Self-healing unitarity is an optical illusion

Archit Vidyarthi[*]

Department of Physics, Indian Institute of Science Education and Research Bhopal,
Madhya Pradesh—462066, India

Among the vast variety of proposals put forward by the community to resolve tree-level unitarity violations in Higgs inflation models, there exists the concept of self-healing. It heals the theory from supposed tree-level violations for elastic scattering processes by summing over successive vacuum polarization loop corrections. In this work, we examine this technique to check whether unitarity is indeed restored and find that there exist underlying constraints in self-healing unitarity that pose the same perturbative unitarity bounds that it was expected to heal.

I. INTRODUCTION

Unitarity is one of several properties at the heart of a quantum theory, and essentially implies that the probability of an event cannot exceed unity. Along with other properties such as positivity, causality, etc., it helps provide us with useful bounds on a theory (for example: perturbative bounds, Froissart bounds, etc. [1]) in the form of constraints on a parameter, or on the domain within which the theory is valid, without needing to introduce new degrees of freedom (DsOF).

Tree-level unitarity violations, estimated using perturbative unitarity bounds, are immensely helpful in pointing out missing pieces in a theory. For a nonrenormalizable theory, these may imply that the loop corrections might become relevant as we approach the apparent violation scale in describing the complete process [1]. For others, they may indicate that the theory is incomplete. Beyond Standard Model (BSM) physics helps fill in gaps stemming from the incompatibility of the Standard Model and gravity, and provides us with possible candidates for the missing DsOF, often motivated by the existence of dark matter and dark energy that make up the majority of the energy content of the universe.

Given how Higgs driven inflation has been one of the prime candidates for an inflaton field (check [2–4] and references therein), the fact that it faces unitarity violations far below the Planck scale is something the scientific community has been trying to explain away for a long

time (see [4–9] and references therein for more info). After several decades of search, though, we have as of yet not been able to resolve the issue completely. Among the several approaches suggested toward resolving the issue is self-healing of unitarity proposed in [10] and later applied in the context of Higgs inflation in [6], which are at the heart of what we discuss in this work.

This paper is organized as follows: in Sec. II, we introduce the reader to the optical theorem and partial wave unitarity bounds as presented in [1]; in Sec. III, we briefly review the idea of self-healing as it was put forward in [10] while briefly introducing [6,11]; in Sec. IV we critically examine [6,10] and assess whether self-healing unitarity mechanism does play a role in the context of unitarizing nonminimally coupled scalar-tensor theories (STTs); and lastly, we conclude in Sec. V.

II. PERTURBATIVE UNITARITY BOUNDS

Imposing that the action is unitary, we obtain the famous optical theorem, which equates the imaginary part of the scattering amplitude to the total scattering cross section.

$$\mathcal{M}(i \to f) - \mathcal{M}^*(f \to i)$$
$$= i\sum_X \int d\Pi_X (2\pi)^4 \delta^4(p_i - p_X)\mathcal{M}(i \to X)\mathcal{M}^*(f \to X).$$
(1)

where $\mathcal{M}$ represents the scattering amplitude, $|iu\rangle$, $|f\rangle$, $|X\rangle$ are initial, final and arbitrary intermediate states, respectively, p_n represents the momentum of state $|n\rangle$, and $d\Pi_X$ is the momentum integral measure. In its generalized form (1), this theorem states that order-by-order in perturbation theory, imaginary parts of higher loop amplitudes are determined by lower loop amplitudes. For instance, the

[*]archit17@iiserb.ac.in

imaginary part of one-loop amplitude could be determined by the tree-level amplitude. A special case arises from this using the assumption that the initial and final states are the same (i.e., $|i\rangle = |f\rangle = |A\rangle$):

$$\text{Im}\mathcal{M}(A \to A) = 2E_{CM}|\vec{p}_i|\sum_X \sigma(A \to X). \quad (2)$$

where E_{CM} is the center of mass energy of the system and $\sigma(A \to X)$ is the scattering cross section for the enclosed process. Optical theorem puts a constraint on how large a scattering amplitude can be. From the approximate form,

$$\text{Im}\mathcal{M} \leq |\mathcal{M}|^2 \Rightarrow |\mathcal{M}| < 1. \quad (3)$$

Now, using the partial wave expansion of the scattering amplitude to impose constraints on coefficients of the Legendre polynomials. To recap, we first expand the scattering amplitude as:

$$\mathcal{M}(\theta) = 16\pi \sum_j a_j (2j+1) P_j(\cos\theta), \quad (4)$$

where a_j are complex-valued coefficients, and $P_j(\cos\theta)$ are Legendre polynomials with $P_j(1) = 1$ and

$$\int_{-1}^{1} P_j(\cos\theta) P_k(\cos\theta) d\cos\theta = \frac{2}{2j+1}\delta_{jk}. \quad (5)$$

For a case where the initial and final states are the same, we can write the total scattering cross section in the center of mass frame as:

$$\sigma_{CM_{tot}} = \frac{16\pi}{E_{CM}^2}\sum_j |a_j|^2 (2j+1). \quad (6)$$

Employing the optical theorem at $\theta = 0$, we have,

$$\text{Im}\mathcal{M}(AB \to AB \quad \text{at } \theta = 0)$$
$$= 2E_{CM}|\vec{p}_i|\sum_X \sigma_{tot}(AB \to X) \geq 2E_{CM}|\vec{p}_i|\sigma_{tot}(AB \to AB), \quad (7)$$

where an inequality has been introduced owing to the fact that $|AB\rangle \in |X\rangle$. Then,

$$\sum_{j=0}^{\infty}(2j+1)\text{Im}(a_j) \geq \frac{2|\vec{p}_i|}{E_{CM}}\sum_{j=0}^{\infty}(2j+1)|a_j|^2. \quad (8)$$

This, coupled with the inequality $|a_j| \geq \text{Im}(a_j)$, means that the magnitude of a_j is now constrained as $|a_j| \leq 1$, $0 \leq \text{Im}(a_j) \leq 1$, and $|\text{Re}(a_j)| \leq 1/2$. These three conditions constitute the perturbative unitarity bounds and can be found from the Argand plane in Fig. 1. The information
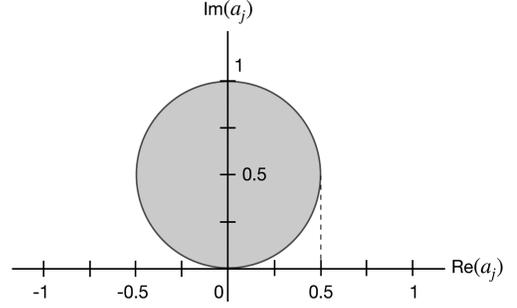

FIG. 1. Argand diagram representing the condition $|a_j|^2 \leq \text{Im}(a_j)$. The area within the circumference represents the acceptable region where unitarity is obeyed.

presented in this section is explained in much greater detail in [1].

III. SELF-HEALING UNITARITY

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{16\pi G_N}R + \partial^\mu\Phi^\dagger\partial_\mu\Phi - m^2|\Phi|^2 + \lambda R|\Phi|^2\right] \quad (9)$$

Preceding [10], authors of [11] worked with a set of complex scalar fields nonminimally coupled with gravity as in Eq. (9), and tried to estimate the scattering amplitude for the process $s\bar{s} \to s'\bar{s}'$, where they set $s \neq s'$ to make sure that only the s-channel graviton exchange diagram contributed to the process, and they could avoid collinear divergences in the t and u channels. They claimed that in the limit where the number of particles is large, the leading order loop corrections are successive vacuum polarization diagrams and that any tree-level violations could be fixed by considering such higher-loop corrections.

Using partial wave expansion as in the previous section, they estimated the scale of unitarity violations for $j = 2$:

$$s = \frac{20}{G_N N}, \quad (10)$$

where s is a Mandelstam variable, G_N is the Newton's constant and N is the number of particles in the theory. They obtained the corresponding scales for the standard model of particle physics ($\sqrt{s} = 6 \times 10^{18}$ GeV) and the minimal supersymmetric standard model ($\sqrt{s} = 4.6 \times 10^{18}$ GeV), both coupled with gravity.

Following this, authors of [10] considered a similar Lagrangian as [11] involving a nonminimal coupling between gravity and multiple scalar fields and provide a useful confirmation for the results presented in [11]. They go a step further and claim that the perturbative unitarity bound (10) is false, and that summing over the infinite loop

contributions in the same manner as done in [11], they could use the relation:

$$a_j = \frac{a_j^{\text{tree}}}{1 - \frac{\text{Re}(a_j^{\text{1-loop}})}{a_j^{\text{tree}}} - ia_j^{\text{tree}}} \quad (11)$$

to show that there are no unitarity violations. Authors of [6] expanded on this work and verified the results for a theory involving the Higgs' doublet They expanded the Higgs' boson around a large background which caused the coupling constants of the Higgs' and Goldstone bosons to differ. They, then, proceeded to show that self-healing phenomenon could be applied to $j = 0$ level as well.

There are two equivalent approaches to observe self-healing of unitarity and we shall examine them both using results from [6,10] in the following section. It should be noted that even though the primary analyses have been performed in the Jordan frame, an equivalent self-healing mechanism can be found in the Einstein frame as well [12].

IV. UNITARITY VIOLATIONS IN NONMINIMALLY COUPLED SCALAR-TENSOR THEORIES

For this section, we consider the action:

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{16\pi G_N} R + \partial^\mu \phi \partial_\mu \phi + \lambda R \phi^2 \right] \quad (12)$$

as used in [10] where we consider N_s scalar fields and work in the large N_s limit. Considering tree-level amplitudes, for $\lambda = 0$ (minimal coupling) and $\lambda = -1/3$ (conformal coupling), the theory is well behaved up to the Planck scale, but more generally, looking at this action from dimensional grounds, we expect the perturbative unitarity scales for all processes $\phi_A \phi_A \rightarrow \phi_B \phi_B$ to be $\approx (G_N \lambda)^{-1}$. Now, for $\lambda > 1$, it implies that perturbative unitarity is being violated below $\approx (G_N)^{-1}$. Authors in [11] found that the perturbative unitarity bound in these theories was dependent on N_s, as stated in (10). The works [6,10] suggest, however, that the self-healing mechanism takes care of any supposed violations through infinite summation of vacuum polarization corrections.

A. Partial wave amplitude approach

The tree-level $\phi_A \phi_A \rightarrow \phi_B \phi_B$ scattering amplitude is,

$$\mathcal{M}^{\text{tree}} = \frac{8\pi G_N}{s} [2s^2 \lambda (3\lambda + 1) + ut] \quad (13)$$

and corresponding partial wave amplitudes for all possible combinations of scalars can then be easily found to be,

$$a_0^{\text{tree}} = \frac{G_N N_s s}{48} [24\lambda(3\lambda + 1) + 1], \quad (14)$$

$$a_2^{\text{tree}} = -\frac{G_N N_s s}{120}. \quad (15)$$

Similarly, at 1-loop level,

$$\mathcal{M}^{\text{1-loop}} = \frac{G_N^2 N_s}{15} [s^2 F(\lambda) - ut] \quad (16)$$

$$a_0^{\text{1-loop}} = \frac{-G_N^2 N_s^2 s^2}{2880\pi} [6F(\lambda) - 1] \ln(-s), \quad (17)$$

$$a_2^{\text{1-loop}} = -\frac{-G_N^2 N_s^2 s^2}{14400\pi} \ln(-s), \quad (18)$$

where $F(\lambda) = 1 + 20\lambda + 180\lambda^2 + 720\lambda^3 + 1080\lambda^4$ and s, t, u represent the Mandelstam variables. The appearance of N_s in the expressions (14), (15), (17), and (18) can be explained simply: the authors in [11] worked with normalized two-particle states, such that the normalization factor for each state was $1/\sqrt{N_s}$. Combined with the combinatorial factor $N_s^2/2$ for large N_s, we are left with $N_s/2$ as the factor. For 1-loop results, we need to attach another N_s factor for the scalars running in the loop. Note that $\lambda = 0$ returns the expressions to the minimal coupling case, and $\lambda = -\frac{1}{3}$ gives us the conformal coupling case. For both of these, unitarity is safe up to the Planck scale. Since a_2 are completely independent of λ, the interesting case is clearly a_0.

The authors in [10] only tackled the $j = 2$ case and proved through $|a_2^{\text{tree}}|^2 = \text{Im}(a_2^{\text{1-loop}})$ that the unitarity violations at tree-level can be taken care of by 1-loop corrections. It is clear from the expressions above, however, that since $j = 2$ does not contain any λ dependence, their results only prove that the minimal theory which appeared to have tree-level unitarity violations at the Planck scale is completely healed when considering the 1-loop corrections. This result was improved in [6], where the authors extended the result to the more relevant $j = 0$ case. Their result holds even when we consider the coupling constants to be the same, as in (12).

The primary claim of [10] is that owing to the condition $|a_2^{\text{tree}}|^2 = \text{Im}(a_2^{\text{1-loop}})$, we can write:

$$a_j = \frac{a_j^{\text{tree}}}{1 + \frac{\text{Re}(a_j^{\text{1-loop}})}{a_j^{\text{tree}}} - ia_j^{\text{tree}}}. \quad (19)$$

The authors state the same for (12) as well, which is corroborated and extended to $j = 0$ in [6]. They imply that perturbative unitarity violations have been dealt with completely and the dependence of violation scale on N_s in [11] is just a misinterpretation of the result. Now, as stated in Sec. II, the consequences of $|a_2^{\text{tree}}|^2 = \text{Im}(a_2^{\text{1-loop}})$ are the constraints: $|a_j| \leq 1$, $0 \leq \text{Im}(a_j) \leq 1$, and $|\text{Re}(a_j)| \leq 1/2$. Considering

the first constraint at tree-level (see the appendix) for the partial wave amplitudes, we find,

$$|a_2^{\text{tree}}| \leq 1 \Rightarrow s \leq \frac{120}{G_N N_s}, \tag{20}$$

$$|a_0^{\text{tree}}| \leq 1 \Rightarrow s \leq \frac{48}{G_N N_s [24\lambda(3\lambda+1)+1]}, \tag{21}$$

i.e., there still exists a perturbative unitarity bound on the theory that depends on N_s and λ, similar to the result of [11] (up to a multiplicative factor). Therefore, we claim that the condition (11) as proposed in [10] does not imply healed tree-level perturbative unitarity violations considering all loop levels, even though they appear to be healed when we consider 1-loop corrections. This can also be seen when we perform a summation over infinite vacuum polarization corrections. In order for the geometric progression to be convergent, we require that the common ratio be <1. This translates in our present case to the condition:

$$\left|\frac{a_2^{\text{1-loop}}}{a_2^{\text{tree}}}\right| = \frac{G_N N_s s}{120\pi} < 1 \Rightarrow s < \frac{120\pi}{G_N N_s} \tag{22}$$

$$\left|\frac{a_0^{\text{1-loop}}}{a_0^{\text{tree}}}\right| = \frac{G_N N_s s}{60\pi}\left[\frac{6F(\lambda)-1}{24\lambda(3\lambda+1)+1}\right] < 1$$

$$\Rightarrow s < \frac{60\pi}{G_N N_s}\left[\frac{24\lambda(3\lambda+1)+1}{6F(\lambda)-1}\right] \tag{23}$$

where we have ignored $\ln(-s)$ contributions since we're working in the UV limit. We see a similar dependence of the perturbative unitarity bound on N_s and λ as seen earlier in (20) and (21).

B. Dressed propagator approach

Further, the authors in [10] verify their results using the dressed propagator approach for $j = 2$, extended to $j = 0$ in [6]. Here we present the 1-loop corrected dressed graviton propagator as follows,

$$i\mathcal{D}^{\alpha\beta,\mu\nu} = \frac{i}{2q^2}(1+2B(q^2))[L^{\alpha\mu}L^{\beta\nu}+L^{\alpha\nu}L^{\beta\mu}-L^{\alpha\beta}L^{\mu\nu}]$$

$$-i\frac{A(q^2)}{4}L^{\alpha\beta}L^{\mu\nu}. \tag{24}$$

where

$$A(q^2) = -\frac{1}{30\pi}G_N N_s(1+10\lambda+30\lambda^2)\ln\left(\frac{-q^2}{\mu^2}\right), \tag{25}$$

$$B(q^2) = \frac{1}{240\pi}G_N N_s \ln\left(\frac{-q^2}{\mu^2}\right), \tag{26}$$

$$L^{\mu\nu} = \eta^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}, \tag{27}$$

where q represents momentum and μ is related to the renormalization scale of the theory. Base graviton propagator can be recovered by setting $A(q^2)$, $B(q^2) = 0$. Again, setting $\lambda = 0, -\frac{1}{3}$ returns the $j = 2$ dressed graviton propagator, and the λ dependent terms correspond to the off-shell $j = 0$ part. Authors in [10] mistakenly assume that $A(q^2)$ goes to zero completely in the aforementioned limits and due to this, their result for the infinite 1-loop summed dressed propagator is incorrect.

Also, in order to proceed with the summation, they assume that $G_N N_s$ is small. Since this is a dimensionful quantity, a more complete statement would be $G_N N_s \ll s^{-1}$. This again gives us a similar upper bound on energy as (10), (20), meaning its dependence on N_s is still present. Further, they only proceed with the $j = 2$ part for the rest of the calculation, which as stated earlier, does not hold any information about the violation scales dependent on λ.

The authors of [6] come to the rescue here. They assume $\lambda_1 = \lambda_2 = \lambda$ (i.e., small Higgs' background limit) and focus on the $j = 0$ part of the dressed propagator (24) by assuming $\lambda \gg 1$, which is a valid limit as mentioned earlier in this section. This limit is suggested by the authors to avoid any contributions from $B(q^2)$, and consequently also ignore parts of the base propagator contribution. The dressed propagator in this work looks like,

$$i\mathcal{D}^{\alpha\beta\mu\nu} \approx -\frac{i}{2q^2}\left(1+\frac{q^2 A(q^2)}{2}\right)L^{\alpha\beta}L^{\mu\nu}$$

$$\approx -\frac{i}{2q^2}\left[1-\frac{q^2 G_N N_s \lambda^2}{2\pi}\ln\left(\frac{-q^2}{\mu^2}\right)\right]L^{\alpha\beta}L^{\mu\nu}. \tag{28}$$

Typographical errors aside, these are the results of [6]. Later, similar to [10], they assume $\lambda^2 G_N N_s \ll q^{-2}$ to be able to sum over the infinite series [which again reinforces the dependence of the perturbative unitarity bound on N_s and λ as in (21) and (23)]. This, however, does not make sense because even though taking the two aforementioned limits simultaneously means that we can ignore $B(q^2)$ (since $\lambda \gg 1$) in favor of $A(q^2)$ in (24), all parts of the base graviton propagator must contribute to the dressed result since the constraint $q^2 G_N N_s \lambda^2 \ll 1$ implies that $A(q^2)$ is the leading perturbative correction. The actual form of the propagator, as per their assumptions, should look like,

$$i\mathcal{D}^{\alpha\beta,\mu\nu} \approx \frac{i}{2q^2}[L^{\alpha\mu}L^{\beta\nu}+L^{\alpha\nu}L^{\beta\mu}-L^{\alpha\beta}L^{\mu\nu}]$$

$$-i\frac{G_N N_s \lambda^2}{4\pi}\ln\left(\frac{-q^2}{\mu^2}\right)L^{\alpha\beta}L^{\mu\nu}, \tag{29}$$

for which summing the infinite series is a rather difficult task. Therefore, we need to confine ourselves with the 1-loop result in this method as well. It can be verified that the partial wave amplitude for the sum of $\phi_A\phi_A \to \phi_B\phi_B$ type processes involving the 1-loop dressed graviton

propagator (29) for all A, B is the same as that obtained in (17) assuming $\lambda \gg 1$ (up to a multiplicative factor), i.e. $|a_0^{\text{tree}}|^2 = \text{Im}(a_0^{\text{1-loop}})$ as verified using the partial wave amplitude approach in [6].

V. DISCUSSION

The self-healing mechanism was defined in [10] to operate under specific conditions that were first discovered in [11] and listed in Sec. III. After examining the claims made in the paper in both partial wave and dressed propagator approaches, we conclude through this work that the assessment made by the authors of [11] that the unitarity violation scale depends on the number of particles N_s is indeed true and complete healing of tree-level violations works only if the bounds described in (20) and (21) are obeyed strictly.

In conclusion, we found that tree-level unitarity violations are indeed healed using 1-loop corrections, but the conditions required to effectively apply the self-healing mechanism themselves impose bounds on the energy scale of the theory that are dependent on the number of particles in the theory, as found by the authors of [11] previously.

APPENDIX

The authors in [10] claim that for a unitary theory obeying $\text{Im}(a_j^{\text{1-loop}}) = |a_j^{\text{tree}}|^2$, we can write,

$$a_j = a_j^{\text{tree}} + a_j^{\text{1-loop}} = \frac{a_j^{\text{tree}}}{1 - \frac{\text{Re}(a_j^{\text{1-loop}})}{a_j^{\text{tree}}} - i a_j^{\text{tree}}}. \quad \text{(A1)}$$

Expanding $a_j^{\text{1-loop}} = \text{Re}(a_j^{\text{1-loop}}) + i\text{Im}(a_j^{\text{1-loop}})$, we find that the equation above holds true if and only if,

$$\text{Re}(a_j^{\text{1-loop}})^2 + 2a_j^{\text{tree}}\text{Re}(a_j^{\text{1-loop}}) + (a_j^{\text{tree}})^4 = 0 \quad \text{(A2)}$$

Real solutions exist for $\text{Re}(a_j^{\text{1-loop}})$ only when $|a_j^{\text{tree}}| \leq 1$.

[1] Matthew D. Schwartz, *Quantum Field Theory and the Standard Model* (Cambridge University Press, Cambridge, England, 2014), ISBN 978-1-107-03473-0, 978-1-107-03473-0.

[2] Michael Atkins and Xavier Calmet, Remarks on Higgs inflation, Phys. Lett. B **697**, 37 (2011).

[3] Rose N. Lerner and John McDonald, Higgs inflation and naturalness, J. Cosmol. Astropart. Phys. 04 (2010) 015.

[4] Javier Rubio, Higgs inflation, Front. Astron. Space Sci. **5**, 50 (2019).

[5] Sukanta Panda, Abbas Altafhussain Tinwala, and Archit Vidyarthi, Ultraviolet unitarity violations in non-minimally coupled scalar-Starobinsky inflation, J. Cosmol. Astropart. Phys. 01 (2023) 029.

[6] Xavier Calmet and Roberto Casadio, Self-healing of unitarity in Higgs inflation, Phys. Lett. B **734**, 17 (2014).

[7] Rose N. Lerner and John McDonald, A unitarity-conserving Higgs inflation model, Phys. Rev. D **82**, 103525 (2010).

[8] Albert Escrivà and Cristiano Germani, Beyond dimensional analysis: Higgs and new Higgs inflations do not violate unitarity, Phys. Rev. D **95**, 123526 (2017).

[9] Ignatios Antoniadis, Anthony Guillen, and Kyriakos Tamvakis, Ultraviolet behaviour of Higgs inflation models, J. High Energy Phys. 08 (2021) 018; J. High Energy Phys. 05 (2022) 74.

[10] Ufuk Aydemir, Mohamed M. Anber, and John F. Donoghue, Self-healing of unitarity in effective field theories and the onset of new physics, Phys. Rev. D **86**, 014025 (2012).

[11] Tao Han and Scott Willenbrock, Scale of quantum gravity, Phys. Lett. B **616**, 215 (2005).

[12] Yohei Ema, Dynamical emergence of scalaron in Higgs inflation, J. Cosmol. Astropart. Phys. 09 (2019) 027.