}}{\langle\lambda_1|P|\tilde\lambda_n]} = \frac{\big|\lambda_P\big\rangle\big[\tilde\lambda_P\big|}{\langle 1|P|n]} \equiv |\hat P\rangle[\hat P|. \tag{6.3.4}$$

134

The little-group ambiguity amounts to associating the denominator with either λ_P, $\tilde{\lambda}_P$, or some combination of them. In what follows, we find it easiest to associate it entirely with the anti-holomorphic spinor, $|\hat{P}] = |\tilde{\lambda}_P]/\langle 1|P|n]$—see Eq. (6.3.5), below.

With this in hand, we now turn to the large z scalings of the various BCFW terms, subjected to the secondary z-shifts in Eq. (6.3.2). There will be two different types of BCFW terms: those with both i and j within the same subamplitude, and those with i and j separated by the propagator. The former inherit all z dependence from the lower point amplitudes in the theory, since the secondary shift acts like a usual BCFW shift on the subamplitude. The latter are more complicated, since the z shift affects the subamplitudes in several ways besides the simple shifts on i and j.

Specifically, both w_P and the factorized form of the internal propagator acquire z dependence:

$$
\begin{aligned}
w_P &= \frac{P^2}{\langle 1|P|n]} & &\longrightarrow & &\frac{P^2 + z\langle i|P|j]}{\langle 1|P|n] + z\langle 1i\rangle[jn]}, \\[2mm]
|\hat{P}\rangle^\alpha &\equiv \frac{\left|\left(\langle n|P\rangle\right)^\alpha\right.}{[jn]} & &\longrightarrow & &|\hat{P}\rangle^\alpha + z|i\rangle^\alpha, \\[2mm]
|\hat{P}]^{\dot{\alpha}} &\equiv \frac{\left|\left(\langle 1|P\rangle\right)^{\dot{\alpha}}\right.}{\langle 1|P|n]/[jn]} & &\longrightarrow & &\frac{|\tilde{\lambda}_P]^{\dot{\alpha}} - z\langle 1i\rangle|j]^{\dot{\alpha}}}{\langle 1|P|n]/[jn] - z\langle 1i\rangle}.
\end{aligned}
\qquad (6.3.5)
$$

With this factorized form of the propagator, it turns out that the left- and right-hand subamplitudes have well defined individual z scalings, which depend only on the helicity choices for i^{h_i}, j^{h_j} and P^h:

$$
\begin{aligned}
M_L(i^- P^-) &\sim z^{-2} & M_R(j^- P^-) &\sim z^{+2} \\[1mm]
M_L(i^- P^+) &\sim z^{-2} & M_R(j^- P^+) &\sim z^{+2} \\[1mm]
M_L(i^+ P^-) &\sim z^{+6} & M_R(j^+ P^-) &\sim z^{+2} \\[1mm]
M_L(i^+ P^+) &\sim z^{-2} & M_R(j^+ P^+) &\sim z^{-6}.
\end{aligned}
\qquad (6.3.6)
$$

The scaling of a full BCFW term $M_L M_R / P^2$ can then be easily determined from these values, which we prove in two steps.

First, note that the large z scalings on the left of Eq. (6.3.6) match the familiar BCFW scalings of full amplitudes. We prove this by showing that the large z behavior of the left-hand subamplitude maps isomorphically onto a BCFW shift of M_L. Looking at Eq. (6.3.5), we see that, in the large z limit, the spinors of i and P become

$$\lambda_i \longrightarrow \lambda_i \qquad\qquad\qquad \lambda_P \longrightarrow z\lambda_i$$

$$\tilde{\lambda}_i \longrightarrow -z\tilde{\lambda}_j \qquad\qquad\qquad \tilde{\lambda}_P \longrightarrow \tilde{\lambda}_j, \qquad (6.3.7)$$

which is just a regular BCFW $[i, P\rangle$ shift within the left-hand subamplitude.

Now we turn to the slightly unusual scalings on the right-hand side of Eq. (6.3.6). With the little-group choice in Eq. (6.3.5), the left-hand subamplitude has exactly the correct spinor variables to map onto the usual BCFW shift. Now observe that, starting with the other little-group choice for the spinors on the z shifted internal propagator, we obtain the usual BCFW scalings on this side:

$$M_R(j^- P^-) \sim z^{-2}$$

$$M_R(j^- P^+) \sim z^{+6}$$

$$M_R(j^+ P^-) \sim z^{-2}$$

$$M_R(j^+ P^+) \sim z^{-2}. \qquad (6.3.8)$$

Proving these results is identical to the previous reasoning for the left-hand subamplitude.

It becomes clear now that to get the other half of the scalings, we need only account for the change in z scaling when switching the $1/\langle 1|P(z)|n]$ factor between

λ_P and $\tilde{\lambda}_P$. Assume that the spinors of the propagator appear with weights[2]:

$$M_R \propto \left(|P\rangle\right)^a \left(|P]\right)^b, \tag{6.3.9}$$

where $-a + b = 2h_P$, and h_P is the helicity of the internal propagator as it enters the right-hand subamplitude. Now, in the limiting cases where $1/\langle 1|P(z)|n]$ is entirely associated with $|\lambda_P\rangle$ or $|\tilde{\lambda}_P]$ the amplitude scales as:

$$M_R \propto \left(\frac{|\lambda_P\rangle}{\langle 1|P|n]}\right)^a \left(|\tilde{\lambda}_P]\right)^b \to z^s, \quad \text{or} \tag{6.3.10}$$

$$M_R \propto \left(|\lambda_P\rangle\right)^a \left(\frac{|\tilde{\lambda}_P]}{\langle 1|P|n]}\right)^b \to z^t, \tag{6.3.11}$$

where s is the BCFW large z scaling exponent, obtained in Eq. (6.3.8), and t is the related scaling, for the other internal little-group choice. It follows that $s - t = b - a = 2h_P$, and so the t scalings can be easily derived as $t = s \pm 4$, depending on the helicity of the propagator.

Having proven all eight scaling relations in Eq. (6.3.6), we can classify the scaling behavior of all possible types of BCF terms with i and j in different subamplitudes. For these terms the propagator contributes a z^{-1} to each term, and so from Eq. (6.3.6) we obtain eight possible types of terms:

- $M_L(i^+P^-)M_R(j^-P^+)/P^2$ scales as z^{+7}, (6.3.12)

- $M_L(i^-P^-)M_R(j^+P^+)/P^2$ scales as z^{-9}, (6.3.13)

- The other six BCFW terms scale as z^{-1} . (6.3.14)

In the next section we will see how pairing terms improves these scalings by one power of z, such that we recover the required z^{-2} and z^6 scalings.

---

[2]In general, the spinors need not appear with uniform homogeneity. The analysis below still holds, but must be applied term by term. The same caveat applies to Eqs. (6.4.6) and (6.4.8).

Finally, while the individual scalings in Eq. (6.3.6) are not invariant under z dependent little-group rescalings on the internal line $\widehat{P}(z)$, the above results for full BCFW terms are invariant under these rescalings.

6.4 Improved behavior from symmetric sums

We first study $[+,+\rangle$ and $[-,-\rangle$ shifts, with scalings in Eq. (6.3.14). Define $M_L(K_L, i, P) \times M_R(-P, j, K_R)/P^2 \equiv M(i|j)$, where K_L is the momenta from the other external states on the left-hand subamplitude. We wish to show that in the large z limit

$$M(i|j) = -M(j|i). \tag{6.4.1}$$

so the leading z^{-1} pieces cancel in the symmetric sum of BCFW terms, $M(i|j) + M(j|i)$.

Because i and j have the same helicity, $M(j|i)$ is obtained directly from $M(i|j)$ by simply swapping labels:

$$M(i|j) = M(\lambda_i, \tilde{\lambda}_i, \lambda_j, \tilde{\lambda}_j) \tag{6.4.2}$$

$$M(j|i) = M(\lambda_j, \tilde{\lambda}_j, \lambda_i, \tilde{\lambda}_i) \tag{6.4.3}$$

In the large z limit, these become

$$M(i|j) = M(\lambda_i, -z\tilde{\lambda}_j, z\lambda_i, \tilde{\lambda}_j) \tag{6.4.4}$$

$$M(i|j) = M(z\lambda_i, \tilde{\lambda}_j, \lambda_i, -z\tilde{\lambda}_j) \tag{6.4.5}$$

The two have equal z scaling, and so can only differ by a relative sign. The spinors appear with weights

$$M(i|j) \propto \langle ij \rangle^F [ij]^G (\lambda_i)^a (\tilde{\lambda}_i)^b (\lambda_j)^c (\tilde{\lambda}_j)^d$$

$$M(j|i) \propto \langle ji \rangle^F [ji]^G (\lambda_j)^a (\tilde{\lambda}_j)^b (\lambda_i)^c (\tilde{\lambda}_i)^d, \tag{6.4.6}$$

while in the large z limit, the leading terms are

$$M(i|j) \propto z^{b+c} \left(\langle ij \rangle^F [ij]^G (\lambda_i)^a (-\tilde{\lambda}_j)^b (\lambda_i)^c (\tilde{\lambda}_j)^d \right)$$

$$M(j|i) \propto z^{a+d} \left(\langle ji \rangle^F [ji]^G (\lambda_i)^a (\tilde{\lambda}_j)^b (\lambda_i)^c (-\tilde{\lambda}_j)^d \right). \tag{6.4.7}$$

These cancel if and only if $F + G + b + d = $ odd. First, from Eq. (6.3.14), $M(a|b)$'s scale as z^{odd}. So $b + c = a + d = $ odd. Second, by helicity counting in Eq. (6.4.6), we know $-F + G - c + d = 2h_j = $ even. Therefore, we obtain the required result, and the leading z^{-1} pieces cancel.

For the $[-, +\rangle$ and $[+, -\rangle$ shifts a simple modification of the above argument is required. This is because we now expect the cancellation to occur between the pair terms $M_L(K_L, i^-, P^+) \times M_R(-P^-, j^+, K_R)/P^2$ and $M_L(K_L, j^+, P^-) \times M_R(-P^+, i^-, K_R)/P^2$. Switching different helicity particles requires us to flip the propagator's helicity as well. It can be shown that, in the large z-limit, $M_L(K_L, i^-, P^+) = M_L(K_L, j^+, P^-)$; likewise for the right-hand subamplitude. Note that switching i^- and j^+ requires more care now: functionally, the correct label swaps for M_L are $i \to P$, $P \to j$ while for M_R $j \to P$ and $P \to i$. Therefore we can write, as above,

$$M_L(i^-, P^+) \propto \langle iP \rangle^F [iP]^G (\lambda_i)^a (\tilde{\lambda}_i)^b (\lambda_P)^k (\tilde{\lambda}_P)^l$$

$$M_L(j^+, P^-) \propto \langle Pj \rangle^F [Pj]^G (\lambda_P)^a (\tilde{\lambda}_P)^b (\lambda_j)^k (\tilde{\lambda}_j)^l. \tag{6.4.8}$$

Crucially, the large z limit is also different for the two subamplitudes, since the limits (6.3.7) were obtained with $i \in P$. The second subamplitude instead has $j \in P$, and in this case the limits are $\lambda_P \to -z\lambda_i$ and $\tilde{\lambda}_P \to \tilde{\lambda}_j$. In the large z limit then identical counting as above shows that $a + b =$ even, and the same will hold for M_R. The propagator is antisymmetric in the large z limit under swapping i and j, and therefore the leading z pieces cancel as expected. This cancellation reduces the leading z^{-1} and z^{+7} scalings for the opposite helicity shifted BCF terms in the previous section, down to the well known z^{-2} and z^{+6} BCFW scalings for GR. This completes the proof of the bonus scaling for GR, and closes the final gap in the on-shell proof of BCFW in GR Ref. [21].

6.5   Analysis of the full amplitude.

The simple argument we used above can be applied directly to the whole amplitude, if we restrict to like-helicity shifts. Consider

$$A_n(i,j) \propto \langle ij \rangle^F [ij]^G (\lambda_i)^a (\tilde{\lambda}_i)^b (\lambda_j)^c (\tilde{\lambda}_j)^d. \tag{6.5.1}$$

If this amplitude is manifestly symmetric under exchange of two (bosonic) particle labels, then $A_n(i,j) = A_n(j,i)$, which fixes $a = c$, $b = d$, and $F + G =$ even. By helicity counting, $-F + G - a + b = 2h_i =$ even, and then $a + b =$ even. So, under a $[i,j\rangle$ shift,

$$A_n(i(z), j(z)) \sim z^{b+c} = z^{a+b} = z^{\text{even}}. \tag{6.5.2}$$

This same logic holds in Eq. (6.5.1), even if the shifted lines are identical fermions. Permuting labels i and j again forces $a = c$, and $b = d$, and $F + G =$ odd. But so

must $2h_i = -F + G - a + b$. Hence $a + b$ remains even. BCFW shifts of identical particles, bosons or fermions, fix z^{even} scaling at large z.

To understand the opposite-helicity shifts, we are led to consider pure GR as embedded within maximal $\mathcal{N} = 8$ SUGRA. Amplitudes in maximal supergravity do not distinguish between positive and negative helicity graviton states. Using the methods of [97] to truncate to pure GR, we recover the usual BCFW scalings.

As an interesting corollary of our four-dimensional analysis, the large z scaling of gravity amplitudes in three dimensions is drastically improved to z^{-4}. Due to the fact that the little group in three dimensions is a discrete group, the BCFW deformation is non-linear. In particular the three dimensional spinors shift as [98]:

$$\lambda_i(z) = \text{ch}(z)\lambda_i + \text{sh}(z)\lambda_j \,, \ \lambda_j(z) = \text{sh}(z)\lambda_i + \text{ch}(z)\lambda_j \tag{6.5.3}$$

where $\text{ch}(z) = (z + z^{-1})/2$ and $\text{sh}(z) = (z - z^{-1})/2i$. Thus, momenta shift as

$$p_i(z) = \overline{P_{ij}} + yq + \frac{1}{y}\tilde{q} \,, \ \ p_j(z) = \overline{P_{ij}} - yq - \frac{1}{y}\tilde{q} \tag{6.5.4}$$

where $\overline{P_{ij}} = \frac{p_i + p_j}{2}$, $y = z^2$, and $q, \tilde{q}$ can be read off from Eq. (6.5.3). Now let's consider three-dimensional gravity amplitudes that arise from the dimension reduction of four-dimensional gravity theory. The degrees of freedom are given by a dilaton and a scalar. Since both are bosons, little group dictates that one must have even power of λ_i. Thus the large z behavior of gravity amplitudes is completely dictated by Eq. (6.5.4). Permutation invariance then requires the function to be symmetric under $y \leftrightarrow -y$, and so must be an even power of y. Thus if gravity amplitudes can be constructed via BCFW shift, the large z asymptotic behavior must be at most $y^{-2} = z^{-4}$. Indeed it is straightforward to check that the four-point $\mathcal{N} = 16$ supergravity amplitude behaves as z^{-4} under a super-BCFW shift. This is to be compared with the z^{-1} scaling of superconformal Chern-Simons theory [98].

More generally, BCFW shifts in $d \geq 4$ take the form,

$$p_i^\mu(z) = p_i^\mu + z\,q^\mu \qquad p_j^\mu(z) = p_j^\mu - z\,q^\mu, \qquad (6.5.5)$$

where q is null and orthogonal to p_i and to p_j. External wave-functions of shifted boson lines also shift [39]. For identical bosons, Bose-symmetry disallows z^{odd} scaling, as it would introduce a sign change under label swaps. Identical fermions shift similarly; here the antisymmetric contraction of the identical spinor wave-functions absorbs their exchange-sign. BCFW shifts of identical particles must scale as z^{even} for large z in dimensions $d \geq 4$.

Symmetry between identical particles is crucial for these cancellations to occur. Gluon partial amplitudes are not permutation invariant: distinct gluons generally have different colors. This spoils the permutation invariance—as is clear from z^{-1} drop-off of adjacent shifts of a color-ordered tree amplitude in Yang-Mills. Gravitons, however, are unique: they cannot have different "colors" [61]. Thus graviton amplitudes are invariant under permutations from the outset: the discrete symmetry group of graviton amplitudes is larger than for gluon amplitudes. Consequently, gravity amplitudes are softer in the deep-UV than Yang-Mills amplitudes.

6.6   Bose-symmetry and color in Yang-Mills

Finally, we explore the interplay between color and the large z structure of Yang-Mills amplitudes. For ease, we focus on $A_4^{\mathrm{tree}}(1^-, 2^-, 3^+, 4^+)$. It can be written in terms of color-ordered partial amplitudes as

$$\frac{A_4(1^-2^-3^+4^+)}{\langle 12 \rangle^2 [34]^2} = \frac{Tr(1234)}{st} + \frac{Tr(1243)}{su} + \frac{Tr(1324)}{tu}. \qquad (6.6.1)$$

Under a $[1, 2\rangle$ shift, only t and u shift, and in opposite directions: $\hat{t}(z) = t + z\langle 1|4|2]$, and $\hat{u}(z) = u - z\langle 1|4|2]$. The term proportional to $Tr(1324)$ scales as z^{-2}, while the other two scale as z^{-1}. The leading z terms,

$$\frac{A_4(\hat{1}^-, \hat{2}^-, 3^+, 4^+)}{\langle 12\rangle^2 [34]^2} \sim \frac{Tr(1234) - Tr(1243)}{z\langle 1|4|2]\, s} + \cdots, \tag{6.6.2}$$

cancel when gluons 1 and 2 are identical, and $T_1 = T_2$.

Cancellation of z^{-1} terms must hold for general tree amplitudes when the gluons have the same color labels. However, only BCFW shifts of lines that are *adjacent* in color-ordering cancel pairwise as in Eq. (6.6.2). For color-orderings where this shift is non-adjacent, there are no pairs of BCF terms with canceling z^{-1}-terms. This implies that good non-adjacent BCFW shifts in gluon partial amplitudes must scale as z^{-2}.

6.7   Future directions and concluding remarks.

We have shown the z^{-2} bonus scalings/relations, crucial for consistent on-shell contraction of Gravitational S-matrices, follow from Bose-symmetry. Similar z^{-1} cancellations occur in QED and GR [99]. Further, Bose-symmetry alone implies z^{-2} drop-off of non-adjacent BCFW shifts in Yang-Mills. More broadly, BCFW shifts of identical particles—bosons and fermions—must scale as z^{even} in general settings, beyond $d = 4$.

Graviton amplitudes in Refs. [100][101][102], which manifest permutation symmetry, also manifest z^{-2} drop-off. This is not a coincidence: permutation symmetry automatically implies bonus behavior. A better understanding of gravity should be tied to more natural manifestations of permutation invariance. However, not all improved scalings obviously come from permutation invariance. Notably, Hodges' observation that BCFW-terms, built from "bad" "opposite helicity" z^{-1} $\mathcal{N} = 7$ SUGRA

shifts, term-by-term scale as z^{-2} [103]. As the legs are not identical, permutation invariance is not prominent in the proof [104].

Permutation invariance has unrecognized and powerful consequences even at tree level. Do new constraints appear when accounting for it in other shifts? Does it have non-trivial consequences at high-loop orders in $\mathcal{N} = 8$ SUGRA, or $\mathcal{N} = 4$ SYM? Would mandating it expose new facets of the "Amplituhedron" of Ref. [90]?

Appendix A

Appendix

A.1 Inducting the $[k, i\rangle$ shift

In this section we show that the a shift $[k, i\rangle$ corresponding to case (4.3.5a), imposes case (4.3.5c) constraints on B_n^k, since B_n^k is not a function of e_k. Succinctly, we want to show:

$$[k, i\rangle[B_{n+1}] \propto z^{-1} \Rightarrow [\overline{k}, i\rangle[B_n^k] \propto z^1 \qquad (A.1.1)$$

The numerator becomes (4.3.11):

$$\mathcal{O}(z^{-1}) \propto \sum_{r \neq i,k} e.e_r B_n^r + (e.e_i + ze.p_k \frac{e_k.p_i}{p_i.p_k}) B_n^i + e.e_k B_n^k$$
$$+ \sum_{r \neq i,k} e.p_r C_n^r + (e.p_i - ze.e_k) C_n^i + (e.p_k + ze.e_k) C_n^k \qquad (A.1.2)$$

in this case $e.e_k$ is no longer unique, so our constraint now involves other functions:

$$e.e_k(B_n^k + z(C_n^k - C_n^i)) \propto \mathcal{O}(z^{-1}) \qquad (A.1.3)$$

However we can still obtain an upper bound for B_n^k, if we find one for C_n^k and C_n^i.

Observe that under this shift, the prefactors $e.e_i$ and $e.p_i$ are unique, so $B^i, C^i \propto z^{-1}$. Next the term proportional to $e.p_k$ is:

$$e.p_k(C_n^k + z\frac{e_k.p_i}{p_i.p_k}B_n^i) \propto e.p_k(C_n^k + \mathcal{O}(z^0)) \propto \mathcal{O}(z^{-1}) \tag{A.1.4}$$

which implies C_n^k can be at most $\propto \mathcal{O}(z^0)$. Therefore eq. (A.1.3) becomes:

$$B_n^k + z(C_n^k - C_n^i) \propto B_n^k + \mathcal{O}(z^1) - \mathcal{O}(z^0) \propto \mathcal{O}(z^{-1}) \tag{A.1.5}$$

so B_n^k is at most $\propto \mathcal{O}(z^1)$ under this shift, as required.

A.2 Inducting the $[1, n\rangle$ shift

In this section we consider the $[1, n\rangle$ and $[n, 1\rangle$ shifts. These are non-adjacent in B_{n+1}, while in B_n^k they are adjacent, and are also affected by the pole shift. We will consider just the case where both B_{n+1} and B_n^k are functions of e_1 and e_n. Therefore if $B_{n+1} \propto z^{-2}$ under the non-adjacent $[1, n\rangle$ shift, we expect to obtain $B^k \propto z^0$: one power from the shift in the denominator, and one from becoming an adjacent shift. So we want to show:

$$[1, n\rangle[B_{n+1}] \propto z^{-2} \Rightarrow [i, j\rangle[B_n^k] \propto z^0 \tag{A.2.1}$$

Since under this shift B_n^k can cancel against B_1^k, we can no longer treat the numerators as independent, so we must consider both functions in (4.3.11):

$$z^{-2} \propto [1, n\rangle \left[\frac{e.e_k B_1^k}{q.p_1} + \frac{e.e_k B_n^k}{q.p_n}\right] = e.e_k \left(\frac{B_1^k}{q.p_1 + zq.e_1} + \frac{B_n^k}{q.p_n - zq.e_1}\right) = \tag{A.2.2}$$

$$= \frac{1}{z}\frac{B_1^k - B_n^k}{q.e_1} - \frac{1}{z^2}\frac{q.p_1 B_1^k - q.p_n B_n^k}{(q.e_1)^2} + \mathcal{O}(z^{-3}) \tag{A.2.3}$$

146

which implies $\frac{1}{z^2}\frac{q.p_n B_n^k}{(q.e_1)^2} \propto z^{-2}$ due to the unique prefactor, so $B_n^k \propto z^0$, which is the expected result.

A.3 Ruling out $B'(a)$ functions with extra momenta

We will show that $B'(a)$ functions (4.3.17) vanish under $G_n^h(a)$ constraints if $B(a)$ functions (4.3.4) also vanish. We can express the former in terms of the latter as:

$$B'(a) = B(a)\sum_{r,s} a_{rs}p_r.p_s \tag{A.3.1}$$

Then we must show:

$$[i,j\rangle[B'(a)] \propto z^m \Rightarrow [i,j\rangle[B(a)] \propto z^m \tag{A.3.2}$$

Under a shift $[i,j\rangle$, with $r,s \neq i,j$, we have:

$$
\begin{aligned}
p_r.p_s &\to p_r.p_s \\
p_i.p_j &\to p_i.p_j \\
p_i.p_r &\to p_i.p_r - ze_i.p_k \\
p_j.p_r &\to p_j.p_r + ze_i.p_k
\end{aligned}
\tag{A.3.3}
$$

Assume that under this shift $B'(a) \propto \mathcal{O}(z^m)$, but that $B(a)$ goes like a higher power, $\mathcal{O}(z^{m+1})$. Equation (A.3.1) becomes:

$$z^m B'^0 + \mathcal{O}(z^{m-1}) = \left(z^{m+1} B^1 + z^m B^0 + \mathcal{O}(z^{m-1}) \right) \left(z \sum_r (a_{jr} - a_{ir}) e_i.p_r + \sum_{rs} a_{rs} p_r.p_s \right) =$$

$$= z^{m+2} \left(B^1 \sum_r e_i.p_r (a_{jr} - a_{ir}) \right)$$

$$+ z^{m+1} \left(B^1 \sum_{rs} a_{rs} p_r.p_s + B^0 \sum_r (a_{jr} - a_{ir}) e_i.p_r \right)$$

$$+ \mathcal{O}(z^m) \tag{A.3.4}$$

On the left side we have only $\mathcal{O}(z^m)$, so the higher orders on the right must vanish. Order z^{m+2} vanishing implies either $B^1 = 0$, or $\sum_r e_i.p_r (a_{jr} - a_{ir}) = 0$. In the latter case, vanishing at order z^{m+1} implies $B^1 = 0$. Either way then we must have $B \propto z^m$, proving eq. (A.3.2), and so functions of the type (4.3.17) are ruled out.

A.4 Constructing minimal numerators

Here, we prove that the minimal numerator for a four-point amplitude of massless particles satisfies Eq. (5.3.5) for the special case where the sum of all four helicities vanishes.[1] Given $A_4(1^{h_1}, 2^{h_2}, 3^{h_3}, 4^{h_4})$, we re-label the external states by increasing helicity:

$$H_1 \geq H_2 \geq H_3 \geq H_4 \tag{A.4.1}$$

The total helicity vanishes, and thus $H_1 \geq 0$ and $H_4 \leq 0$. Now, define $H_1^+ = |H_1|$ and $H_4^- = |H_4|$: the numerator has at least $2H_1^+$ $\tilde{\lambda}$s, and at least $2H_4^-$ λs.

---

[1] All four-point tree-amplitudes constructed from a set of three-point amplitudes and their conjugate amplitudes, A_3 & $\bar{A}_3$, have this property.

Now there must exist

$$N_{\tilde{\lambda}}^{\text{ext.}} = 2H_1^+ + N_{\tilde{\lambda}}^{\text{rest}} \geq 2H_1^+ \tag{A.4.2}$$

total external $\tilde{\lambda}$s in the numerator. Similarly, the numerator must contain a total of

$$N_{\lambda}^{\text{ext.}} = 2H_4^- + N_{\lambda}^{\text{rest}} \geq 2H_4^- \tag{A.4.3}$$

external λs.

By definition, the numerator is both (a) Lorentz-invariant, and (b) little-group covariant, and (c) encodes all of the helicity information of the asymptotic scattering states. Therefore, it must be of the form

$$\text{Numerator} \sim \langle \, \rangle_{(1)} ... \langle \, \rangle_{(n)} \; [\,]_{(1)} ... [\,]_{(m)} \; . \tag{A.4.4}$$

Notably, requiring $\sum_{a=1}^{4} H_a = 0$ directly implies that the numerator contains equal number of holomorphic and anti-holomorphic spinor-helicity variables,

$$\sum_{a=1}^{4} H_a = 0 \Leftrightarrow \left\{ N_{\lambda}^{\text{ext.}} = N_{\tilde{\lambda}}^{\text{ext.}} \; , \; \text{and} \; N_{\lambda}^{\text{total}} = N_{\tilde{\lambda}}^{\text{total}} \right\} \tag{A.4.5}$$

We note that because the numerator contains the same number of λs as $\tilde{\lambda}$s, and must be a product of spinor-brackets, then it must have the same number of each type of spinor-product:

$$\mathcal{N}_b = N_{\langle \, \rangle} = N_{[\,]} \; . \tag{A.4.6}$$

The reason is as follows. First, note that only inner-products of spinor-helicity variables are both (a) Lorentz-invariant and (b) little-group covariant. Because the numerator has both of these properties, all of the λs and $\tilde{\lambda}$s which encapsulate the

149

helicity information of the asymptotic scattering states must be placed within spinor-brackets. If we take $N_{\langle\,\rangle} \neq N_{[\,]}$, then there would be a mis-match between the number of λs and $\tilde\lambda$s in the numerators. This contradicts the statement that the numerator must contain the same number of positive- and negative- chirality spinor-helicity variables. This proves Eq. (A.4.6).

Now, the *minimal* number of spinor-brackets $\mathcal{N}$ is simply given by

$$\mathcal{N}_b = 2 \times \max\{H_1^+, H_4^-\} \tag{A.4.7}$$

This can be seen in the following way. At the minimum, there must be $2H_1^+$ [ ]s and $2H_4^-$ $\langle\,\rangle$s within the numerator. Otherwise, at least two of the $2H_1^+$ copies of $\tilde\lambda_1$ within the numerator would have to within the same spinor bracket, $[\tilde\lambda_1, \tilde\lambda_1]$. But this would force the numerator to vanish. As we are only concerned with non-trivial amplitudes, we thus require $N_{[\,]} \geq 2H_1^+$. The same logic requires $N_{\langle\,\rangle} \geq 2H_4^-$.

But, by Eq. (A.4.6), we must have $N_{[\,]} = N_{\langle\,\rangle}$. So we must have $\mathcal{N}_b \geq 2 \times \max\{H_1^+, H_4^-\}$. The minimal numerator saturates this inequality. This proves Eq. (A.4.7).

It is important to note that this minimal number of spinor-brackets of each type, $\mathcal{N}_b = 2\max\{H_1^+, H_4^-\}$, mandated by Eq. (A.4.7) to be present within the numerator is "large" enough to encode the helicity information of all of the external scattering states—not just the helicity information of $1^{+H_1^+}$ and $4^{-H_4^-}$.

In other words, there are "enough" $\langle\,\rangle$s and [ ]s already present in the numerator to fit in the remaining λs and $\tilde\lambda$s required to encode the helicity information of the other two particles. I.e.,

$$N_{[\,]} \geq N_{\tilde\lambda}^{\text{rest}} \text{ , and } N_{\langle\,\rangle} \geq N_{\lambda}^{\text{rest}} \tag{A.4.8}$$

Before proving this, first recall eqs. (A.4.2), (A.4.3), and (A.4.7). By (A.4.7), $N_{[\ ]} = N_{\langle\ \rangle} = 2\max\{H_1^+, H_4^-\}$. Now, how many λs and $\tilde{\lambda}$s must be present in the numerator to ensure all external helicity data is properly entered into the numerator? There are only three cases to consider. In all cases, Eq. (A.4.8) holds:

1. Particles 1 and 2 have positive helicity, while particles 3 and 4 have negative helicity. Now, by definition, we would like to show that of the $2\max\{H_1^+, H_4^-\}$ [ ]s required by Eq. (A.4.7) are sufficiently numerous to allow inclusion of $2|H_2|$ more $\tilde{\lambda}_2$s. This is guaranteed by the orderings: $H_1 \geq H_2$. So there are enough empty slots in the anti-holomorphic spinor-brackets to encode the helicity of all positive-helicity particles. The same holds for H_3. For this case, Eq. (A.4.8) holds.

2. Only particle 4 has negative helicity. All others have positive helicity. Because the amplitudes under consideration have total helicity zero, we know that the sum of helicities of the particles with positive helicity must equal H_4^-. Hence there must be $2H_4^- \lambda_4$s and $2H_4^-$ physical $\tilde{\lambda}$s in the numerator. Further, as only particle 4 has negative helicity, it follows that $\mathcal{N}_b = 2\max\{H_1^+, H_4^-\} = 2H_4^-$. And so there are $2H_4^-$ spinor-brackets of each kind. For this case Eq. (A.4.8) holds.

3. Only particle 1 has positive helicity. This case is logically equivalent to the above.

This proves Eq. (A.4.8), and therefore proves Eq. (5.3.5):

$$N \sim \langle\ \rangle_{(1)}...\langle\ \rangle_{(2H)} \ [\ ]_{(1)}...[\ ]_{(2H)} \Rightarrow [N] = (K^2)^{2H} \ , \tag{A.4.9}$$

where $H = \max\{|h_1|, ..., |h_4|\}$. Establishing this result concludes the proof.

A.5 Ruling out theories with $N_p = 2$, for $A \geq 3$

In this section, we rule out self-interacting theories constructed from three-point amplitudes which necessitate two poles in the four-point amplitudes, for $A > 2$. This is simple pole-counting, augmented by constraint (5.3.1) and the results of subsection 5.3.3 for $N_p = 3$. Recall, for amplitudes within a self-interacting sector of a theory, we have $\sum_{i=1}^{4} h_i^{\text{ext}} = 0$, and $N_p = 2H + 1 - A$. For $N_p = 2$, we must have $H = \max\{|f|, |H|, |g|\} = (A+1)/2$.

Within this sector we may construct $A_4(1^{+H}, 2^{-H}, 3^{+f}, 4^{-f})$ from $A_3(H, g, f)$. By assumption, it has two factorization channels, specifically the t- and u-channels.[2] Without loss of generality, we take $f > 0$. The intermediary in the u-channel pole has spin $g = A - (H + f) = (A-1)/2 - f < H = (A+1)/2$, and poses no barrier to a unitary & local interpretation of the amplitude/theory.

However, for the t-channel's intermediary must have helicity $\tilde{H} = A + f - H = (A-1)/2 + f$. And so, for $f > 1$, we must include a new state with larger helicity $\tilde{H} = H + (f-1) > H$. As discussed in subsection 5.3.3, this does not a priori spell doom for the theory. However, in this case it does: inclusion of this new, larger helicity, state within the theory forces inclusion of four-point amplitudes with these states on external lines. These *new* amplitudes have a larger number of poles: $2\tilde{h} + 1 - A = (2H + 1 - A) + 2f = 2 + 2f \geq 4 > 3$, for $f > 1$.

Theories with $H = (A+1)/2$, and $f > 1$ (or $g > 1$) cannot be consistent with unitarity and locality. Inspection reveals that all theories with $N_p = 2$ and $A > 2$ are of this type, save for three special examples. Explicitly, for $A = 3, 4, 5$, and $A = 6$,

---

[2]The s-channel in this amplitude is disallowed, as it would require a new particle with helicity $\tilde{H} = \pm A$, which would lead to amplitudes with $N_p = A + 1$.

the $N_p = 2$ theories are defined by (H, g, f)s of the following types,

$$A = 3: \quad (2, 2, -1) \,, \quad \left(2, \frac{3}{2}, -\frac{1}{2}\right) \,, \quad (2, 1, 0) \,, \quad \left(2, \frac{1}{2}, \frac{1}{2}\right) \,, \tag{A.5.1}$$

$$A = 4: \quad \left(\frac{5}{2}, \frac{5}{2}, -1\right) \,, \quad \left(\frac{5}{2}, 2, -\frac{1}{2}\right) \,, \quad \left(\frac{5}{2}, \frac{3}{2}, 0\right) \,, \quad \left(\frac{5}{2}, 1, \frac{1}{2}\right) \,, \tag{A.5.2}$$

$$A = 5: \quad (3, 3, -1) \,, \quad \left(3, \frac{5}{2}, -\frac{1}{2}\right) \,, \quad (3, 2, 0) \,, \quad \left(3, \frac{3}{2}, \frac{1}{2}\right) \,, \quad (3, 1, 1) \,, \tag{A.5.3}$$

$$A = 6: \quad \left(\frac{7}{2}, \frac{7}{2}, -1\right) \,, \quad \left(\frac{7}{2}, 3, -\frac{1}{2}\right) \,, \quad \left(\frac{7}{2}, \frac{5}{2}, 0\right) \,, \quad \left(\frac{7}{2}, 2, \frac{1}{2}\right) \,, \quad \left(\frac{7}{2}, \frac{3}{2}, 1\right) \,, \tag{A.5.4}$$

and their conjugate amplitudes. Clearly, all but the last two entries on line (A.5.1) and the last entry on line (A.5.2) have $f > 1$ (thus $\tilde{H} > (A+1)/2$), and are inconsistent. Further, it is clear that all higher-A $N_p = 2$ theories may only have pathological three-point amplitudes, which indirectly lead to this same tension with locality and unitarity: except for those three special cases, all $N_p = 2$ theories must have fs that are larger than unity.

It is a simple exercise to show that these three pathological examples are inconsistent: playing around with the factorization channels of $A_4(H, -H, f, -f)$s reveals that again the t-channel is the problem. The t-channel requires the three-point amplitudes on lines (A.5.1) or (A.5.2) which are *directly* ruled out, as they have $H = (A+1)/2$ and $f > 1$.

No three-point amplitude with $N_p = 2H + 1 - A = 2$ can lead to a constructible S-matrix consistent with locality and unitarity, for any A larger than two.

A.6 Uniqueness of spin-3/2 states

In this appendix, we will use similar arguments to those in section 5.4 to show that massless spin-3/2 states can only couple consistently to massless particles with helicities $|H| \leq 2$. Recall that for $A > 2$, no constructible theory can be consistent with unitarity and locality unless $A/3 \leq H \leq A/2$. To see whether-or-not the grav-

itino discussed in section 5.6 can couple to any higher-A amplitude, we simply study four-point amplitudes with factorization channels of the type,

$$A_4(1^{+2}, 2^{-\frac{3}{2}}, 3^{-c}, 4^{-d}) \to A_3^{(GR)}(1^{+2}, 2^{-\frac{3}{2}}, P^{+\frac{3}{2}}) \frac{1}{s} A_3^{(A)}(P^{-\frac{3}{2}}, 3^{-c}, 4^{-d}). \qquad \text{(A.6.1)}$$

Note that only two three-point amplitudes are consistent with the dual requirements (a) $A/3 \le H \le A/2$ and (b) $3/2 \in \{h_1, h_2, h_3\}$ for $A > 2$. These theories, and their corresponding four-particle amplitudes, are:

$$A_3(A/2 - 1/2, 3/2, A/2 - 1) \Rightarrow A_4(1^{+2}, 2^{-\frac{3}{2}}, 3^{-(A/2-1/2)}, 4^{-(A/2-1)}), \text{ and}$$

$$A_3(A/2, 3/2, A/2 - 3/2) \Rightarrow A_4(1^{+2}, 2^{-\frac{3}{2}}, 3^{-A/2}, 4^{-(A/2-3/2)}). \qquad \text{(A.6.2)}$$

Now, the minimal numerator which encodes the helicity information of, for instance, the first amplitude is,

$$N \sim [1|P|2\rangle^3 [1|Q|3\rangle \left(\langle 3, 4\rangle^2\right)^{A/2-1} \Rightarrow [N] = \left(K^2\right)^{(A/2)+3}. \qquad \text{(A.6.3)}$$

However, by power-counting, the kinematic part of the amplitude must have mass-dimension,

$$\left[\frac{N}{f(s,t,u)}\right] = \left[\frac{1}{K^2} A_{\text{Left}}^{(GR)} A_{\text{Right}}^{(A)}\right] = \frac{(K^2)^{1/2} (K^2)^{A/2}}{(K^2)} = (K^2)^{A/2}, \qquad \text{(A.6.4)}$$

and thus the denominator, $f(s,t,u)$ must have mass-dimension three:

$$[f(s,t,u)] = (K^2)^3 \Rightarrow f(s,t,u) = s\,t\,u\,! \qquad \text{(A.6.5)}$$

However, as is obvious from inspection of any amplitude for $A \ge 4$, two of these factorization channels require inclusion of states with helicities which violate the most basic constraint (5.3.1). Thus, no spin-3/2 state in any three-point amplitude with

154

$A \geq 4$ can be identified with the gravitino of section 5.6. The sole exception to this is the three-point amplitude $(H, A) = (3/2, 3)$:

$$A_3(1^0, 2^{+\frac{3}{2}}, 3^{+\frac{3}{2}}) \Rightarrow A_4(1^{+2}, 2^{-\frac{3}{2}}, 3^{-\frac{3}{2}}, 4^0). \tag{A.6.6}$$

Factorization channels in this putative amplitude necessitate only either scalar or gravitino exchange, and are thus not in obvious violation of the consistency condition (5.3.1).

A.7 F^3- and R^3-theories and SUSY

Basic counting arguments show us that the F^3- and R^3- theories, i.e. the S-matrices constructed from $A_3(1, 1, 1)$ and $A_3(2, 2, 2)$ and their conjugates, are not compatible with leading-order (SUGRA) interactions with spin-3/2 states. The argument is simple.

First, we show that F^3-theories are not supersymmetrizable. Begin by including the minimal $\mathcal{N} = 1$ SUGRA states, together with the three-particle amplitudes which couple gluons to the single species of spin-3/2 (gravitino) state that construct the $\mathcal{N} = 1$ SYM multiplet. Additionally, allow the F^3-three-point amplitude as a building-block of the S-matrix. In other words, begin consider the four-particle S-matrix constructed from,

$$A_3\left(2, \frac{3}{2}, -\frac{3}{2}\right), \ A_3\left(\frac{3}{2}, 1, -\frac{1}{2}\right), \ A_3\left(1, \frac{1}{2}, -\frac{3}{2}\right), \ \text{and } A_3\left(1, 1, 1\right), \tag{A.7.1}$$

where all spin-1 states are gluons, and all spin-1/2 states are gluinos. Now, consider the four-particle amplitude, $A_4(+\frac{3}{2}, -\frac{1}{2}, -1, -1)$. On the s-channel, it factorizes

nicely:

$$A_4\left(1^{+\frac{3}{2}}, 2^{-\frac{1}{2}}, 3^{-1}, 4^{-1}\right)\bigg|_{s\to 0} \to \frac{1}{s} A_3\left(1^{\frac{3}{2}}, 2^{-\frac{1}{2}}, P^{+1}\right) A_3\left(P^{-1}, 3^{-1}, 4^{-1}\right). \quad \text{(A.7.2)}$$

Clearly, it fits into the theory defined in Eq. (A.7.1). To proceed further, we note that its minimal numerator must have the form,

$$N \sim [1|P|2\rangle[1|Q|3\rangle[1|K|4\rangle\langle 34\rangle, . \quad \text{(A.7.3)}$$

Now, this amplitude must have kinematic mass-dimension,

$$\left[\frac{A_3^{\text{SUGRA}} A_3^{F^3}}{K^2}\right] = \left[\frac{[\ ]^2\langle\ \rangle^3}{\langle\rangle[\ ]}\right] = \left(K^2\right)^{1+\frac{1}{2}}. \quad \text{(A.7.4)}$$

Combining Eq. (A.7.3) and Eq. (A.7.4), we see that $1/f(s,t,u)$ must have two poles. On the *other* pole, say on the $t \to 0$ pole, it takes the form

$$A_4\left(1^{+\frac{3}{2}}, 2^{-\frac{1}{2}}, 3^{-1}, 4^{-1}\right)\bigg|_{t\to 0} \to \frac{1}{t} A_3\left(1^{\frac{3}{2}}, 4^{-1}, P^{+\Delta}\right) A_3\left(P^{-\Delta}, 2^{-\frac{1}{2}}, 3^{-1}\right). \quad \text{(A.7.5)}$$

Now, one of these two sub-amplitudes must have $A = 3$. However, recall that in section 5.4 we showed that the only three-point amplitude which may have spin-1 states identified with the gluons is the $A_3(1,1,1)$ amplitude. Observe that, regardless of which amplitude has $A = 3$, *both* amplitudes contain one spin-1 state and another state with spin-$s \neq 1$. Therefore *neither* amplitude can consistently couple to the $A = 1$ gluons. Therefore we conclude that $\mathcal{N} = 1$ supersymmetry is incompatible with F^3-type interactions amongst gluons.

Similar arguments show that the three-point amplitudes arising from R^3-type interactions cannot lead to consistent S-matrices, once spin-3/2 gravitinos are included in the spectrum. Again, we first specify the four-particle S-matrix as constructed

156

from the following primitive three-particle amplitudes:

$$A_3\left(2, \frac{3}{2}, -\frac{3}{2}\right), \text{ and } A_3\left(2, 2, 2\right),$$ (A.7.6)

Now, consider the four-particle amplitude, $A_4(+\frac{3}{2}, -\frac{3}{2}, -2, -2)$, an analog to that considered in the F^3-discussion. On the s-channel, it factorizes nicely:

$$A_4\left(1^{+\frac{3}{2}}, 2^{-\frac{3}{2}}, 3^{-2}, 4^{-2}\right)\Big|_{s \to 0} \to \frac{1}{s} A_3\left(1^{\frac{3}{2}}, 2^{-\frac{3}{2}}, P^{+2}\right) A_3\left(P^{-2}, 3^{-2}, 4^{-2}\right).$$ (A.7.7)

Clearly, it fits into the theory defined in Eq. (A.7.6). To proceed further, we note that its minimal numerator must have the form,

$$N \sim [1|P|2\rangle^3 \left(\langle 34\rangle^2\right)^2.$$ (A.7.8)

Now, this new amplitude must have kinematic mass-dimension,

$$\left[\frac{A_3^{\text{SUGRA}} A_3^{R^3}}{K^2}\right] = \left[\frac{[\ ]^2 \langle\ \rangle^6}{\langle\rangle[\ ]}\right] = \left(K^2\right)^3.$$ (A.7.9)

Combining Eq. (A.7.8) and Eq. (A.7.9), we see that $1/f(s, t, u)$ must have two poles, again, as in the F^3-discussion above. On the *other* pole, say on the $t \to 0$ pole, it takes the form

$$A_4\left(1^{+\frac{3}{2}}, 2^{-\frac{3}{2}}, 3^{-2}, 4^{-2}\right)\Big|_{t \to 0} \to \frac{1}{t} A_3\left(1^{\frac{3}{2}}, 4^{-2}, P^{+\Delta}\right) A_3\left(P^{-\Delta}, 2^{-\frac{3}{2}}, 3^{-2}\right).$$ (A.7.10)

Reasoning isomorphic to that which disallowed $\mathcal{N} = 1$ SUSY and F^3-gluonic interactions rules out the compatibility of this given factorization channel with $\mathcal{N} = 1$ SUSY and R^3-effective gravitational interactions. Namely, it must be that one of these two sub-amplitudes must have $A = 6$. However, recall that in section 5.4 we showed that the only three-point amplitude which may have spin-1 states identified

with the gluons is the $A_3(1,1,1)$ amplitude. Observe that, regardless of which amplitude has $A = 6$, *both* amplitudes contain one spin-1 state and another state with spin-$s \neq 2$. Therefore *neither* amplitude can consistently couple to the $A = 2$ gravitons. Therefore we conclude that $\mathcal{N} = 1$ supersymmetry is *also* incompatible with R^3-type interactions amongst gravitons.

A.8   Justifying the complex deformation in section 5.5

One might worry about the validity of such a shift, and how it could be realized in practice. In other words, one could wonder whether-or-not shifting the Mandlsetam invariants, $(s, t, u) \rightarrow (s + z\tilde{s}, t + z\tilde{t}, u + z\tilde{u})$ would not also shift the numerator of the amplitude. Here, we prove that such a shift must always exist.

First, a concrete example. Suppose one desired to study the constraints on the f_{abc}s characterizing $A_3(1_a^{+1}, 2_b^{-1}, 3_c^{-1}) = f_{abc}\langle 23\rangle^3/\{\langle 31\rangle\langle 12\rangle\}$, through looking at the four-particle amplitude $A_4(1^{-1}, 2^{-1}, 3^{+1}, 4^{+1})$. The numerator must be $\langle 12\rangle^2[34]^2$. So, recognizing that $u = -s - t$ and $\tilde{u} = -\tilde{s} - \tilde{t}$, we see if we shift

$$s = \langle 21\rangle[12] \rightarrow \langle 21\rangle([12] + z\,\tilde{s}/\langle 21\rangle) = s + z\tilde{s} \tag{A.8.1}$$

$$t = \langle 41\rangle[14] \rightarrow (\langle 41\rangle + z\,\tilde{t}/[14])[14] = t + z\tilde{t} \tag{A.8.2}$$

$$u = -s - t \rightarrow u + z\,\tilde{u} = -(s+t) - z(\tilde{s} + \tilde{t})\,. \tag{A.8.3}$$

Deforming the anti-holomorphic part of s and the holomorphic part of t allows the z-shift to probe the s-, t-, and u-poles of the amplitude while leaving the numerator $\langle 12\rangle^2[34]^2$ unshifted.

This is the general case for amplitudes with higher-spin poles, i.e. for amplitudes with $3 \geq 2H + 1 - A \geq 2$ (the only cases amenable to this general analysis); we prove

this by contradiction. By virtue of having two or three poles in each term, we are guaranteed that the numerator does not have any complete factors of s, t, and/or u: if it did, then this would knock out one of the poles in a term, in violation of the assumption that $N_p = 2$ or 3.

Bibliography

[1] Stephen J. Parke and T. R. Taylor. An Amplitude for n Gluon Scattering. *Phys. Rev. Lett.*, 56:2459, 1986.

[2] Ruth Britto, Freddy Cachazo, Bo Feng, and Edward Witten. Direct proof of tree-level recursion relation in Yang-Mills theory. *Phys. Rev. Lett.*, 94:181602, 2005.

[3] James M. Drummond, Johannes M. Henn, and Jan Plefka. Yangian symmetry of scattering amplitudes in N=4 super Yang-Mills theory. *JHEP*, 05:046, 2009.

[4] Andrew Hodges. Eliminating spurious poles from gauge-theoretic amplitudes. *JHEP*, 05:135, 2013.

[5] Nima Arkani-Hamed and Jaroslav Trnka. The Amplituhedron. *JHEP*, 10:030, 2014.

[6] H. Kawai, D. C. Lewellen, and S. H. H. Tye. A Relation Between Tree Amplitudes of Closed and Open Strings. *Nucl. Phys.*, B269:1–23, 1986.

[7] Z. Bern, J. J. M. Carrasco, and Henrik Johansson. New Relations for Gauge-Theory Amplitudes. *Phys. Rev.*, D78:085011, 2008.

[8] Zvi Bern, Lance J. Dixon, and Radu Roiban. Is N = 8 supergravity ultraviolet finite? *Phys. Lett.*, B644:265–271, 2007.

[9] Steven Weinberg. Infrared photons and gravitons. *Phys. Rev.*, 140:B516–B524, 1965.

[10] Freddy Cachazo and Andrew Strominger. Evidence for a New Soft Graviton Theorem. 2014.

[11] Freddy Cachazo, Song He, and Ellis Ye Yuan. New Double Soft Emission Theorems. *Phys. Rev.*, D92(6):065030, 2015.

[12] Stephen W. Hawking, Malcolm J. Perry, and Andrew Strominger. Soft Hair on Black Holes. *Phys. Rev. Lett.*, 116(23):231301, 2016.

[13] Zvi Bern and Yu-tin Huang. Basics of Generalized Unitarity. *J. Phys.*, A44:454003, 2011.

[14] Freddy Cachazo, Song He, and Ellis Ye Yuan. Scattering of Massless Particles in Arbitrary Dimensions. *Phys. Rev. Lett.*, 113(17):171601, 2014.

[15] Edward Witten. Perturbative gauge theory as a string theory in twistor space. *Commun. Math. Phys.*, 252:189–258, 2004.

[16] Carlos R. Mafra, Oliver Schlotterer, Stephan Stieberger, and Dimitrios Tsimpis. A recursive method for SYM n-point tree amplitudes. *Phys. Rev.*, D83:126012, 2011.

[17] Stephen L. Adler. Consistency conditions on the strong interactions implied by a partially conserved axial vector current. *Phys. Rev.*, 137:B1022–B1033, 1965.

[18] Karol Kampf, Jiri Novotny, and Jaroslav Trnka. Tree-level Amplitudes in the Nonlinear Sigma Model. *JHEP*, 05:032, 2013.

[19] Clifford Cheung, Karol Kampf, Jiri Novotny, Chia-Hsien Shen, and Jaroslav Trnka. A Periodic Table of Effective Field Theories. *JHEP*, 02:020, 2017.

[20] Paolo Benincasa and Freddy Cachazo. Consistency Conditions on the S-Matrix of Massless Particles. 2007.

[21] Philip C. Schuster and Natalia Toro. Constructing the Tree-Level Yang-Mills S-Matrix Using Complex Factorization. *JHEP*, 06:079, 2009.

[22] Nima Arkani-Hamed, Laurentiu Rodina, and Jaroslav Trnka. Locality and Unitarity from Singularities and Gauge Invariance. 2016.

[23] Laurentiu Rodina. Uniqueness from gauge invariance and the Adler zero. 2016.

[24] Laurentiu Rodina. Uniqueness from locality and BCFW shifts. 2016.

[25] David A. McGady and Laurentiu Rodina. Higher-spin massless S-matrices in four-dimensions. *Phys. Rev.*, D90(8):084048, 2014.

[26] Marcus Spradlin, Anastasia Volovich, and Congkao Wen. Three Applications of a Bonus Relation for Gravity Amplitudes. *Phys. Lett.*, B674:69–72, 2009.

[27] Steven Weinberg. The Quantum theory of fields. Vol. 1: Foundations. 2005.

[28] Nima Arkani-Hamed and Jaroslav Trnka. The Amplituhedron. *JHEP*, 10:030, 2014.

[29] Luiz Antonio Barreiro and Ricardo Medina. RNS derivation of N-point disk amplitudes from the revisited S-matrix approach. *Nucl. Phys.*, B886:870–951, 2014.

[30] Rutger H. Boels and Ricardo Medina. Graviton and gluon scattering from first principles. *Phys. Rev. Lett.*, 118(6):061602, 2017.

[31] Steven Weinberg. Infrared photons and gravitons. *Phys. Rev.*, 140:B516–B524, 1965.

[32] Johannes Broedel, Marius de Leeuw, Jan Plefka, and Matteo Rosso. Constraining subleading soft gluon and graviton theorems. *Phys. Rev.*, D90(6):065024, 2014.

[33] Z. Bern, J. J. M. Carrasco, and Henrik Johansson. New Relations for Gauge-Theory Amplitudes. *Phys. Rev.*, D78:085011, 2008.

[34] Clifford Cheung, Karol Kampf, Jiri Novotny, and Jaroslav Trnka. Effective Field Theories from Soft Limits of Scattering Amplitudes. *Phys. Rev. Lett.*, 114(22):221602, 2015.

[35] Clifford Cheung, Karol Kampf, Jiri Novotny, Chia-Hsien Shen, and Jaroslav Trnka. On-Shell Recursion Relations for Effective Field Theories. *Phys. Rev. Lett.*, 116(4):041601, 2016.

[36] Clifford Cheung, Karol Kampf, Jiri Novotny, Chia-Hsien Shen, and Jaroslav Trnka. A Periodic Table of Effective Field Theories. *JHEP*, 02:020, 2017.

[37] Freddy Cachazo, Song He, and Ellis Ye Yuan. Scattering Equations and Matrices: From Einstein To Yang-Mills, DBI and NLSM. *JHEP*, 07:149, 2015.

[38] Karol Kampf, Jiri Novotny, and Jaroslav Trnka. Tree-level Amplitudes in the Nonlinear Sigma Model. *JHEP*, 05:032, 2013.

[39] Nima Arkani-Hamed and Jared Kaplan. On Tree Amplitudes in Gauge Theory and Gravity. *JHEP*, 04:076, 2008.

[40] Paolo Benincasa and Freddy Cachazo. Consistency Conditions on the S-Matrix of Massless Particles. 2007.

[41] Freddy Cachazo, Song He, and Ellis Ye Yuan. Scattering of Massless Particles in Arbitrary Dimensions. *Phys. Rev. Lett.*, 113(17):171601, 2014.

[42] Laurentiu Rodina. Uniqueness from locality and BCFW shifts. 2016.

[43] Freddy Cachazo, Song He, and Ellis Ye Yuan. Scattering of Massless Particles in Arbitrary Dimensions. *Phys. Rev. Lett.*, 113(17):171601, 2014.

[44] Freddy Cachazo, Song He, and Ellis Ye Yuan. Scattering Equations and Matrices: From Einstein To Yang-Mills, DBI and NLSM. *JHEP*, 07:149, 2015.

[45] Andrew Strominger. On BMS Invariance of Gravitational Scattering. *JHEP*, 07:152, 2014.

[46] Temple He, Vyacheslav Lysov, Prahar Mitra, and Andrew Strominger. BMS supertranslations and Weinberg's soft graviton theorem. *JHEP*, 05:151, 2015.

[47] Hui Luo and Congkao Wen. Recursion relations from soft theorems. *JHEP*, 03:088, 2016.

[48] Luiz Antonio Barreiro and Ricardo Medina. RNS derivation of N-point disk amplitudes from the revisited S-matrix approach. *Nucl. Phys.*, B886:870–951, 2014.

[49] Rutger H. Boels and Ricardo Medina. Graviton and gluon scattering from first principles. *Phys. Rev. Lett.*, 118(6):061602, 2017.

[50] Ruth Britto, Freddy Cachazo, and Bo Feng. New recursion relations for tree amplitudes of gluons. *Nucl. Phys.*, B715:499–522, 2005.

[51] Nima Arkani-Hamed and Jared Kaplan. On Tree Amplitudes in Gauge Theory and Gravity. *JHEP*, 04:076, 2008.

[52] Philip C. Schuster and Natalia Toro. Constructing the Tree-Level Yang-Mills S-Matrix Using Complex Factorization. *JHEP*, 06:079, 2009.

[53] Song He and Hong-bao Zhang. Consistency Conditions on S-Matrix of Spin 1 Massless Particles. *JHEP*, 07:015, 2010.

[54] Clifford Cheung, Karol Kampf, Jiri Novotny, Chia-Hsien Shen, and Jaroslav Trnka. On-Shell Recursion Relations for Effective Field Theories. *Phys. Rev. Lett.*, 116(4):041601, 2016.

[55] Steven Weinberg. Photons and Gravitons in s Matrix Theory: Derivation of Charge Conservation and Equality of Gravitational and Inertial Mass. *Phys. Rev.*, 135:B1049–B1056, 1964.

[56] Steven Weinberg. Infrared photons and gravitons. *Phys. Rev.*, 140:B516–B524, 1965.

[57] Marcus T. Grisaru and H. N. Pendleton. Soft Spin 3/2 Fermions Require Gravity and Supersymmetry. *Phys. Lett.*, B67:323–326, 1977.

[58] Christian Fronsdal. Massless Fields with Integer Spin. *Phys. Rev.*, D18:3624, 1978.

[59] A. Sagnotti and M. Tsulaia. On higher spins and the tensionless limit of string theory. *Nucl. Phys.*, B682:83–116, 2004.

[60] C. Aragone and Stanley Deser. Constraints on gravitationally coupled tensor fields. *Nuovo Cim.*, A3:709–720, 1971.

[61] Steven Weinberg and Edward Witten. Limits on Massless Particles. *Phys. Lett.*, B96:59–62, 1980.

[62] Juan Maldacena and Alexander Zhiboedov. Constraining Conformal Field Theories with A Higher Spin Symmetry. *J. Phys.*, A46:214011, 2013.

[63] Vasyl Alba and Kenan Diab. Constraining conformal field theories with a higher spin symmetry in d=4. 2013.

[64] Zvi Bern, Lance J. Dixon, and David A. Kosower. On-Shell Methods in Perturbative QCD. *Annals Phys.*, 322:1587–1634, 2007.

[65] Gerard 't Hooft, C. Itzykson, A. Jaffe, H. Lehmann, P. K. Mitter, I. M. Singer, and R. Stora. Recent Developments in Gauge Theories. Proceedings, Nato Advanced Study Institute, Cargese, France, August 26 - September 8, 1979. *NATO Sci. Ser. B*, 59:pp.1–438, 1980.

[66] Michael B. Green and John H. Schwarz. Anomaly Cancellation in Supersymmetric D=10 Gauge Theory and Superstring Theory. *Phys. Lett.*, B149:117–122, 1984.

[67] Yu-tin Huang and David McGady. Consistency Conditions for Gauge Theory S Matrices from Requirements of Generalized Unitarity. *Phys. Rev. Lett.*, 112(24):241601, 2014.

[68] Paolo Benincasa and Eduardo Conde. Exploring the S-Matrix of Massless Particles. *Phys. Rev.*, D86:025007, 2012.

[69] A. Fotopoulos and Mirian Tsulaia. On the Tensionless Limit of String theory, Off - Shell Higher Spin Interaction Vertices and BCFW Recursion Relations. *JHEP*, 11:086, 2010.

[70] Paul Dempster and Mirian Tsulaia. On the Structure of Quartic Vertices for Massless Higher Spin Fields on Minkowski Background. *Nucl. Phys.*, B865:353–375, 2012.

[71] Dario Francia, J. Mourad, and A. Sagnotti. Current Exchanges and Unconstrained Higher Spins. *Nucl. Phys.*, B773:203–237, 2007.

[72] Nima Arkani-Hamed, Jacob L. Bourjaily, Freddy Cachazo, Simon Caron-Huot, and Jaroslav Trnka. The All-Loop Integrand For Scattering Amplitudes in Planar N=4 SYM. *JHEP*, 01:041, 2011.

[73] Freddy Cachazo, Peter Svrcek, and Edward Witten. MHV vertices and tree amplitudes in gauge theory. *JHEP*, 09:006, 2004.

[74] Kasper Risager. A Direct proof of the CSW rules. *JHEP*, 12:003, 2005.

[75] Nima Arkani-Hamed. Pitp 2011 lectures. 2011.

[76] C. Aragone and Stanley Deser. Consistency Problems of Spin-2 Gravity Coupling. *Nuovo Cim.*, B57:33–49, 1980.

[77] Stanley Deser and B. Zumino. Consistent Supergravity. *Phys. Lett.*, 62B:335, 1976.

[78] Nima Arkani-Hamed, Freddy Cachazo, and Jared Kaplan. What is the Simplest Quantum Field Theory? *JHEP*, 09:016, 2010.

[79] Timothy Cohen, Henriette Elvang, and Michael Kiermaier. On-shell constructibility of tree amplitudes in general field theories. *JHEP*, 04:053, 2011.

[80] Paolo Benincasa, Camille Boucher-Veronneau, and Freddy Cachazo. Taming Tree Amplitudes In General Relativity. *JHEP*, 11:057, 2007.

[81] Karol Kampf, Jiri Novotny, and Jaroslav Trnka. Recursion relations for tree-level amplitudes in the $SU(N)$ nonlinear sigma model. *Phys. Rev.*, D87(8):081701, 2013.

[82] Karol Kampf, Jiri Novotny, and Jaroslav Trnka. Tree-level Amplitudes in the Nonlinear Sigma Model. *JHEP*, 05:032, 2013.

[83] Allan Adams, Nima Arkani-Hamed, Sergei Dubovsky, Alberto Nicolis, and Riccardo Rattazzi. Causality, analyticity and an IR obstruction to UV completion. *JHEP*, 10:014, 2006.

[84] Sidney R. Coleman and J. Mandula. All Possible Symmetries of the S Matrix. *Phys. Rev.*, 159:1251–1256, 1967.

[85] Rudolf Haag, Jan T. Lopuszanski, and Martin Sohnius. All Possible Generators of Supersymmetries of the s Matrix. *Nucl. Phys.*, B88:257, 1975.

[86] Z. Bern, J. J. Carrasco, Lance J. Dixon, H. Johansson, and R. Roiban. The Ultraviolet Behavior of N=8 Supergravity at Four Loops. *Phys. Rev. Lett.*, 103:081301, 2009.

[87] Zvi Bern, Tristan Dennen, Yu-tin Huang, and Michael Kiermaier. Gravity as the Square of Gauge Theory. *Phys. Rev.*, D82:065003, 2010.

[88] Niklas Beisert, Henriette Elvang, Daniel Z. Freedman, Michael Kiermaier, Alejandro Morales, and Stephan Stieberger. E7(7) constraints on counterterms in N=8 supergravity. *Phys. Lett.*, B694:265–271, 2011.

[89] Rutger H. Boels and Reinke Sven Isermann. On powercounting in perturbative quantum gravity theories through color-kinematic duality. *JHEP*, 06:017, 2013.

[90] Nima Arkani-Hamed and Jaroslav Trnka. The Amplituhedron. *JHEP*, 10:030, 2014.

[91] Paolo Benincasa and Freddy Cachazo. Consistency Conditions on the S-Matrix of Massless Particles. 2007.

[92] Rutger H. Boels and Reinke Sven Isermann. New relations for scattering amplitudes in Yang-Mills theory at loop level. *Phys. Rev.*, D85:021701, 2012.

[93] Bo Feng, Rijun Huang, and Yin Jia. Gauge Amplitude Identities by On-shell Recursion Relation in S-matrix Program. *Phys. Lett.*, B695:350–353, 2011.

[94] Z. Bern, J. J. Carrasco, D. Forde, H. Ita, and Henrik Johansson. Unexpected Cancellations in Gravity Theories. *Phys. Rev.*, D77:025010, 2008.

[95] Yu-tin Huang, David A. McGady, and Cheng Peng. One-loop renormalization and the S-matrix. *Phys. Rev.*, D87(8):085028, 2013.

[96] N. E. J. Bjerrum-Bohr, David C. Dunbar, Harald Ita, Warren B. Perkins, and Kasper Risager. The No-Triangle Hypothesis for N=8 Supergravity. *JHEP*, 12:072, 2006.

[97] Henriette Elvang, Yu-tin Huang, and Cheng Peng. On-shell superamplitudes in N¡4 SYM. *JHEP*, 09:031, 2011.

[98] Dongmin Gang, Yu-tin Huang, Eunkyung Koh, Sangmin Lee, and Arthur E. Lipstein. Tree-level Recursion Relation and Dual Superconformal Symmetry of the ABJM Theory. *JHEP*, 03:116, 2011.

[99] Simon Badger, N. E. J. Bjerrum-Bohr, and Pierre Vanhove. Simplicity in the Structure of QED and Gravity Amplitudes. *JHEP*, 02:038, 2009.

[100] Dung Nguyen, Marcus Spradlin, Anastasia Volovich, and Congkao Wen. The Tree Formula for MHV Graviton Amplitudes. *JHEP*, 07:045, 2010.

[101] David A. McGady. Soft-exactness of tree-level gluon and graviton MHV amplitudes. *Nucl. Phys. Proc. Suppl.*, 216:254–256, 2011.

[102] Andrew Hodges. A simple formula for gravitational MHV amplitudes. 2012.

[103] Andrew Hodges. New expressions for gravitational scattering amplitudes. *JHEP*, 07:075, 2013.

[104] Jin-Yu Liu and En Shih. Bonus scaling and BCFW in $\mathcal{N} = 7$ supergravity. *Phys. Lett.*, B740:151–157, 2015.