PHYSICAL REVIEW D **92,** 083013 (2015)

Conditionally extended validity of perturbation theory: Persistence of AdS stability islands

Fotios Dimitrakopoulos and I-Sheng Yang

IOP and GRAPPA, Universiteit van Amsterdam, Science Park 904, 1090 GL Amsterdam, Netherlands

(Received 20 July 2015; published 22 October 2015)

Approximating nonlinear dynamics with a truncated perturbative expansion may be accurate for a while, but it, in general, breaks down at a long time scale that is one over the small expansion parameter. There are interesting cases in which such breakdown does not happen. We provide a mathematically general and precise definition of those cases, in which we prove that the validity of truncated theory trivially extends to the long time scale. This enables us to utilize numerical results, which are only obtainable within finite times, to legitimately predict the dynamics when the expansion parameter goes to zero, and thus the long time scale goes to infinity. In particular, this shows that existing noncollapsing solutions in the AdS (in)stability problem persist to the zero-amplitude limit, opposing the conjecture by Dias, Horowitz, Marolf, and Santos that predicts a shrinkage to measure zero [O. J. Dias *et al.*, Classical Quantum Gravity 29, 235019 (2012)]. We also point out why the persistence of collapsing solutions is harder to prove, and how the recent interesting progress by Bizon, Maliborski, and Rostoworowski has not yet proven this [P. Bizon, M. Maliborski, and A. Rostworowski, arXiv:1506.03519].

DOI: 10.1103/PhysRevD.92.083013 PACS numbers: 95.30.Sf

I. INTRODUCTION AND SUMMARY

A. Truncated perturbative expansion

A linear equation of motion $D\phi = 0$ often has close-form analytical solutions. A nonlinear equation, $D\phi = F_{\mathrm{nonlinear}}(\phi)$, on the other hand, usually does not. One can attempt to expand $F_{\mathrm{nonlinear}}$ when ϕ is small. For example,

$$D\phi = F_{\mathrm{nonlinear}}(\phi) = \phi^2 + \mathcal{O}(\phi^3). \quad (1.1)$$

When the amplitude is small, $|\phi| < \epsilon \ll 1$, one can solve the truncated equation of motion that includes the ϕ^2 term as a perturbative expansion of $|\phi^2/\phi| < \epsilon$ from the linear solutions. For a small enough choice of ϵ, this can be a good enough approximation to the fully nonlinear theory. Unfortunately, this will only work for a short amount of time. After some time $T \sim \epsilon^{-1}$, the correction from the first nonlinear order accumulates and becomes comparable to the original amplitude. Thus, the actual amplitude can exceed ϵ significantly to invalidate the expansion.[1]

A slightly more subtle question arises while applying such a truncated perturbation theory. Occasionally, there can be accidental cancellations while solving it. Thus, during the process, the amplitude may stay below ϵ for $T \sim \epsilon^{-1}$. Are we then able to trust these solutions?

It is very tempting to directly answer "no" to the above question. When $T \sim \epsilon^{-1}$, not only does the accumulated contribution from ϕ^2, which the theory does take into account, modify ϕ significantly, but the ϕ^3 term that the theory discarded also modifies ϕ^2, and so on. Since we have truncated all those even higher order terms that may have significant effects, the validity of the expansion process seems to unsalvageably break down.

The above logic sounds reasonable but not entirely correct. In this paper, we demonstrate that at exactly the $T \sim \epsilon^{-1}$ time scale, the opposite is true. These "nice" solutions we occasionally find in the truncated theory indeed faithfully represent similar solutions in the full nonlinear theory. This idea is not entirely new. We are certainly inspired by the application of the two-time formalism and the renormalization flow technique in the AdS-(in)stability problem, and both of them operate under this same concept [1,2].[2] However, one may get the impression from those examples that additional techniques are required to maintain the approximation over the long time scale. One main point of this paper is to establish that the validity of truncated theory extends *trivially* in those cases. As long as the truncated theory is implemented recursively, which is the natural way to solve any time evolution anyway, it remains trustworthy in those cases.[3]

In Sec. II, we state and prove a theorem that guarantees a truncated perturbative expansion, implemented recursively,

[1]The notion of "time" here is just to make connections to practical problems for physicists. The general idea is valid whenever one tries to solve perturbation theory from some limited boundary conditions to a far-extended domain.

[2]We thank Luis Lehner for pointing out that some post-Newtonian expansions to general relativity also show validity at this long time scale [3].

[3]It is extremely likely that a capable mathematician can directly point to a textbook material to back up this claim. However, such a reference is difficult for us to find and may not be very transparent to physicists. Furthermore, the actual mathematical proofs are quite simple, so we will simply construct and present them in this paper. Any suggestion to include a mathematical reference is welcomed.

to approximate the full nonlinear theory accurately in the long time scale in the relevant cases. More concretely, this theorem leads to the following two facts:

(i) If one solves the truncated theory and finds solutions in which the amplitude remains small during the long time scale, then *similar* solutions exist in the full nonlinear theory.[4]

(ii) If numerical evolution of the full nonlinear theory provides solutions in which the amplitude remains small during the long time scale, then a truncated theory can reproduce *similar* solutions.

Formally, the meaning of "similar" in the above statements means that the difference between two solutions goes to zero faster than their amplitudes in the zero amplitude limit. This theorem provides a two-way bridge between numerical and analytical works. Anything of this nature can be quite useful. For example, numerical results are usually limited to finite amplitudes and times, while the actual physical questions might involve taking the limit of zero amplitude and infinite time. With this theorem, we can start from known numerical results and extend them to the limiting case with analytical techniques.

In Sec. III, we prove another theorem that enables us to do just that in the AdS (in)stability problem. The key is that the truncated theory does not need to be exactly solved to be useful. Since its form is simpler than the full nonlinear theory, it can manifest useful properties, such as symmetries, to facilitate further analysis. Since it is a truncated theory, the symmetry might be an approximation itself, and it might be naïvely expected to break down at the long time scale. Not surprisingly, using a similar process, we can again prove that such symmetry remains trustworthy in the relevant cases.

It is interesting to note that the conventional wisdom, which suggested an unsalvageable breakdown at $T \sim \epsilon^{-1}$, is not entirely without merits. We can prove that both theorems hold for $T = \alpha\epsilon^{-1}$ an arbitrarily but ϵ-independent value of α. However, pushing it further to a slightly longer, ϵ-dependent time scale, for example, $T \sim \epsilon^{-1.1}$, the proofs immediately become obsolete. The situation for $T \sim \epsilon^{-1}(-\ln\epsilon)$ is also delicate and will not always hold. Naturally, for time scales *longer* than $T \sim \epsilon^{-1}$, one needs to truncate the theory at an even higher order to maintain its validity. A truncated theory up to the ϕ^m term will only be valid up to $T \sim \epsilon^{1-m}$.

B. AdS (in)stability problem

In Sec. IV, we apply both theorems to the AdS-(in)stability problem [1,2,4–20]. Currently, the main focus of this problem is indeed the consequence of the nonlinear dynamics of gravitational self-interaction, at the time scale that the leading-order expansion should generically break down. Some have tried to connect such a breakdown to the formation of black holes and further advocate that such instability of AdS space is generic. In particular, Dias, Horowitz, Marolf, and Santos made the stability island conjecture [4]. Although at finite amplitudes, there are numerical evidence and analytical arguments to support measure-nonzero sets of noncollapsing solutions, they claimed that the sets of these solutions shrink to measure zero at the zero amplitude limit.

Since the relevant time scale goes to infinity at the zero amplitude limit, such a conjecture cannot be directly tested by numerical efforts. Nevertheless, by the two theorems we prove in this paper, it becomes straightforward to show that such a conjecture is in conflict with existing evidence. The physical intuition of our argument was already outlined in Ref. [18], and here we establish the rigorous mathematical structure behind it.

(i) Theorem I allows us to connect noncollapsing solutions [1,4,11,13] to analogous solutions in a truncated theory, both at finite amplitudes.

(ii) Theorem II allows us to invoke the rescaling symmetry in the truncated theory and establish those solutions at arbitrarily smaller amplitudes.

(iii) Using Theorem I again, we can establish those noncollapsing solutions in the full nonlinear theory at arbitrarily smaller amplitudes.

Thus, if noncollapsing solutions form a set of measure nonzero at finite amplitudes as current evidence implies, then they persist to be a set of measure nonzero when the amplitude approaches zero. Since the stability island conjecture states that stable solutions should shrink to sets of measure zero, it is in conflict with existing evidence.

It is important to note that defeating the stability island conjecture is not the end of the AdS (in)stability problem. Another important question is whether collapsing solutions, which likely also form a set of measure nonzero at finite amplitudes, also persist down to the zero-amplitude limit. It is easy to see why that question is harder to answer. Truncated expansions of gravitational self-interaction, at least all those that have been applied to the problem, do break down at a certain point during black hole formation. Thus, theorem I does not apply, one cannot establish a solid link between the truncated dynamics to the fully nonlinear one, and the AdS (in)stability problem remains unanswered.

In order to make an equally rigorous statement about collapsing solutions, one will first need to pose a weaker claim. Instead of arguing for the generality of black hole formation, one should be content with "energy density exceeding certain bounds" or something similar. This type of claim is then more suited to be studied within the validity of theorem I, and it is also a reasonable definition of AdS instability. If arbitrarily small initial energy always

[4]Note that sometimes, especially in gauge theories, the full nonlinear theory might impose a stronger constraint on acceptable initial conditions. One should start from those acceptable initial conditions in order to apply our theorem. We thank Ben Freivogel for pointing this out.

evolves to have finite energy density somewhere, it is a clear sign of a runaway behavior due to gravitational attraction.[5]

Finally, we should note that the truncated theory is already nonlinear and may be difficult to solve directly. If one invokes another approximation while solving the truncated theory, such as time averaging, then the process becomes vulnerable to an additional form of breakdown, such as the oscillating singularity seen in [22]. Even if numerical observations in some cases demonstrate a coincidence between such breakdown and black hole formation, the link between them is not yet as rigorous as the standard established in this paper for noncollapsing solutions.

II. THEOREM I: CONDITIONALLY EXTENDED VALIDITY

Consider a linear space $\mathcal{H}$ with a norm satisfying triangular inequality,

$$\|x+y\| \le \|x\| + \|y\|, \quad \text{for all } x, y \in \mathcal{H}. \tag{2.1}$$

Then consider three smooth functions L, f, g all from $\mathcal{H}$ to itself. We require that $L(x) = 0$ if and only if $x = 0$, and it is "semi-length-preserving,"

$$\|L(x)\| \le \|x\|. \tag{2.2}$$

Note that this condition on the length is at no cost to generality. Given any smooth function $\bar{L}$ meeting the first requirement, we can always rescale it so that it is exactly length preserving and that it maintains its smoothness,

$$L(x) \equiv \frac{\|x\|}{\|\bar{L}(x)\|}\bar{L}(x), \quad \text{if } x \neq 0;$$
$$L(x) \equiv 0, \quad \text{when } x = 0. \tag{2.3}$$

Within some radius $r < 1$, f and g are two functions that are both close to L but even closer to each other.

(1) Close to L: $\forall \|x\| < r$,

$$\|f(x) - L(x)\| < a\|x\|^m,$$
$$\|g(x) - L(x)\| < a\|x\|^m, \quad \text{for some } a > 0, \quad m > 1. \tag{2.4}$$

Doing so smoothly: $\forall \|x\|$, $\|y\| < r$ and some $b > 0$,

$$\|[f(x) - L(x)] - [f(y) - L(y)]\|$$
$$< b\|x-y\|\mathrm{Max}(\|x\|, \|y\|)^{m-1},$$
$$\|[g(x) - L(x)] - [g(y) - L(y)]\|$$
$$< b\|x-y\|\mathrm{Max}(\|x\|, \|y\|)^{m-1}. \tag{2.5}$$

(2) Even closer to each other: $\forall \|x\| < r$,

$$\|f(x) - g(x)\| < c\|x\|^l, \quad \text{for some } c > 0, \quad l > m. \tag{2.6}$$

Since this is a physics paper, we make the analogy to the physical problem more transparent by using an example. Choose a finite time Δt to evolve the linear equation of motion $D\phi = 0$; L is given by $L[\phi(t)] = \phi(t + \Delta t)$. Similarly, evolving the full nonlinear theory $D\phi = F_{\mathrm{nonlinear}}(\phi)$ leads to a different solution ϕ that defines $f[\phi(t)] = \phi(t + \Delta t)$, and $D\phi = \phi^2$ defines $g[\phi(t)] = \phi(t + \Delta t)$. Furthermore, the norm can often be defined as the square root of conserved energy in the linear evolution, which satisfies both the triangular inequality and the semipreserving requirement.

From this analogy, evolution to a longer time scale is naturally given by applying these functions recursively. We thus define three sequences accordingly.

$$f_0 = g_0 = L_0 = x, \qquad L_n = L(L_{n-1}),$$
$$f_n = f(f_{n-1}), \qquad g_n = g(g_{n-1}). \tag{2.7}$$

We prove a theorem which guarantees that after a time long enough for both g_n and f_n to deviate significantly from L_n, they can still stay close to each other.

Theorem I: For any finite $\delta > 0$ and $\alpha > 0$, there exists $0 < \epsilon < r$ such that if $\|f_n\| < \epsilon$ for all $0 \le n < \alpha\epsilon^{1-m}$, then $\|f_n - g_n\| < \delta\epsilon$.

Since f_n is known to be of order ϵ, when its difference from g_n is arbitrarily smaller than ϵ, one remains a good approximation of the other.

Proof: First, we define

$$\Delta_n \equiv c\epsilon^l \sum_{i=0}^{n-1} (1 + b\epsilon^{m-1})^i = c\epsilon^l \frac{(1 + b\epsilon^{m-1})^n - 1}{b\epsilon^{m-1}}. \tag{2.8}$$

Within the range of n stated in theorem I, it is easy to see that

$$\Delta_n \le \Delta_{\lfloor \alpha\epsilon^{1-m} \rfloor} < c\epsilon^l \frac{(1 + b\epsilon^{m-1})^{\alpha\epsilon^{1-m}} - 1}{b\epsilon^{m-1}}$$
$$< \frac{c\epsilon^{1+l-m}}{b}(e^{b\alpha} - 1). \tag{2.9}$$

Since $l > m$, there is always a choice of ϵ such that $\Delta_n < \delta\epsilon$. We choose an ϵ small enough for that case, and also small enough such that

[5]It is then natural to believe that black hole formation follows, though it is still not guaranteed and it is hard to prove. For example, a Gauss-Bonnet theory can behave the same up to this point, but its mass gap forbids black hole formation afterward [21].

$$\|f_n\| + a\epsilon^m < \epsilon + a\epsilon^m < r, \tag{2.10}$$

$$\|f_n\| + \Delta_n < \epsilon + \delta\epsilon < r. \tag{2.11}$$

Next, we use mathematical induction to prove that given such choice of ϵ,

$$\|f_n - g_n\| \le \Delta_n. \tag{2.12}$$

For $n = 0$, this is trivially true,

$$\|f_0 - g_0\| = 0 \le \Delta_0 = 0. \tag{2.13}$$

Assume this is true for $(n-1)$,

$$\|f_{n-1} - g_{n-1}\| \le \Delta_{n-1}. \tag{2.14}$$

Combining it with Eqs. (2.6) and (2.5), we can derive the next term in the sequence,

$$\|f_n - g_n\| \le \|f(f_{n-1}) - g(f_{n-1})\| + \|g(f_{n-1}) - g(g_{n-1})\| \le c\epsilon^l + (1 + b\epsilon^{m-1})\Delta_{n-1} = \Delta_n. \tag{2.15}$$

Thus, by mathematical induction, we have proven the theorem.

Note that although in the early example for physical intuitions we took f as the full nonlinear theory and g as the truncated theory, their roles are actually interchangeable in theorem I. Thus, we can use the theorem in both ways. If a fully nonlinear solution, presumably obtained by numerical methods, stays below ϵ, then theorem I guarantees that a truncated theory can reproduce such a solution. The reverse is also true. If the truncated theory leads to a solution that stays below ϵ, then theorem I guarantees that this is a true solution reproducible by numerical evolution of the full nonlinear theory.

Also note that the truncated theory might belong to an expansion that does not really converge to the full nonlinear theory. It is quite common in field theories that a naive expansion is asymptotic instead of convergent. Theorem I is not concerned with whether such a full expansion is convergent or not. It only requires that the truncated theory is a good approximation to the full theory up to some specified order, as stated in Eq. (2.6). Divergence of an expansion scheme at higher orders does not invalidate our result.[6]

Finally, if one takes a closer look at Eq. (2.8), one can see that if n is allowed to be larger than the ϵ^{1-m} time scale, for example, $n \sim \epsilon^{-s}$ with $s > m-1$, then Δ_n fails to be bounded from above in the $\epsilon \to 0$ limit. Since the upper bound we use is already quite optimal, we believe that the truncated theory breaks down at any longer time scale. In particular, the theory does not concern l. Namely, independent of how small the truncated error is, accumulation beyond the ϵ^{1-m} time scale always makes the truncated dynamics a bad approximation for the full theory. Thus, the conventional wisdom only requires a small correction. Usually, the truncated theory breaks down at the ϵ^{1-m} time scale. Occasionally, it can still hold at exactly this time scale but breaks down at any longer time scale.

[6]We thank Jorge Santos for pointing out the importance of this point.

III. THEOREM II: CONDITIONALLY PRESERVED SYMMETRY

We consider an example in which the truncated theory has an approximate scaling symmetry. Let $L(x) = x$, $g(x) = L(x) + G(x)$, such that for all $\|x\|$, $\|y\| < r$,

$$\|G(x)\| < a\|x\|^m, \tag{3.1}$$

$$\|G(x) - G(y)\| < b \cdot \|x - y\| \mathrm{Max}(\|x\|, \|y\|)^{m-1}, \tag{3.2}$$

$$\|G(x) - N^m G(x/N)\| < d\|x\|^p, \tag{3.3}$$

for a given $p > m$ and any $N > 1$. Namely, the linear theory is trivial where $L_n = x$ does not evolve with n. The only evolution for g_n comes from the function $G(x)$, which is for many purposes effectively an "x^m term." In this case, it is reasonable to expect a rescaling symmetry: Reducing the amplitude by a factor of N, but evolving for a time longer by a factor of N^{m-1}, leads to roughly the same result.

Theorem II: For any finite $\delta > 0$ and $\alpha > 0$, there exists $0 < \epsilon < r$ such that if $\|g_n(x)\| < \epsilon$ for all $0 \le n < \alpha\epsilon^{1-m}$, then

$$\|N g_n(x/N) - (1-\beta) g_{n'}(x) - \beta g_{n'+1}(x)\| < \delta\epsilon. \tag{3.4}$$

Here, $n' = \lfloor (nN^{1-m}) \rfloor$ is the largest integer smaller than or equal to (nN^{1-m}), and $\beta = (nN^{1-m}) - n'$. This should be valid for any $N > 1$ and for $0 \le n < \alpha(\epsilon/N)^{1-m}$.

The physical intuition is as follows. Every term in the rescaled sequence stays arbitrarily close to some weighted average of the terms in the original sequence, which exactly corresponds to the appropriate "time" (number of steps) of the rescaling. We first prove this for a special case, $N = 2^{\frac{1}{m-1}}$. This case is particularly simple since such rescaling exactly doubles the length of the sequence, and β will be either 0 or 1/2, which leads to two specific inequalities to prove:

$$\left\| 2^{\frac{1}{m-1}} g_{2n-1}(x/2^{\frac{1}{m-1}}) - \frac{g_{n-1}(x) + g_n(x)}{2} \right\| \le C \cdot \epsilon^q, \tag{3.5}$$

$$\| 2^{\frac{1}{m-1}} g_{2n}(x/2^{\frac{1}{m-1}}) - g_n(x) \| \le C \cdot \epsilon^q, \tag{3.6}$$

for some $C > 0$ and $q > 1$. This will again be done by a mathematical induction.

During the process, it should become clear that the proof can be generalized to any $N > 1$. We will not present such a proof because the larger variety of β values makes it more tedious, although it is still straightforward. However, for the self-completeness of this paper, what we need next is for N to be arbitrarily large. Through another mathematical induction, we can easily prove theorem II for $N = 2^{\frac{k}{m-1}}$ an arbitrarily positive integer k. It is still a bit tedious, so we will present this in the Appendix.

Proof for $N = 2^{\frac{1}{m-1}}$: We start by defining the monotonically increasing function

$$\Delta_n = \left(\frac{d}{b}2^{m-1}\epsilon^{p-m+1} + \frac{a}{2}\epsilon^m\right)\left[\left(1 + \frac{b}{2}(2\epsilon)^{m-1}\right)^n - 1\right], \tag{3.7}$$

with the properties

$$\Delta_n < \Delta_{\alpha(\epsilon/N)^{1-m}} \tag{3.8}$$

$$= \left(\frac{d}{b}2^{m-1}\epsilon^{p-m+1} + \frac{a}{2}\epsilon^m\right)\left[\left(1 + \frac{b}{2}(2\epsilon)^{m-1}\right)^{\alpha(\epsilon/N)^{1-m}} - 1\right]$$

$$< \left(\frac{d}{b}2^{m-1}\epsilon^{p-m+1} + \frac{a}{2}\epsilon^m\right)[e^{2^{m-1}\alpha b/4} - 1] < C\cdot\epsilon^q,$$

$$\Delta_{n+1} = \Delta_n\left(1 + \frac{b}{2}(2\epsilon)^{m-1}\right) + \frac{d}{2}\epsilon^p + \frac{ab}{4}\epsilon^m(2\epsilon)^{m-1}. \tag{3.9}$$

The meaning of Eq. (3.8) is that, in the range we are concerned with, Δ_n is bounded from above by a power of ϵ higher than 1 since $q = \mathrm{Min}\{p - m + 1, m\}$. Therefore, we can always choose ϵ small enough such that

$$\|g_n(x)\| + \Delta_n < \epsilon + C\cdot\epsilon^q < r. \tag{3.10}$$

Given our choice of ϵ, we can employ mathematical induction to prove that

$$\left\|Ng_{2n-1}(x/N) - \frac{g_{n-1}(x) + g_n(x)}{2}\right\| \le \Delta_{2n-1} \tag{3.11}$$

$$\|Ng_{2n}(x/N) - g_n(x)\| \le \Delta_{2n}, \tag{3.12}$$

which proves Eqs. (3.5) and (3.6).

First, we observe that for $n = 0$,

$$\|Ng_0(x/N) - g_0(x)\| = \left\|n\frac{x}{N} - x\right\| = 0 < C\cdot\epsilon^q \tag{3.13}$$

is obviously true. Then, we assume that Eq. (3.12) is true for n in the original sequence and for $2n$ in the rescaled sequence. We can next prove the $(2n+1)$ term in the rescaled sequence,

$$\begin{aligned}\left\|Ng_{2n+1}(x/N) - \frac{g_n(x)+g_{n+1}(x)}{2}\right\| &= \left\|Ng_{2n}(x/N) + NG(g_{2n}(x/N)) - g_n(x) - \frac{1}{2}G(g_n(x))\right\| \\ &= \left\|Ng_{2n}(x/N) + NG(g_{2n}(x/N)) - g_n(x) - \frac{1}{2}G(g_n(x)) + NG(g_n(x)/N) - NG(g_n(x)/N)\right\| \\ &< \|Ng_{2n}(x/N) - g_n(x)\| + N\|G(g_{2n}(x/n)) - G(g_n(x)/N)\| \\ &\quad + \frac{1}{N^{m-1}}\|G(g_n(x)) - N^mG(g_n(x)/N)\| < \Delta_{2n} + \Delta_{2n}\frac{b}{N^{m-1}}(\|\Delta_{2n}\| + \|g_n(x)\|)^{m-1} + \frac{d}{2}\epsilon^p \\ &< \Delta_{2n} + \Delta_{2n}\frac{b}{2}(\Delta_{2n} + \epsilon)^{m-1} + \frac{d}{2}\epsilon^p < \Delta_{2n} + \Delta_{2n}\frac{b}{2}(2\epsilon)^{m-1} + \frac{d}{2}\epsilon^p < \Delta_{2n+1}.\end{aligned} \tag{3.14}$$

Similarly, we can prove the $(2n+2)$ term in the rescaled sequence, which is the $(n+1)$ term in the original sequence,

$$\begin{aligned}\|Ng_{2n+2}(x/N) - g_{n+1}(x)\| = &\left\|Ng_{2n+1}(x/N) + NG(g_{2n+1}(x/N)) - \frac{1}{2}g_{n+1}(x)\right. \\ &- \frac{1}{2}g_n(x) - \frac{1}{2}G(g_n(x)) + NG\left(\frac{g_n(x)+g_{n+1}(x)}{N^m}\right) - NG\left(\frac{g_n(x)+g_{n+1}(x)}{N^m}\right) \\ &\left. + NG\left(\frac{g_n(x)}{N}\right) - NG\left(\frac{g_n(x)}{N}\right)\right\| \\ < &\left\|Ng_{2n+1}(x/N) - \frac{g_{n+1}(x)+g_n(x)}{2}\right\| + N\left\|G(g_{2n+1}(x/N)) - G\left(\frac{g_n(x)+g_{n+1}(x)}{N^m}\right)\right\| \\ &+ N\left\|G\left(\frac{g_n(x)+g_{n+1}(x)}{N^m}\right) - G\left(\frac{g_n(x)}{N}\right)\right\| + N\left\|G\left(\frac{g_n(x)}{N}\right) - \frac{1}{N^{m-1}}G(g_n(x))\right\| \\ < &\ \Delta_{2n+1} + \Delta_{2n+1}\frac{b}{2}(2\epsilon)^{m-1} + \frac{d}{2}\epsilon^p + \frac{ab}{4}\epsilon^m(2\epsilon)^{m-1} = \Delta_{2n+2}.\end{aligned} \tag{3.15}$$

This completes the mathematical induction.

Equation (3.9) takes basically the same form as Eq. (2.8). Thus, theorem II also only holds up to exactly the ϵ^{1-m} time scale, but not any longer.

IV. APPLICATION: PERSISTENCE OF STABILITY ISLANDS

First, we review the "stability island conjecture" argued by Dias, Horowitz, Marolf, and Santos in Ref. [4]. Numerical simulations suggest that given a small but finite initial amplitude $\phi_{\text{init}} \sim \epsilon$ in AdS space with the Dirichlet boundary condition, dynamical evolution can lead to black hole formation at the long time scale $T \sim \epsilon^{-2}$ [5].[7] Meanwhile, some initial conditions do not lead to black holes at the same time scale. In particular, there are special solutions (set of measure zero) that stay exactly as they are and do not collapse. These especially stable solutions are called geons (in pure gravity) or boson stars or oscillons (scalar field)[6,11,14].[8] At finite amplitudes, they are not only stable themselves, but they also stabilize an open neighborhood in the phase space, forming stability islands that prevent nearby initial conditions from collapsing into black holes in the $\sim\epsilon^{-2}$ time scale.

Dias, Horowitz, Marolf, and Santos argued that such a stabilization effect can be understood as breaking the AdS resonance.[9] It should lose strength as the geon's own amplitude decreases. Thus, such stability islands disappear in the limit of zero amplitude. The easiest way to summarize their conjecture is shown in the cuspy phase-space diagram in Fig. 1. Other than the measure-zero set of exact geons or boson stars, noncollapsing solutions at finite amplitude will all end up collapsing as the amplitude goes to zero.

Next, we show that the requirements of both theorems I and II are applicable to the AdS (in)stability problem. For simplicity, we present the analysis on a massless scalar field in global AdS space of the Dirichlet boundary condition. The metric fluctuation in pure gravity will also meet the requirements [4,16]. We avoid going into specific details of the AdS dynamics, but we provide the relevant papers where those details can be found.

(i) The linear space $\mathcal{H}$ we used to state both theorems (see the beginning of Sec. II) contains all smooth functions $\phi(\vec{r})$ on the domain of the entire spatial slice of the AdS space with one global time between.

(ii) The function L evolves one such function forward for one "AdS period," namely, $T = 2\pi R_{\text{AdS}}$ in the explanation right below Eq. (2.6), using the fixed background equation of motion. It includes no gravitational self-interaction and is a linear function. Actually, since the AdS spectrum has integer eigenvalues, the evolution is exactly periodic [25,26]. $L(x) = x$ is trivial, automatically conserves length, and also meets the requirement for theorem II.

(iii) The definition of the norm is trickier. We first evolve x, using the fixed-background evolution, for exactly $2\pi R_{\text{AdS}}$, and examine the maximum local energy density that occurred during such an evolution. The norm is defined to be the square root of this value, $\sqrt{\rho_{\max}}$. The evolution is linear, and the quantity is both a maximum and effectively an absolute value; thus, it satisfies the triangular inequality.[10]

(iv) The actual dynamics, including Einstein equations, is clearly nonlinear. When the maximum energy density is small, the gravitational backreaction is well bounded. One can perform a recursive expansion in which the leading-order correction to the linear dynamics comes from coupling to its own energy, $\rho\phi \propto \phi^3$ [5,18,27]. A theory truncated at this order and the full nonlinear theory can be our f and g, interchangeably, in theorem I with $m = 3$.[11]

(v) The ϕ^3 contributions calculated in different approximation methods might be different [5,18,27], but they all satisfy the approximate rescaling symmetry required for the function G in theorem II.

Now that we have established the applicability of both theorems in this paper, the stability island conjecture can be disproved in three simple steps.

(1) At a small but finite amplitude where measure-nonzero sets of noncollapsing solutions exist (the outermost thick arc in Fig. 1), apply theorem I to translate them into solutions in the truncated theory.

(2) Use theorem II to scale down the above solutions to arbitrarily small amplitudes. That means projecting radially in Fig. 1 into an arc of the same angular span.

(3) Use theorem I again to translate these rescaled solutions in the truncated theory back to the full nonlinear theory. This establishes the existence of noncollapsing solutions as a set of measure nonzero (an arc of finite angular span in Fig. 1).[12]

[7]Note that for this purpose, $m = 3$; thus, ϵ^{-2} is the relevant time scale.

[8]There are also quasiperiodic solutions that do not stay exactly the same but demonstrate a long-term periodic behavior, and the energy density never gets large [1].

[9]In some sense, this argument [4] provides stronger support for noncollapsing solutions to have a nonzero measure because it goes beyond spherical symmetry. Current numerical results are limited to spherical symmetry; thus, strictly speaking, we cannot establish a nonzero measure for either collapsing or noncollapsing results. This is why controversies over some numerical results [23,24] should not undermine the belief that noncollapsing solutions form a set of nonzero measure at a finite amplitude.

[10]The reason why we adopt this tortuous definition of norm is to guarantee that the gravitational interaction during one AdS time stays weak when the norm is small; thus, we can apply both theorems. Note that defining total energy as the norm would not serve this purpose.

[11]Such expansion, continued to higher orders, is likely to be only asymptotic instead of convergent. As explained in Sec. II, that does not cause a problem for our theorems.

[12]This only works for rescaling down to smaller amplitudes. Rescaling to larger amplitudes can easily exceed the radius of validity of perturbative expansion even at short time scales.

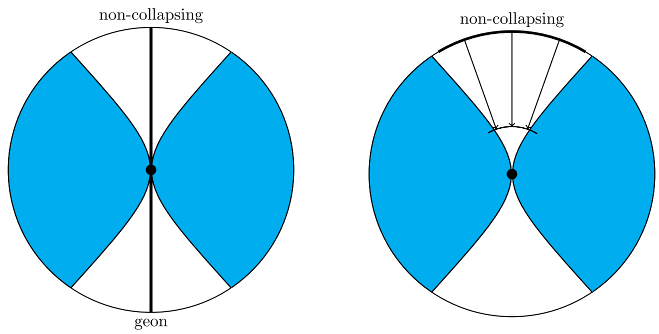

FIG. 1 (color online). Phase-space diagrams of small initial perturbations around empty AdS (central black dot) according to the stability island conjecture. The radial direction represents field amplitude (total energy), and the angular direction represents field profile shape (energy distribution). Initial perturbations in the shaded (blue) region will collapse into black holes at the $\sim\epsilon^{-2}$ time scale, while those in the unshaded region, around the exactly stable geons (thick black line), will not. The unshaded region is cuspy, showing that according to the conjecture, the angular span of noncollapsing perturbations goes to zero as the amplitude goes to zero. The right panel demonstrates the usage of both theorems we proved in this paper. We can transport the known, noncollapsing solutions, directly in the radial direction, to an arc of identical angular span at an arbitrarily smaller radius. This is in direct contradiction with the cuspy nature of the unshaded region.

Thus, we have established that the measure-nonzero neighborhood stabilized by a geon at finite amplitude, provided that it never evolves to high local energy density during the long time scale, directly guarantees the same measure-nonzero, noncollapsing neighborhood at arbitrarily smaller amplitudes. This directly contradicts the stability island conjecture.

It is interesting to note that the collapsing solutions always have large energy density at a certain point; thus, neither theorem we proved here is applicable. As a result, one cannot establish their existence at arbitrarily small amplitudes through a similar process. Therefore, the opposite possibility to the stability island conjecture, that collapsing solutions disappear into a set of measure zero at zero amplitude, is still consistent with current evidence.

ACKNOWLEDGMENTS

We thank the Kavli Institute for Theoretical Physics at Santa Barbara for hospitality during the workshop "Quantum Gravity Foundation, From UV to IR." In particular, we were inspired by our discussion with Ben Freivogel, Don Marolf, and Gary Horowitz during the workshop. We are also grateful for the discussions with Piotr Bizon, Alex Buchel, Matt Lippert, Luis Lehner, and Andrzej Rostworowski.

APPENDIX: ARBITRARILY SMALL RESCALING

In Sec. III, we have proven theorem II for $N = 2^{\frac{1}{m-1}}$. Now, we generalize it to arbitrary $N' = 2^{\frac{k}{m-1}} = N^k$, for any $k \in \mathbb{N}^+$:

$$\|N^k g_n(x/N^k) - (1-\beta_k(n)) g_{\lfloor \frac{n}{2^k} \rfloor}(x) - \beta_k(n) g_{\lfloor \frac{n}{2^k} \rfloor + 1}(x)\| \leq C' \cdot \epsilon^q, \tag{A1}$$

where we have written down explicitly the dependence of β on k and n:

$$\beta_k(n) = \frac{2}{n^k} - \left\lfloor \frac{n}{2^k} \right\rfloor, \tag{A2}$$

which possesses the following properties for positive integers j and l:

$$\beta_{k+1}(2l) = \beta_k(l), \qquad \text{this is always true;} \tag{A3}$$

$$\beta_{k+1}(2l+1) = \frac{1}{2}\beta_k(l) + \frac{1}{2}\beta_k(l+1), \quad \text{for } l+1 \neq j \cdot 2^k; \tag{A4}$$

$$\beta_{k+1}(2l+1) = \frac{1}{2}\beta_k(l) + \frac{1}{2}[1-\beta_k(l+1)], \quad \text{for } l+1 = j \cdot 2^k. \tag{A5}$$

These follow naturally from the properties of the floor function that

$$\left\lfloor \frac{2l+1}{2^{k+1}} \right\rfloor = \left\lfloor \frac{l}{2^k} \right\rfloor \qquad \text{is always true,} \tag{A6}$$

and

$$\left\lfloor \frac{2l+1}{2^{k+1}} \right\rfloor = \left\lfloor \frac{l+1}{2^k} \right\rfloor, \quad \text{when } l \neq j \cdot 2^k - 1; \tag{A7}$$

$$\left\lfloor \frac{2l+1}{2^{k+1}} \right\rfloor = \left\lfloor \frac{l+1}{2^k} \right\rfloor - 1, \quad \text{when } l = j \cdot 2^k - 1. \tag{A8}$$

We now define

$$F_k \equiv C \cdot \epsilon^q \sum_{i=0}^{k} N^{i(1-q)} = \frac{C \cdot \epsilon^q}{1 - N^{1-q}} (1 - N^{k(1-q)}) \leq C' \cdot \epsilon^q, \tag{A9}$$

for $C' = C/(1 - N^{1-q})$. This converges as $k \to \infty$, since $1 - q < 0$, and satisfies the recursive relation

$$F_{k+1} = F_k + C \cdot N^{k(1-q)}. \tag{A10}$$

Now we prove Eq. (A1) by induction. We have already shown that it holds for $k = 1$ in Sec. III; hence, assuming that it holds for arbitrary k, we want to show that it holds for $k + 1$ as well.

It is helpful to split the proof into three parts: one for $n = 2l$, one for $n = 2l + 1$, with $l \neq j \cdot 2^k - 1$, and one for $n = 2l + 1$, with $l = j \cdot 2^k - 1$.

(1) *Part 1:* $n = 2l$

$$\begin{aligned} &\|N^{k+1}g_{2l}(x/N^{k+1}) - (1-\beta_{k+1}(2l))g_{\lfloor\frac{2l}{2^{k+1}}\rfloor}(x) - \beta_{k+1}(2l)g_{\lfloor\frac{2l}{2^{k+1}}\rfloor+1}(x)\| \\ &= \left\|N\cdot N^k g_{2l}\left(\frac{x/N^k}{N}\right) - (1-\beta_k(l))g_{\lfloor\frac{2l}{2^{k+1}}\rfloor}(x) - \beta_k(l)g_{\lfloor\frac{2l}{2^{k+1}}\rfloor+1}(x)\right\| < N^k\left\|Ng_{2l}\left(\frac{x/N^k}{N}\right) - g_l(x/N^k)\right\| \\ &\quad + \|N^k g_l(x/N^k) - (1-\beta_k(l))g_{\lfloor\frac{l}{2^k}\rfloor}(x) - \beta_k(l)g_{\lfloor\frac{l}{2^k}\rfloor+1}(x)\| < C\cdot N^{k(1-q)} + F_k = F_{k+1}. \end{aligned} \tag{A11}$$

(2) *Part 2:* $n = 2l+1$, *with* $l \neq j\cdot 2^k - 1$

$$\begin{aligned} &\|N^{k+1}g_{2l+1}(x/N^{k+1}) - (1-\beta_{k+1}(2l+1))g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x) - \beta_{k+1}(2l+1)g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\| \\ &= \Bigg\|N\cdot N^k g_{2l+1}\left(\frac{x/N^k}{N}\right) - \left(1-\frac{1}{2}\beta_k(l) - \frac{1}{2}\beta_k(l+1)\right)g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x) \\ &\quad - \frac{1}{2}(\beta_k(l)+\beta_k(l+1))g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\Bigg\| < N^k\left\|Ng_{2l+1}\left(\frac{x/N^k}{n}\right) - \frac{g_l(x/N^k)+g_{l+1}(x/N^k)}{2}\right\| \\ &\quad + \frac{1}{2}\|N^k g_l(x/N^k) - (1-\beta_k(l))g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x) - \beta_k(l)g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\| \\ &\quad + \frac{1}{2}\|N^k g_{l+1}(x/N^k) - (1-\beta_k(l+1))g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x) - \beta_k(l+1)g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\| < C\cdot N^{k(1-q)} + 2\frac{1}{2}F_k = F_{k+1}. \end{aligned} \tag{A12}$$

(3) *Part 3:* $n = 2l+1$, *with* $l = j\cdot 2^k - 1$

$$\begin{aligned} &\|N^{k+1}g_{2l+1}(x/N^{k+1}) - (1-\beta_{k+1}(2l+1))g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x) - \beta_{k+1}(2l+1)g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\| \\ &= \Bigg\|N\cdot N^k g_{2l+1}\left(\frac{x/N^k}{N}\right) - \left(1-\frac{1}{2}\beta_k(l) - \frac{1}{2}(1-\beta_k(l+1))\right)g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x) \\ &\quad - \frac{1}{2}(\beta_k(l)+(1-\beta_k(l+1)))g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\Bigg\| \\ &< N^k\left\|Ng_{2l+1}(x/N^k) - \frac{g_l(x/N^k)+g_{l+1}(x/N^k)}{2}\right\| \\ &\quad + \frac{1}{2}\left\|N^k g_l(x/N^k) - \beta_k(l)g_{\lfloor\frac{l}{2^k}\rfloor+1}(x) - (1-\beta_k(l))g_{\lfloor\frac{l}{2^k}\rfloor}(x)\right\| \\ &\quad + \frac{1}{2}\|N^k g_{l+1}(x/N^k) - \beta_k(l+1)g_{\lfloor\frac{l+1}{2^k}\rfloor-1}(x) - (1-\beta_k(l+1))g_{\lfloor\frac{l+1}{2^k}\rfloor}(x)\| \\ &< C\cdot N^{k(1-q)} + 2\frac{1}{2}F_k = F_{k+1}. \end{aligned} \tag{A13}$$

Here, we have used the fact that

$$\beta_k(l+1)g_{\lfloor\frac{l+1}{2^k}\rfloor-1}(x) = \beta_k(l+1)g_{\lfloor\frac{l+1}{2^k}\rfloor+1}(x), \tag{A14}$$

since $\beta_k(j\cdot 2^k) = 0$.

[1] V. Balasubramanian, A. Buchel, S. R. Green, L. Lehner, and S. L. Liebling, Holographic Thermalization, Stability of Anti–de Sitter Space, and the Fermi-Pasta-Ulam Paradox, Phys. Rev. Lett. **113**, 071601 (2014).
[2] B. Craps, O. Evnin, and J. Vanhoof, Renormalization group, secular term resummation and AdS (in)stability, J. High Energy Phys. 10 (2014) 48.
[3] C. M. Will, Post-Newtonian effects in N-body dynamics: Conserved quantities in hierarchical triple systems, Classical Quantum Gravity **31**, 244001 (2014).
[4] O. J. Dias, G. T. Horowitz, D. Marolf, and J. E. Santos, On the nonlinear stability of asymptotically anti-de Sitter solutions, Classical Quantum Gravity **29**, 235019 (2012).
[5] P. Bizon and A. Rostworowski, Weakly Turbulent Instability of Anti–de Sitter Spacetime, Phys. Rev. Lett. **107**, 031102 (2011).
[6] O. J. Dias, G. T. Horowitz, and J. E. Santos, Gravitational turbulent instability of anti-de Sitter space, Classical Quantum Gravity **29**, 194002 (2012).
[7] H. de Oliveira, L. A. Pando Zayas, and E. Rodrigues, A Kolmogorov-Zakharov spectrum in AdS gravitational collapse, Phys. Rev. Lett. **111**, 051101 (2013).
[8] S. L. Liebling, Nonlinear collapse in the semilinear wave equation in AdS space, Phys. Rev. D **87**, 081501 (2013).
[9] M. Maliborski, Instability of Flat Space Enclosed in a Cavity, Phys. Rev. Lett. **109**, 221101 (2012).
[10] A. Buchel, L. Lehner, and S. L. Liebling, Scalar collapse in AdS, Phys. Rev. D **86**, 123011 (2012).
[11] A. Buchel, S. L. Liebling, and L. Lehner, Boson stars in AdS Spacetime, Phys. Rev. D **87**, 123006 (2013).
[12] P. Bizon, Is AdS stable?, Gen. Relativ. Gravit. **46**, 1724 (2014).
[13] M. Maliborski and A. Rostworowski, Lecture notes on turbulent instability of anti-de Sitter spacetime, Int. J. Mod. Phys. A **28**, 1340020 (2013)..
[14] M. Maliborski and A. Rostworowski, Time-Periodic Solutions in an Einstein AdS–Massless-Scalar-Field System, Phys. Rev. Lett. **111**, 051102 (2013).
[15] M. Maliborski and A. Rostworowski, What drives AdS unstable?, Phys. Rev. D **89**, 124006 (2014).
[16] G. T. Horowitz and J. E. Santos, Geons and the instability of anti-de Sitter spacetime, arXiv:1408.5906.
[17] P. Basu, C. Krishnan, and A. Saurabh, A stochasticity threshold in holography and the instability of AdS, Int. J. Mod. Phys. A **30**, 1550128 (2015).
[18] F. V. Dimitrakopoulos, B. Freivogel, M. Lippert, and I.-S. Yang, Position space analysis of the AdS (in)stability problem, J. High Energy Phys. 08 (2015) 077.
[19] I.-S. Yang, Missing top of the AdS resonance structure, Phys. Rev. D **91**, 065011 (2015).
[20] P. Basu, C. Krishnan, and P. Bala Subramanian, AdS (In) stability: Lessons from the scalar field, Phys. Lett. B **746**, 261 (2015).
[21] A. Buchel, S. R. Green, L. Lehner, and S. L. Liebling, Universality of non-equilibrium dynamics of CFTs from holography, arXiv:1410.5381.
[22] P. Bizon, M. Maliborski, and A. Rostworowski, Resonant dynamics and the instability of anti-de Sitter spacetime, arXiv:1506.03519.
[23] P. Bizon and A. Rostworowski, Comment on "Holographic Thermalization, Stability of AdS, and the Fermi-Pasta-Ulam-Tsingou Paradox" by V. Balasubramanian et al., Phys. Rev. Lett. **115**, 049101 (2015).
[24] A. Buchel, S. R. Green, L. Lehner, and S. L. Liebling, Reply to Comment on Two-Mode Stability Islands around AdS, Phys. Rev. Lett. **115**, 049102 (2015)
[25] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Holographic representation of local bulk operators, Phys. Rev. D **74**, 066009 (2006).
[26] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Local bulk operators in AdS/CFT: A holographic description of the black hole interior, Phys. Rev. D **75**, 106001 (2007).
[27] P. Bizon, T. Chmaj, and A. Rostworowski, Late-time tails of a self-gravitating massless scalar field revisited, Classical Quantum Gravity **26**, 175006 (2009).