

## Perturbations of the almost Killing equation and their implications

Sumanta Chakraborty<sup>1,2</sup> and Justin C. Feng<sup>3</sup>

<sup>1</sup>*School of Mathematical and Computational Sciences, Indian Association for the Cultivation of Science, Kolkata-700032, India*

<sup>2</sup>*School of Physical Sciences, Indian Association for the Cultivation of Science, Kolkata-700032, India*

<sup>3</sup>*CENTRA, Departamento de Física, Instituto Superior Técnico IST, Universidade de Lisboa UL, Avenida Rovisco Pais 1, 1049 Lisboa, Portugal*



(Received 7 December 2020; accepted 12 March 2021; published 12 April 2021)

Killing vectors play a crucial role in characterizing the symmetries of a given spacetime. However, realistic astrophysical systems are in most cases only approximately symmetric. Even in the case of an astrophysical black hole, one might expect Killing symmetries to exist only in an approximate sense due to perturbations from external matter fields. In this work, we consider the generalized notion of Killing vectors provided by the almost Killing equation, and study the perturbations induced by a perturbation of a background spacetime satisfying exact Killing symmetry. To first order, we demonstrate that for nonradiative metric perturbations (that is, metric perturbations with nonvanishing trace) of symmetric vacuum spacetimes, the perturbed almost Killing equation avoids the problem of an unbounded Hamiltonian for hyperbolic parameter choices. For traceless metric perturbations, we obtain similar results for the second-order perturbation of the almost Killing equation, with some additional caveats. Thermodynamical implications are also explored.

DOI: [10.1103/PhysRevD.103.084020](https://doi.org/10.1103/PhysRevD.103.084020)

### I. INTRODUCTION

Symmetries are central to our understanding of the physical world and play a key role in describing a wide range of physical systems, from the determination of the Lagrangian for a mechanical system to the lattice structure of crystalline substances. This extends to general relativity and relativistic theories of gravity: Symmetries and the Killing vectors that formalize them are useful for characterizing the properties of spacetime and matter. For example, the existence of a timelike Killing vector field ensures that the spacetime is time-translation invariant, leading to conserved definitions for energy for test particles and matter. Similarly, the existence of a closed spacelike Killing vector field ensures rotational invariance of the spacetime geometry, leading to a conserved definition for angular momentum. Moreover, many astrophysical systems are approximately described by spacetime geometries admitting such Killing vectors. However, the exact nature of these symmetries is lost in realistic systems due to dynamical behavior of and irregularities in the matter configurations. This scenario can arise in various contexts; e.g., when one drops a cup of coffee into a black hole (considering the gravitational backreaction), the resulting perturbed spacetime no longer inherits the exact Killing symmetry. Fortunately, one may still construct certain generalizations of Killing vector fields in such circumstances, which are useful for understanding generalizations

of conserved quantities (such as energy and momentum) for gravitating systems that lack exact symmetries.

The literature contains several approaches for defining generalized Killing vectors and symmetries. Specific examples include Matzner's eigenvector approach [1], which has recently been of interest for studying quantum geometries in causal dynamical triangulations [2], symmetry-seeking coordinates [3], affine collineations [4], and the almost Killing equation (henceforth, AKE) [5,6]. The latter approach, the generalized Killing vectors defined by the AKE, forms the subject of this paper.

The generalized Killing vector fields (henceforth, GKV) associated with the AKE may be used to define conserved charges in spacetimes with no exact Killing symmetries. Given some notion of the GKV, the generalized Komar current, as defined in [7,8], may be used to construct generalizations of the usual Komar charges—explicit examples have been constructed and studied in [9,10] (see also [11] for further generalizations of the Komar current). For example, in [12], it was shown that the generalized Komar current for solutions of the AKE, which are the GKVs, can provide a measure of the matter content of the physical system under consideration. It was also demonstrated in [12] that GKVs may be used with the generalized Komar current to obtain a Gauss law for systems of black holes in vacuum and matter distributions with compact support if the GKV is divergenceless or for a certain choice of parameters associated with the AKE.

Though one might hope that, for sufficiently small perturbations of symmetric spacetimes, the solutions of the AKE are close to that of an exact Killing vector field, GKV's do not necessarily approximate Killing vectors in the sense that the components of  $\nabla_{(\mu}\chi_{\nu)}$  can be large compared to that of  $\chi^\alpha$  (where  $\chi^\alpha$  is a GKV), even in Minkowski spacetime. One might postulate that an appropriate choice for initial data for the AKE will ensure that  $\nabla_{(\mu}\chi_{\nu)}$  is small in the sense that the components of  $\nabla_{(\mu}\chi_{\nu)}$  are much smaller than that of  $\chi^\alpha$  for some normalization. This matter was studied to some degree in [12], which examines the hyperbolicity and Hamiltonian stability of the system described by the AKE. There, parameter choices were identified in which the AKE is strongly and weakly hyperbolic, and also in which the system admits ghost modes and unbounded Hamiltonians. Ghosts and unbounded Hamiltonians are potentially dangerous, as they may correspond to runaway behavior which can potentially drive the solution far from the Killing condition  $\nabla_{(\mu}\chi_{\nu)} = 0$ , even if the initial data approximately satisfy this condition (and its time derivative). Though there is no parameter choice for which the generic AKE system is both hyperbolic and has a bounded Hamiltonian, it was shown in [12] that in vacuum ( $R_{\mu\nu} = 0$ ) spacetime and for initial data satisfying  $\nabla \cdot \chi = 0$  and its derivative, the system yields a constraint which renders it dynamically equivalent to a system with a bounded Hamiltonian and simultaneously equivalent to a strongly hyperbolic system. Moreover, it was argued that for appropriate initial data and falloff conditions, the AKE can provide a notion of an approximate Killing vector in a neighborhood of spatial infinity of asymptotically flat spacetimes.

Despite the promising results presented in [12] for the vacuum case, these do not in general extend to the non-vacuum ( $R_{\mu\nu} \neq 0$ ) case. Therefore, it is not immediately apparent that the AKE can be simultaneously well posed and equivalent to a system with a bounded Hamiltonian for spacetimes containing matter. On the other hand, for perturbations of spacetimes that admit an exact Killing vector, one might expect the AKE for the perturbed spacetime to admit solutions that approximate Killing vectors. Thus, one of the primary aims of this article is to construct perturbative solutions to the AKE for perturbations of spacetimes which admit Killing vectors and to study their properties and the interpretation of the resulting generalized Komar currents and charges. Additionally, we would like to explore the connection of the perturbed Komar current and charges with the thermodynamic behavior of black hole spacetimes, e.g., the first law. As we will show, there is a close correspondence between the AKE and black hole thermodynamics.

The paper is organized as follows: In Sec. II we will review the AKE and shall present a physically interesting scenario, namely, that of the Vaidya spacetime, where some

of the key aspects of the AKE will be demonstrated. Subsequently, the evolution of the GKV's in the perturbed spacetime will be presented in Sec. III from both the action formalism and also from the perturbation of the AKE itself. The stability of the perturbed AKE, as well as its hyperbolicity, will be studied in Sec. IV, before discussing the nature of the solution of the AKE for both first- and second-order perturbations in Sec. V. Finally, the thermodynamic interpretation of the AKE will be depicted in Sec. VI, before presenting the concluding remarks in Sec. VII.

Notations and conventions: Throughout this paper, we will use the mostly positive signature convention, such that the Minkowski metric in the Cartesian coordinates has the following form:  $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ . The four-dimensional spacetime indices will be denoted by the greek letters  $\mu, \nu, \alpha, \dots$ . We will work in units such that the fundamental constants have the values  $G = c = \hbar = 1$ . Throughout the article, indices on quantities which appear in arguments will be denoted with superscript and subscript dots; for instance, the arguments in  $\mathcal{A}[\chi]$  and  $\mathcal{O}(h..)$  represent the quantities  $\chi^\mu$  and  $h_{\mu\nu}$ .

## II. THE ALMOST KILLING EQUATION: A BRIEF REVIEW

In this section, we will briefly review the almost Killing equation, where the motivation for its construction and its various properties will be discussed in detail. In addition, we will also present the Vaidya geometry as an example of arriving at a solution of the almost Killing equation.

### A. Motivation, construction, and properties

A Killing vector field  $\xi^\mu$  is defined as one which satisfies the Killing equation  $\mathcal{L}_\xi g_{\mu\nu} = 2\nabla_{(\mu}\xi_{\nu)} = 0$ . The divergence of the Killing equation takes the form

$$\square \xi^\alpha + R^\alpha{}_\beta \xi^\beta = 0. \quad (2.1)$$

As evident, Eq. (2.1) takes the form of a wave equation; on geometries that do not admit Killing vectors, one can nonetheless construct generalizations of the Killing equation by solving Eq. (2.1) for an appropriate set of initial data. The AKE is a generalization of Eq. (2.1), and is given by the following formula:

$$\square \chi^\alpha + R^\alpha{}_\beta \chi^\beta + \nabla^\alpha [(1 - \mu) \nabla \cdot \chi] = 0, \quad (2.2)$$

where  $\mu$  is a scalar, which in previous literature is assumed to be a constant; for generality, we do not assume this to be the case here. The solution of Eq. (2.2), i.e.,  $\chi^\alpha$ , is the GKV. It is straightforward to verify that Killing vectors satisfy the AKE; it is in this sense that solutions of the AKE may be regarded as generalizations of Killing vectors. As discussed in [12], GKV's are not necessarily approximate Killing vectors in the following sense. The vector  $\chi^\alpha$  satisfies the

AKE if the tensor  $Q_{\mu\nu} = \nabla_{(\mu}\chi_{\nu)}$  is transverse and trace-free; however, the components of transverse and trace-free tensors need not be small.<sup>1</sup>

It is instructive to derive any evolution equation from an action principle, and the AKE is no different. The AKE presented in Eq. (2.2) may be derived from the following action (see the Appendix for a derivation of the AKE from this action functional):

$$\mathcal{A}[\chi] = \int_{\mathcal{M}} d^4x \sqrt{-g} \left( -\nabla^{(\alpha}\chi^{\beta)} \nabla_{(\alpha}\chi_{\beta)} + \frac{1}{2}\mu(\nabla \cdot \chi)^2 \right). \quad (2.3)$$

Here,  $\mathcal{M}$  denotes the spacetime volume of interest, and as evident, it immediately follows that the action  $\mathcal{A}[\chi]$  vanishes if  $\chi^\alpha$  is a Killing vector field. It was shown in [12] that the AKE is strongly hyperbolic when the parameter  $\mu = 1$ ; however, it fails to be hyperbolic when  $\mu \neq 2$ , and is weakly hyperbolic for all other (constant) values for  $\mu$ . It was also argued that in general, the AKE may suffer from dynamical instabilities; a Hamiltonian analysis reveals the presence of ghosts for  $\mu < 2$ , and unbounded terms for  $\mu > 1/3$ . There is no parameter choice for  $\mu$  in which the AKE avoids these potential instabilities and is also hyperbolic. However, for vacuum spacetimes, at least one exception exists, which we will discuss shortly.

In any spacetime manifold, given some vector field  $V^\alpha$ , it follows from differential geometry that it is possible to construct a conserved current and hence a conserved charge. This conserved current takes the following form:

$$J^\alpha = \nabla_\beta (\nabla^\alpha V^\beta - \nabla^\beta V^\alpha). \quad (2.4)$$

For Killing vector fields, the conserved current  $J^\alpha$  is known as the Komar current, and the associated charges are known as the Komar charges, and we have so far referred to these, respectively, as the generalized Komar current and the generalized Komar charges. However, if the vector field  $V^\alpha$  is considered as a generator of the diffeomorphism, then  $J^\alpha$  is in fact the conserved current corresponding to the invariance under said diffeomorphism.<sup>2</sup> For this reason, it is perhaps more appropriate to call this the Noether current since it arises out of the diffeomorphism invariance of the gravitational system [15–19]. For the remainder of this article, we shall use this terminology and shall refer to the conserved charges associated with  $J^\alpha$  as Noether charges. As [20–22] demonstrate, Noether charges defined in this manner have interesting thermodynamical

<sup>1</sup>Alternatively, one can show that even on a Minkowskian background, the AKE admits wavelike solutions for which the components  $Q_{\mu\nu} \gg 0$ , so that they cannot be considered as approximate Killing vectors by any means.

<sup>2</sup>For further discussion of this point, see [13,14].

interpretations when computed over certain spacelike and null surfaces.

As emphasized before, the Noether current  $J^\alpha$  defined in Eq. (2.4) is identically divergence-free, which when evaluated for solutions of the AKE takes the following form:

$$J_\chi^\alpha = 2R^\alpha{}_\beta \chi^\beta + \nabla^\alpha [(2 - \mu)\nabla \cdot \chi]. \quad (2.5)$$

We note that when  $\mu = 2$ , the Noether current may be interpreted as a measure of the energy and momentum through the use of the trace-reversed Einstein field equations:

$$R_{\alpha\beta} = 8\pi \left( T_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}T \right). \quad (2.6)$$

Moreover, the divergence-free property of the Noether current yields the following expression:

$$\square[(\mu - 2)\nabla \cdot \chi] = \chi^\beta \nabla_\beta R + 2R^{\alpha\beta} \nabla_{(\alpha}\chi_{\beta)}, \quad (2.7)$$

where the contracted Bianchi identity  $\nabla_\alpha R^\alpha{}_\beta = (1/2)\nabla_\beta R$  has been used. The above evolution equation for  $(\nabla \cdot \chi)$  was used in [12] to show that in a vacuum spacetime, the constraint  $\nabla \cdot \chi = 0$  is propagated by the AKE; if the initial data satisfy the constraint  $\nabla \cdot \chi = 0$  and its time derivative, then the time development of the solution satisfies the constraint. Under this constraint, the AKE becomes strongly hyperbolic and is independent of  $\mu$ , so that it is no longer subject to the instabilities associated with ghosts or unbounded terms in the Hamiltonian.

## B. Example: The Vaidya geometry

Here, we review and generalize the solution for the AKE in the Vaidya spacetime, as presented in [10], to gain some insight into the relationship between GKVs, the Smarr relation, and the laws of thermodynamics. The line element associated with Vaidya spacetime takes the following form ( $d\Omega^2$  being the round metric on the 2-sphere):

$$ds^2 = - \left[ 1 - \frac{2M(v)}{r} \right] dv^2 + 2dv dr + r^2 d\Omega^2, \quad (2.8)$$

where  $M(v)$  is an arbitrary positive function of the advanced null coordinate  $v$ . Following [10], here also we construct a solution to the AKE in the Vaidya spacetime presented in Eq. (2.8) for the  $\mu = 2$  case, which has the following form:

$$\chi^\alpha = \left( \frac{M(v)}{M_0}, \frac{rM'(v) + f(v)}{M_0}, 0, 0 \right), \quad (2.9)$$

where  $f(v)$  is an arbitrary function of the advanced null coordinate  $v$ , and  $1/M_0$  is a constant factor; a natural choice for this factor is to set it equal to the ADM mass of

the spacetime. The fact that the GKV  $\chi^\alpha$  depends on an arbitrary function  $f(v)$  is due to the fact that for  $\mu = 2$ , the AKE fails to be hyperbolic. Fortunately, the resulting Noether current and Noether charge are unaffected by the arbitrary function  $f(v)$ , so that one can regard it as a sort of “gauge” potential. There is, however, a criterion that one may use to fix this arbitrary function  $f(v)$ , which will be discussed later. The Noether current associated with the GKV  $\chi^\alpha$  then takes the following form:

$$J_\chi^\alpha = \left( 0, \frac{4M(v)M'(v)}{M_0 r^2}, 0, 0 \right). \quad (2.10)$$

In the Vaidya spacetime, the surface characterized by

$$r_H := 2M(v) \quad (2.11)$$

is a special surface, since the expansion of the outgoing null generators vanishes on this surface and is the apparent horizon. Moreover, it is straightforward to demonstrate that  $\chi^\alpha = (1, 0, 0, 0)$  is null on the surface  $r = r_H$ . Thus, expressing the surface element as  $dS_{\alpha\beta} = \varepsilon_{\alpha\beta\mu\nu} e_\theta^\mu e_\phi^\nu d\theta d\phi$ , where  $e_\theta^\mu = \delta_\theta^\mu$  and  $e_\phi^\nu = \delta_\phi^\nu$  are the basis vector components on the apparent horizon, the mass within the apparent horizon is given by

$$M_H = \frac{1}{8\pi} \oint_H \nabla^\alpha \chi^\beta dS_{\alpha\beta}. \quad (2.12)$$

The above integrand turns out to be independent of the radius of the surface on which it is being evaluated, and thus, one obtains the following expression for the mass enclosed by the apparent horizon:

$$M_H = \frac{M(v)^2}{M_0}. \quad (2.13)$$

Intriguingly, the area of the apparent horizon is given by  $A_H = 4\pi r_H^2$  [with  $r_H$  given by Eq. (2.11)], so that one may rewrite the mass within the apparent horizon as presented in Eq. (2.13) as

$$M_H = \frac{A_H}{16\pi M_0} = \frac{\kappa_0 A_H}{4\pi}, \quad \kappa_0 := \frac{1}{4M_0}, \quad (2.14)$$

where  $\kappa_0$  is the surface gravity associated with the surface  $r = 2M_0$  corresponding to the event horizon of the final black hole spacetime. The fact that  $\kappa_0$  is a constant here is contradictory to the explicit formula  $\kappa^2 = -\frac{1}{2}(\nabla_\mu \chi_\nu)(\nabla^\mu \chi^\nu)$  for the surface gravity, since it depends on the gauge function  $f(v)$ . One may choose the gauge function, such that  $f(v) = -2M(v)M'(v)$ , which corresponds to the requirement that  $\chi$  is null on the apparent horizon, in which case, one obtains

$$\kappa = \frac{\sqrt{1 - 16M'(v)^2}}{4M_0}. \quad (2.15)$$

Even though it appears that the surface gravity is indeed dependent on the mass function, it is straightforward to verify that if the  $v$  dependence is treated as a perturbation, such that  $M(v) = M_0 + \varepsilon \delta M(v)$ , then the  $v$ -dependent part of  $\kappa$  is a second-order term in the perturbation

$$\kappa = \frac{1}{4M_0} - \frac{2\delta M'(v)^2}{4M_0} \varepsilon^2 + O(\varepsilon^3). \quad (2.16)$$

Therefore, it follows that to first order, the perturbation of the surface gravity identically vanishes, and the first law  $\delta M = \kappa \delta A / 8\pi$  holds identically. Even then, at first sight, Eq. (2.15) appears to be puzzling, as it seems to conflict with the expected behavior for the surface gravity, which must satisfy the exact Smarr relation  $M_H = \kappa A_H / 4\pi$  (neglecting angular momentum). However, upon closer inspection, one notes that since the AKE is linear in  $\chi$ , the constant factor  $1/M_0$  is not specified by the AKE. In an asymptotically flat spacetime, a natural choice for  $M_0$  is the ADM mass  $M_{\text{ADM}}$ , and hence, the mass enclosed by the apparent horizon may then be written as  $M_H = (M(v)/M_{\text{ADM}})M(v)$ . This suggests that  $M_H$  can be interpreted as a rescaling of  $M(v)$  by the ratio of  $M(v)$  to  $M_{\text{ADM}}$ .

To better understand this scenario, we assume that  $M'(v)$  has a compact support in  $v$ , such that at late time,  $M(v) \rightarrow M_0$  and  $f(v) \rightarrow 0$ , yielding  $M_0 \rightarrow M_{\text{ADM}}$ , in which case one has  $\chi^\alpha = (1, 0, 0, 0)$ . At early times, again assuming  $M'(v) \rightarrow 0$  and  $f(v) \rightarrow 0$ , one has  $\chi^\alpha = (M_{\text{early}}/M_{\text{ADM}}, 0, 0, 0)$ , where  $M_{\text{early}}$  is the mass of the spacetime before  $M'(v)$  becomes nonzero, and thus,  $\chi^\alpha$  will differ from  $(\partial/\partial t)^\alpha$ . Since the early-time geometry of the Vaidya spacetime approximates that of a Schwarzschild black hole, then it is appropriate to rescale  $\chi^\alpha$  by a factor  $M_{\text{ADM}}/M_{\text{early}}$ ; in doing so, one obtains an early-time horizon mass consistent with the early-time “Schwarzschild mass.” This is an indication that the horizon mass constructed from the Komar integral for solutions of the AKE is not identical to the “local” (in  $v$ ) mass of the black hole.

### III. EVOLUTION OF THE PERTURBED GKVs

In this section, we will consider the perturbation of a background spacetime with Killing symmetry (e.g., Schwarzschild or Kerr), which may or may not be vacuum. Since the perturbation need not respect the symmetry of the background spacetime, the perturbed spacetime does not admit an exact Killing vector field, but the perturbed spacetime will admit GKVs, as long as solutions to Eq. (2.2) exist in the perturbed spacetimes. The perturbed AKE associated with the perturbation of the background spacetime will be derived in two different ways: first from a variational principle, where the perturbation of the action

presented in Eq. (2.3) will be considered, and then from the direct perturbation of the AKE itself. We will verify that the results arising out of these two different approaches match.

### A. Notations and conventions

Before proceeding further, it is perhaps appropriate to settle the notations and the conventions that we will use for this section and the remainder of this article. With the exception of the GKVs, often denoted as  $\chi^\mu$ , “barred” symbols will be used to denote exact quantities; for instance  $\bar{g}_{\alpha\beta}$ ,  $\bar{\nabla}_\alpha$ , and  $\bar{R}_{\alpha\beta}$ , respectively, denote the exact metric, connection, and Ricci tensor. Unbarred geometric quantities will be used to denote background quantities. Metric perturbations will be denoted  $h_{\alpha\beta}$ , and are defined by

$$h_{\alpha\beta} := \bar{g}_{\alpha\beta} - g_{\alpha\beta}. \quad (3.1)$$

Indices are raised and lowered according to the background metric  $g_{\alpha\beta}$ . Killing vector fields for the background spacetime will be denoted  $\xi^\mu$ , and we define  $\delta\xi^\mu$  to be the difference between the GKV and the exact Killing vector in the following manner:

$$\delta\xi^\alpha := \chi^\alpha - \xi^\alpha. \quad (3.2)$$

In general, the prefix  $\delta$  will denote differences between the exact and background quantities (which we will later assume to be small compared to background values), and the prefix  $\Delta$  will denote first-order variations.

### B. Perturbation of the action principle yielding the almost Killing equation

As emphasized before, we assume that the background metric  $g_{\mu\nu}$  admits a Killing vector field  $\xi^\mu$ . Since  $h_{\mu\nu}$  is a perturbation over and above the background spacetime, it is legitimate to assume that  $h_{\mu\nu} \ll g_{\mu\nu}$ , and for some normalization of the background Killing vector field  $\xi^\mu$ , we also assume  $\delta\xi^\mu \ll \xi^\mu$ .<sup>3</sup> The action presented in Eq. (2.3) may then be expanded in  $h_{\mu\nu}$ ,  $\delta\xi^\mu$ , and  $\delta\mu := \bar{\mu} - \mu$ , keeping terms up to quadratic order in each. We treat the expansions in  $h_{\mu\nu}$  and  $\delta\xi^\mu$  independently, so that we keep terms of the form  $h..h.. \delta\xi^\mu \delta\xi^\nu$ . For simplicity, we assume the quantity  $\delta\mu$  to be independent of the spacetime coordinates. It should be emphasized that  $\delta\xi_\mu \neq \bar{g}_{\mu\nu}(\chi^\nu - \xi^\nu)$ , since indices are raised and lowered with the background metric. We begin by writing down the Lagrangian from the action functional presented in

<sup>3</sup>Technically speaking, this is achieved by introducing two parameters,  $\epsilon_1$  and  $\epsilon_2$ , with  $\epsilon_{1,2} \ll 1$  and then expanding the metric as  $\bar{g}_{\mu\nu} = g_{\mu\nu} + \epsilon_1 h_{\mu\nu}$  and the GKV as  $\chi^\alpha = \xi^\alpha + \epsilon_2 \delta\xi^\alpha$ . Then, keeping terms linear in  $\epsilon_1$  and  $\epsilon_2$  will provide first-order perturbation, while terms  $\mathcal{O}(\epsilon_{1,2}^2)$  yield the second-order perturbations.

Eq. (2.3), which has the following explicit form in the spacetime with metric  $\bar{g}_{\mu\nu}$ :

$$\begin{aligned} \bar{L} = & -\frac{1}{4}(\bar{g}_{\nu\alpha}\bar{\nabla}_\mu\chi^\alpha + \bar{g}_{\mu\alpha}\bar{\nabla}_\nu\chi^\alpha)(\bar{g}^{\mu\beta}\bar{\nabla}_\beta\chi^\nu + \bar{g}^{\nu\beta}\bar{\nabla}_\beta\chi^\mu) \\ & + \frac{\bar{\mu}}{2}(\bar{\nabla}_\mu\chi^\mu)^2. \end{aligned} \quad (3.3)$$

We wish to express the Lagrangian presented above solely in terms of the background metric  $g_{\mu\nu}$ , the perturbation  $h_{\mu\nu}$ , and the perturbation  $\delta\xi^\mu$  along with  $\delta\mu$ . Since the Lagrangian depicted in Eq. (3.3) consists of terms quadratic in  $\chi^\mu$  and its derivatives, it can be rewritten in the following form:

$$\bar{L} = \chi^\alpha\chi^\beta L_{\alpha\beta} + \chi^\mu\nabla^\alpha\chi^\beta L_{\mu\alpha\beta} + \nabla^\nu\chi^\mu\nabla^\beta\chi^\alpha L_{\mu\nu\alpha\beta}, \quad (3.4)$$

where the tensors  $L_{\alpha\beta}$ ,  $L_{\mu\alpha\beta}$ , and  $L_{\mu\nu\alpha\beta}$  depend on the background metric  $g_{\mu\nu}$ , the perturbation  $h_{\mu\nu}$ , the derivatives of  $h_{\mu\nu}$ , and  $\bar{\mu}$ . Note that in writing Eq. (3.4), we make no assumption about  $h_{\mu\nu}$ ; Eq. (3.4) should hold to all orders in the metric perturbation  $h_{\mu\nu}$ . Using the above decomposition of the Lagrangian, the action can similarly be written down in the following form:

$$\begin{aligned} \mathcal{A}[\delta\xi^\mu] = & \int_{\mathcal{V}} d^4x \sqrt{-g} [\chi^\alpha\chi^\beta M_{\alpha\beta} + \chi^\mu\nabla^\alpha\chi^\beta M_{\mu\alpha\beta} \\ & + \nabla^\nu\chi^\mu\nabla^\beta\chi^\alpha M_{\mu\nu\alpha\beta}], \end{aligned} \quad (3.5)$$

where we define

$$M_{\sigma_1\sigma_2\dots} := \sqrt{\bar{g}/g} L_{\sigma_1\sigma_2\dots}. \quad (3.6)$$

One may at this point perform the variation of the action with respect to  $\delta\xi^\alpha$ , the perturbed GKV, without explicit knowledge of the tensors  $M_{\sigma_1\sigma_2\dots}$ . Up to boundary terms, the first-order variation of the action takes the following form:

$$\Delta\mathcal{A} = \int d^4x \sqrt{-g} \Delta\delta\xi^\sigma \mathbb{E}_\sigma, \quad (3.7)$$

where

$$\begin{aligned} \mathbb{E}_\alpha = & \frac{1}{2} \left\{ 2\nabla^\beta\chi^\nu \left[ \frac{1}{2}(L_{\beta\nu\gamma\alpha} + L_{\gamma\alpha\beta\nu})(h^{\sigma\tau}\nabla^\gamma h_{\sigma\tau} - \nabla^\gamma h) \right. \right. \\ & \left. \left. - (\nabla^\gamma L_{\beta\nu\gamma\alpha} + \nabla^\gamma L_{\gamma\alpha\beta\nu}) + L_{\alpha\beta\mu} - L_{\nu\beta\alpha} \right] \right. \\ & + 2\chi^\nu \left[ \frac{1}{2} L_{\nu\beta\alpha}(h^{\gamma\sigma}\nabla^\beta h_{\gamma\sigma} - \nabla^\beta h) - \nabla^\beta L_{\nu\beta\alpha} \right. \\ & \left. \left. + L_{\alpha\nu} + L_{\nu\alpha} \right] - 2\nabla^\gamma\nabla^\beta\chi^\nu (L_{\beta\nu\gamma\alpha} + L_{\gamma\alpha\beta\nu}) \right\}. \end{aligned} \quad (3.8)$$

Keeping in mind  $\chi^\alpha = \xi^\alpha + \delta\xi^\alpha$ , the perturbed AKE may then be written as

$$\frac{\bar{g}^{\mu\alpha}\mathbb{E}_\alpha}{\sqrt{\bar{g}/g}} = 0. \quad (3.9)$$

The factor of  $1/\sqrt{\bar{g}/g}$  is included because one typically factors out the volume element from the functional derivative when deriving field equations [as was done when going from Eq. (A4) to Eq. (A5) in the Appendix]. Again, we emphasize that the analysis presented here does not require that  $h_{\mu\nu}$  be small; the result in Eq. (3.9) holds to all orders in  $h_{\mu\nu}$ . Thus, one may expand the tensors  $L_{\sigma_1\sigma_2\dots}$  in various powers of the gravitational perturbation in the following manner:

$$\begin{aligned} L_{\alpha\beta} &= L_{\alpha\beta}^0 + L_{\alpha\beta}^1 + L_{\alpha\beta}^2 + O(h^3), \\ L_{\mu\alpha\beta} &= L_{\mu\alpha\beta}^0 + L_{\mu\alpha\beta}^1 + L_{\mu\alpha\beta}^2 + O(h^3), \\ L_{\mu\nu\alpha\beta} &= L_{\mu\nu\alpha\beta}^0 + L_{\mu\nu\alpha\beta}^1 + L_{\mu\nu\alpha\beta}^2 + O(h^3). \end{aligned} \quad (3.10)$$

It turns out that to zeroth order, one has  $L_{\alpha\beta}^0 = 0$ ,  $L_{\mu\alpha\beta}^0 = 0$ , while

$$L_{\mu\nu\alpha\beta}^0 = \frac{1}{2}(\bar{\mu} g_{\alpha\beta}g_{\mu\nu} - g_{\alpha\nu}g_{\beta\mu} - g_{\alpha\mu}g_{\beta\nu}). \quad (3.11)$$

To first order in  $h_{\mu\nu}$ , the tensors  $L_{\alpha\beta}^1$ ,  $L_{\mu\alpha\beta}^1$ , and  $L_{\mu\nu\alpha\beta}^1$  can be expressed as linear functions of the gravitational perturbation  $h_{\mu\nu}$  as

$$\begin{aligned} L_{\alpha\beta}^1 &= 0, \quad L_{\mu\alpha\beta}^1 = \frac{1}{2}\{\bar{\mu}g_{\alpha\beta}\nabla_\mu h - 2\nabla_\mu h_{\alpha\beta}\}, \\ L_{\mu\nu\alpha\beta}^1 &= \frac{1}{2}\{h_{\alpha\mu}g_{\beta\nu} - g_{\alpha\mu}h_{\beta\nu}\}. \end{aligned} \quad (3.12)$$

Finally, we present the second-order terms in the perturbation  $h_{\mu\nu}$  as

$$\begin{aligned} L_{\alpha\beta}^2 &= \frac{1}{8}\{\bar{\mu}\nabla_\alpha h\nabla_\beta h - 2\nabla_\alpha h^{\sigma\tau}\nabla_\beta h_{\sigma\tau}\}, \\ L_{\mu\alpha\beta}^2 &= \frac{1}{2}\{2h_\alpha{}^\sigma\nabla_\mu h_{\beta\sigma} - \bar{\mu}g_{\alpha\beta}h^{\sigma\tau}\nabla_\mu h_{\sigma\tau}\}, \\ L_{\mu\nu\alpha\beta}^2 &= \frac{1}{2}\{h_{\alpha\mu}h_{\beta\nu} - g_{\beta\nu}h_\alpha{}^\sigma h_{\mu\sigma}\}. \end{aligned} \quad (3.13)$$

In what follows, we consider in detail the expansion of the Lagrangian to first order in the metric perturbation  $h_{\mu\nu}$ , and as we shall demonstrate, the resulting perturbed AKE is consistent with the expansion coefficients determined above.

### C. Explicit perturbation of the action to first order in the metric

The expression for the AKE given in Eq. (3.8) is rather complicated and somewhat opaque; it is perhaps more illustrative to show explicitly that the derivation of the perturbed AKE to first order in the metric perturbations  $h_{\mu\nu}$ , simplifying the expressions along the way. To obtain the expansion of the Lagrangian to first order in  $h_{\mu\nu}$ , one may use Eqs. (3.11) and (3.12) to obtain the following Lagrangian:

$$\begin{aligned} \bar{L} &= -\frac{1}{4}[(g_{\nu\alpha} + h_{\nu\alpha})\nabla_\mu\chi^\alpha + (g_{\mu\alpha} + h_{\mu\alpha})\nabla_\nu\chi^\alpha + \chi^\rho\nabla_\rho h_{\mu\nu}] \\ &\quad \times [(g^{\mu\beta} - h^{\mu\beta})\nabla_\beta\chi^\nu + (g^{\nu\beta} - h^{\nu\beta})\nabla_\beta\chi^\mu + \chi^\sigma\nabla_\sigma h^{\mu\nu}] \\ &\quad + \frac{\bar{\mu}}{2}[(\nabla_\mu\chi^\mu)^2 + \chi^\alpha\nabla_\alpha h(\nabla_\mu\chi^\mu)]. \end{aligned} \quad (3.14)$$

Having expressed the Lagrangian explicitly in terms of the background metric  $g_{\mu\nu}$  and the perturbation  $h_{\alpha\beta}$ , we now expand the GKV field in terms of the background Killing field  $\xi^\mu$  and the perturbation  $\delta\xi^\mu$ . One can see that Eq. (3.14) has the form of Eq. (3.4), and it is not difficult to verify that the Lagrangians are equivalent for the coefficients given in Eqs. (3.10)–(3.12). Using the Killing equation for  $\xi^\mu$ , i.e., setting  $\nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu = 0$ , the Lagrangian presented in Eq. (3.14) can be further simplified. In particular, it is worth emphasizing that the on-shell value of the action for the background Killing vector field identically vanishes, and thus, the Lagrangian density given in Eq. (3.14) becomes

$$\begin{aligned} \bar{L} &= -\frac{1}{2}[g_{\nu\alpha}\nabla_\mu\delta\xi^\alpha + h_{\nu\alpha}\nabla_\mu\xi^\alpha + h_{\nu\alpha}\nabla_\mu\delta\xi^\alpha + g_{\mu\alpha}\nabla_\nu\delta\xi^\alpha \\ &\quad + h_{\mu\alpha}\nabla_\nu\xi^\alpha + h_{\mu\alpha}\nabla_\nu\delta\xi^\alpha + (\xi^\rho + \delta\xi^\rho)\nabla_\rho h_{\mu\nu}] \\ &\quad \times \left[ \nabla^\mu\delta\xi^\nu - h^{\mu\beta}\nabla_\beta\xi^\nu - h^{\mu\beta}\nabla_\beta\delta\xi^\nu + \frac{1}{2}(\xi^\sigma + \delta\xi^\sigma)\nabla_\sigma h^{\mu\nu} \right] \\ &\quad + \frac{\bar{\mu}}{2}[(\nabla_\mu\delta\xi^\mu)^2 + (\xi^\alpha + \delta\xi^\alpha)\nabla_\alpha h(\nabla_\mu\delta\xi^\mu)] + O(h^2). \end{aligned} \quad (3.15)$$

Even though the above Lagrangian density looks sufficiently complicated, we can reduce it to a very simple form by dividing the above into three categories: (a) terms quadratic in the derivatives of  $\delta\xi^\alpha$ , (b) terms linear in the derivatives of  $\delta\xi^\alpha$ , and (c) terms independent of derivatives of  $\delta\xi^\alpha$ . The terms quadratic in the derivative of  $\delta\xi^\alpha$  yield

$$\begin{aligned} \text{Quadratic terms} &= \frac{1}{2}[(\mu + \delta\mu)\delta_\alpha^\mu\delta_\beta^\nu - (g_{\alpha\beta} + h_{\alpha\beta})(g^{\mu\nu} - h^{\mu\nu}) \\ &\quad - (\delta_\alpha^\nu + h_\alpha^\nu)(\delta_\beta^\mu - h_\beta^\mu)](\nabla_\mu\delta\xi^\alpha\nabla_\nu\delta\xi^\beta), \end{aligned} \quad (3.16)$$

while the terms linear in the derivative of  $\delta\xi^\alpha$  become

$$\begin{aligned}
 \text{Linear terms} = & \frac{\bar{\mu}}{2} [(\xi^\alpha + \delta\xi^\alpha) \nabla_\alpha h(\nabla_\mu \delta\xi^\mu)] - \frac{1}{2} [g_{\nu\alpha} \nabla_\mu \delta\xi^\alpha + g_{\mu\alpha} \nabla_\nu \delta\xi^\alpha] \times \left[ -h^{\mu\beta} \nabla_\beta \xi^\nu + \frac{1}{2} (\xi^\sigma + \delta\xi^\sigma) \nabla_\sigma h^{\mu\nu} \right] \\
 & - \frac{1}{2} \nabla^\mu \delta\xi^\nu [h_{\nu\alpha} \nabla_\mu \xi^\alpha + h_{\mu\alpha} \nabla_\nu \xi^\alpha + (\xi^\rho + \delta\xi^\rho) \nabla_\rho h_{\mu\nu}].
 \end{aligned} \tag{3.17}$$

Finally, as indicated in the expression for  $L_{\alpha\beta}$  in Eq. (3.12) the terms involving no derivatives of  $\delta\xi^\alpha$  is  $\mathcal{O}(h^2)$ , and hence, will not contribute in our subsequent discussion regarding the determination of the action functional of the perturbed AKE. Thus, we have computed the Lagrangian of the GKV field involving linear order terms in the perturbation  $h_{\mu\nu}$  and up to quadratic order terms in the perturbed GKV  $\delta\xi^\mu$ . However, computation of the action functional requires multiplication of the above Lagrangian by a factor of  $\sqrt{-\bar{g}}$ , where  $\bar{g}$  is the determinant of the perturbed metric  $\bar{g}_{\alpha\beta}$ . Therefore, the complete action for the perturbed Killing vector field  $\delta\xi^\mu$  takes the following form:

$$\begin{aligned}
 \mathcal{A}[\delta\xi] = & \int_{\mathcal{V}} d^4x \sqrt{-\bar{g}} \left\{ \left[ \frac{1}{2} (\mu \delta_\alpha^\mu \delta_\beta^\nu - g_{\alpha\beta} g^{\mu\nu} - \delta_\alpha^\nu \delta_\beta^\mu) + \frac{1}{4} h (\mu \delta_\alpha^\mu \delta_\beta^\nu - g_{\alpha\beta} g^{\mu\nu} - \delta_\alpha^\nu \delta_\beta^\mu) \right. \right. \\
 & + \frac{1}{2} \left( 1 + \frac{1}{2} h \right) \delta\mu \delta_\alpha^\mu \delta_\beta^\nu + \frac{1}{2} (g_{\alpha\beta} h^{\mu\nu} - h_{\alpha\beta} g^{\mu\nu} + \delta_\alpha^\nu h_\beta^\mu - h_\alpha^\nu \delta_\beta^\mu) \left. \right] (\nabla_\mu \delta\xi^\alpha \nabla_\nu \delta\xi^\beta) \\
 & + \left[ \frac{\bar{\mu}}{2} [(\xi^\alpha + \delta\xi^\alpha) \nabla_\alpha h(\nabla_\mu \delta\xi^\mu)] - [(\nabla_\mu \delta\xi^\alpha) h_{\alpha\beta} \nabla^\mu \xi^\beta + (\nabla_\mu \delta\xi^\alpha) h^{\mu\beta} \nabla_\alpha \xi_\beta] \right. \\
 & \left. \left. - g_{\nu\alpha} \nabla_\mu \delta\xi^\alpha (\xi^\sigma + \delta\xi^\sigma) \nabla_\sigma h^{\mu\nu} \right] \right\}.
 \end{aligned} \tag{3.18}$$

Having derived the action to linear order in the gravitational perturbation  $h_{\mu\nu}$ , we can determine an arbitrary variation of the action for variation of the perturbed Killing vector field  $\delta\xi^\mu$ , which when set to zero should yield the corresponding perturbed AKE. The final expression for the variation, ignoring any boundary contribution, takes the following form:

$$\begin{aligned}
 \Delta\mathcal{A} = & \int_{\mathcal{V}} d^4x \sqrt{-g} \delta(\delta\xi^\alpha) \left( 1 + \frac{1}{2} h \right) (g_{\alpha\beta} + h_{\alpha\beta}) \left\{ [(1 - \mu) \nabla^\beta (\nabla_\nu \delta\xi^\nu) + \square \delta\xi^\beta + R^\beta{}_\rho \delta\xi^\rho] \right. \\
 & - \delta\mu \nabla^\beta \nabla_\sigma \delta\xi^\sigma - h^{\mu\nu} \nabla_\mu \nabla_\nu \delta\xi^\beta + 2(\nabla^{(\mu} \delta\xi^{\nu)}) \left( \nabla_\mu h_\nu^\beta - \frac{1}{2} \nabla^\beta h_{\mu\nu} \right) - R^\beta{}_{\rho\sigma\mu} h^{\mu\rho} \delta\xi^\sigma \\
 & \left. + \left( \frac{1 - \mu - \delta\mu}{2} \right) \nabla^\beta [(\xi^\rho + \delta\xi^\rho) \nabla_\rho h] - h^{\beta\mu} (1 - \mu - \delta\mu) \nabla_\mu (\nabla_\nu \delta\xi^\nu) \right\},
 \end{aligned} \tag{3.19}$$

where we have neglected all the terms quadratic in the gravitational perturbation  $h_{\mu\nu}$ . Setting the variation  $\Delta\mathcal{A}$  to zero, for arbitrary variation of the perturbation of the Killing vector field  $\delta\xi^\mu$ , we obtain the following dynamical equation for the perturbed GKV field  $\delta\xi^\mu$ :

$$(1 - \mu) \nabla^\beta (\nabla_\nu \delta\xi^\nu) + \square \delta\xi^\beta + R^\beta{}_\rho \delta\xi^\rho = J^\beta, \tag{3.20}$$

$$\begin{aligned}
 J^\beta = & \delta\mu \nabla^\beta \nabla_\sigma \delta\xi^\sigma + h^{\mu\nu} \nabla_\mu \nabla_\nu \delta\xi^\beta - 2(\nabla^{(\mu} \delta\xi^{\nu)}) \left( \nabla_\mu h_\nu^\beta - \frac{1}{2} \nabla^\beta h_{\mu\nu} \right) + R^\beta{}_{\rho\sigma\mu} h^{\mu\rho} \delta\xi^\sigma \\
 & - \left( \frac{1 - \mu - \delta\mu}{2} \right) \nabla^\beta [(\xi^\rho + \delta\xi^\rho) \nabla_\rho h] + h^{\beta\mu} (1 - \mu - \delta\mu) \nabla_\mu (\nabla_\nu \delta\xi^\nu).
 \end{aligned} \tag{3.21}$$

The above provides the dynamical equation for the perturbed Killing vector field  $\delta\xi^\alpha$  arising from the variation of the action. One can verify that this expression is equivalent to that obtained from Eq. (3.8); we have done this using the `xAct` package for *Mathematica*. In the subsequent discussion, we will discuss explicitly the derivation of this equation from the perturbation of the AKE itself to first

order in the metric perturbations. This will depict the internal consistency of the results derived in this work.

#### D. Perturbation of the almost Killing equation to first order in the metric

We have derived the evolution equation for the perturbed GKV field to first order in the metric perturbations, starting

from the variation of the perturbed action for the GKV field. As we will show in this section, the same equation can also be derived from direct perturbation of the AKE itself. As before, we assume that the background metric  $g_{\mu\nu}$  admits a Killing vector field  $\xi^\mu$  and also that  $h_{\mu\nu} \ll g_{\mu\nu}$  and  $\delta\xi^\mu \ll \xi^\mu$  (see also footnote 2). Thus, we will expand the AKE given in Eq. (2.2) in the perturbed spacetime with metric  $\bar{g}_{\mu\nu}$  to first order in  $\delta\xi^\mu$  and  $h_{\mu\nu}$  each, again assuming that the expansions are independent (so that we keep terms of the form  $h_{\mu\nu}\delta\xi^\mu$ ). It is convenient to first present the expansion of the following geometric quantities to linear order in the gravitational perturbation  $h_{\mu\nu}$ :

$$\delta R_{\mu\nu} = \frac{1}{2}(-\square h_{\mu\nu} - \nabla_\mu \nabla_\nu h + \nabla_\mu \nabla_\alpha h_\nu^\alpha + \nabla_\nu \nabla_\alpha h_\mu^\alpha + R_{\beta\mu} h_\nu^\beta + R_{\beta\nu} h_\mu^\beta - 2R_{\alpha\mu\beta\nu} h^{\alpha\beta}), \quad (3.22)$$

$$\bar{\nabla}_\alpha \bar{\nabla}_\beta V^\beta = \delta_\mu^\beta \bar{\nabla}_\alpha \bar{\nabla}_\beta V^\mu = \nabla_\alpha (\nabla_\beta V^\beta) + \nabla_\alpha (\delta\Gamma_{\beta\rho}^\beta V^\rho) + \delta\Gamma_{\alpha\rho}^\beta \nabla_\beta V^\rho - \delta\Gamma_{\alpha\beta}^\rho \nabla_\rho V^\beta = \nabla_\alpha (\nabla_\beta V^\beta) + \nabla_\alpha (\delta\Gamma_{\beta\rho}^\beta V^\rho), \quad (3.23)$$

$$\begin{aligned} \bar{\square} V^\mu &= (g^{\alpha\beta} - h^{\alpha\beta}) \nabla_\alpha \nabla_\beta V^\mu + \left( -\frac{1}{2} \nabla^\mu h_{\alpha\rho} + \nabla_\alpha h_\rho^\mu \right) (\nabla^\alpha V^\rho + \nabla^\rho V^\alpha) - \nabla_\rho V^\mu \left( \nabla_\alpha h^{\alpha\rho} - \frac{1}{2} \nabla^\rho h \right) \\ &+ \frac{1}{2} V^\rho (\square h_\rho^\mu + \nabla_\rho \nabla_\beta h^{\beta\mu} - \nabla^\mu \nabla^\beta h_{\beta\rho} + R_{\sigma\rho} h^{\sigma\mu} - R^{\sigma\mu} h_{\sigma\rho}), \end{aligned} \quad (3.24)$$

where  $V^\mu$  is an arbitrary vector field. In deriving the above identities, we have used various properties of the Riemann tensor, e.g.,  $R_{\alpha\beta\mu\nu} = R_{\nu\mu\beta\alpha}$  among others. Applying all these identities to the AKE in the perturbed spacetime and imposing the Lorenz gauge condition  $\nabla_\alpha h_\rho^\alpha = (1/2)\nabla_\rho h$ , we obtain

$$\begin{aligned} (g^{\alpha\beta} - h^{\alpha\beta}) \nabla_\alpha \nabla_\beta \chi^\mu + \left( -\frac{1}{2} \nabla^\mu h_{\alpha\rho} + \nabla_\alpha h_\rho^\mu \right) (\nabla^\alpha \chi^\rho + \nabla^\rho \chi^\alpha) + R^\mu{}_{\beta\chi} \chi^\beta + (1 - \bar{\mu}) \bar{g}^{\mu\alpha} \left[ \nabla_\alpha (\nabla_\beta \chi^\beta) + \frac{1}{2} \nabla_\alpha (\chi^\rho \nabla_\rho h) \right] \\ - R^\mu{}_{\alpha\beta\sigma} h^{\alpha\sigma} \chi^\beta + \{ \nabla_\alpha (1 - \bar{\mu}) \} \bar{g}^{\mu\alpha} \left[ (\nabla_\beta \chi^\beta) + \frac{1}{2} (\chi^\rho \nabla_\rho h) \right] = 0, \end{aligned} \quad (3.25)$$

which is valid up to linear order in the gravitational perturbation  $h_{\mu\nu}$ .

At this point, we have not yet expanded in the GKV field  $\chi^\mu$ ; we do this now. We make use of the wave equation for the background Killing vector field  $\xi^\mu$  given in Eq. (2.1) and other properties of Killing vectors to obtain

$$\begin{aligned} (g^{\alpha\beta} - h^{\alpha\beta}) \nabla_\alpha \nabla_\beta \delta\xi^\mu + \left( -\frac{1}{2} \nabla^\mu h_{\alpha\rho} + \nabla_\alpha h_\rho^\mu \right) (\nabla^\alpha \delta\xi^\rho + \nabla^\rho \delta\xi^\alpha) + R^\mu{}_{\beta\delta\xi} \delta\xi^\beta + (1 - \bar{\mu}) \bar{g}^{\mu\alpha} \left\{ \nabla_\alpha (\nabla_\beta \delta\xi^\beta) \right. \\ \left. + \frac{1}{2} \nabla_\alpha [(\xi^\rho + \delta\xi^\rho) \nabla_\rho h] \right\} - R^\mu{}_{\alpha\beta\sigma} h^{\alpha\sigma} \delta\xi^\beta + \{ \nabla_\alpha (1 - \bar{\mu}) \} \bar{g}^{\mu\alpha} \left[ (\nabla_\beta \delta\xi^\beta) + \frac{1}{2} (\xi^\rho \nabla_\rho h + \delta\xi^\rho \nabla_\rho h) \right] = 0. \end{aligned} \quad (3.26)$$

This is our result for the perturbed AKE. The above equation has been derived under very general conditions, without any assumptions about the nature of the perturbation. Thus, it is possible to express the above equation in several different ways, under different assumptions, which we will list below. First, we rewrite the above evolution equation for the perturbed GKV field  $\delta\xi^\mu$  in the following form:

$$\square \delta\xi^\mu + R^\mu{}_{\nu} \delta\xi^\nu + g^{\mu\alpha} \nabla_\alpha \{ (1 - \mu) (\nabla_\sigma \delta\xi^\sigma) \} = j^\mu, \quad (3.27)$$

$$\begin{aligned} j^\mu &= h^{\alpha\beta} [\nabla_\alpha \nabla_\beta \delta\xi^\mu + R^\mu{}_{\alpha\beta\sigma} \delta\xi^\sigma + \delta_\alpha^\mu \nabla_\beta \{ (1 - \mu - \delta\mu) (\nabla_\sigma \delta\xi^\sigma) \}] + g^{\mu\alpha} \nabla_\alpha (\delta\mu \nabla_\sigma \delta\xi^\sigma) - 2 \left( \nabla_\alpha h_\rho^\mu - \frac{1}{2} \nabla^\mu h_{\alpha\rho} \right) \nabla^{(\alpha} \delta\xi^{\rho)} \\ &- \frac{1}{2} (1 - \mu - \delta\mu) g^{\mu\alpha} \nabla_\alpha [(\xi^\rho + \delta\xi^\rho) \nabla_\rho h] - \frac{1}{2} \nabla_\alpha (1 - \mu - \delta\mu) g^{\mu\alpha} [(\xi^\rho + \delta\xi^\rho) \nabla_\rho h]. \end{aligned} \quad (3.28)$$

Upon comparison, we find that this evolution equation for the perturbed GKV  $\delta\xi^\mu$  is identical to what we had derived from the action, i.e., to Eq. (3.20), except for the terms involving derivatives of  $\mu$  and  $\delta\mu$ , respectively. This is because, while deriving Eq. (3.20), we have assumed for convenience that  $\mu$  and  $\delta\mu$  are constants, while that is not the case for the derivation presented above. If we assume that  $\mu$  for the background spacetime is constant, and  $\delta\mu$  to be a scalar function, then the dynamics of the perturbed GKV is determined by



$$(g^{\alpha\beta} - h^{\alpha\beta})[\nabla_\alpha \nabla_\beta \delta \xi^\mu + R^\mu_{\alpha\beta} \delta \xi^\rho] = j^\mu, \quad (3.29)$$

$$j^\mu = -(1 - \mu - \delta\mu) \left[ (g^{\mu\alpha} - h^{\mu\alpha}) \nabla_\alpha (\nabla_\sigma \delta \xi^\sigma) + \frac{1}{2} g^{\mu\alpha} \nabla_\alpha \{ (\xi^\rho + \delta \xi^\rho) \nabla_\rho h \} \right] - 2 \left( \nabla_\alpha h^\mu_\rho - \frac{1}{2} \nabla^\mu h_{\alpha\rho} \right) \nabla^{(\alpha} \delta \xi^{\rho)}$$

$$+ (\nabla_\alpha \delta\mu) \left\{ g^{\mu\alpha} \nabla_\alpha (\nabla_\sigma \delta \xi^\sigma) + \frac{1}{2} g^{\mu\alpha} [(\xi^\rho + \delta \xi^\rho) \nabla_\rho h] - (\nabla_\sigma \delta \xi^\sigma) h^{\mu\alpha} \right\}. \quad (3.30)$$

As mentioned before, it will be useful if we write down simplified versions of Eq. (3.29) for various scenarios of physical interest. These can range from the use of the transverse-traceless gauge to setting  $\delta\mu = \text{constant}$ . We discuss below each of these limits explicitly.

- (i) If we choose  $\delta\mu = \text{constant}$ , then the last term in the expression for  $j^\mu$  in Eq. (3.29) identically vanishes. In this context, the dynamical equation for the perturbed GKV becomes identical to that derived from the perturbation of the action, i.e., to Eq. (3.20).
- (ii) If we assume that the background spacetime is vacuum, and the perturbation involves no incoming matter fields, then the use of the transverse-traceless gauge [equivalently setting  $h = 0$  in Eq. (3.29)] yields

$$(g^{\alpha\beta} - h^{\alpha\beta})[\nabla_\alpha \nabla_\beta \delta \xi^\mu + R^\mu_{\alpha\beta} \delta \xi^\rho] = j^\mu, \quad (3.31)$$

$$j^\mu = -(1 - \mu - \delta\mu) [(g^{\mu\alpha} - h^{\mu\alpha}) \nabla_\alpha (\nabla_\sigma \delta \xi^\sigma)]$$

$$- 2 \left( \nabla_\alpha h^\mu_\rho - \frac{1}{2} \nabla^\mu h_{\alpha\rho} \right) \nabla^{(\alpha} \delta \xi^{\rho)}$$

$$+ (\nabla_\alpha \delta\mu) \{ g^{\mu\alpha} \nabla_\alpha (\nabla_\sigma \delta \xi^\sigma) - (\nabla_\sigma \delta \xi^\sigma) h^{\mu\alpha} \}. \quad (3.32)$$

Note that any term involving  $R_{\alpha\beta}$  will not contribute, since for vacuum spacetime the Ricci tensor identically vanishes. Also, if we assume  $\delta\mu$  to be constant, the above equation can be simplified even more, as the last term in  $j^\mu$  will be absent.

- (iii) If we use the fact that the perturbed GKV is really a consequence of the perturbation of the spacetime geometry (we will examine this case in detail later), then we will have  $\chi^\mu = \xi^\mu + \delta\chi^\mu_1 + \delta\chi^\mu_2$ , where  $\delta\chi^\mu_1$  is linear in the gravitational perturbation, while  $\delta\chi^\mu_2$  is quadratic in the gravitational perturbation. An identical decomposition will work for the  $\bar{\mu}$  as well. If we keep terms linear in the gravitational perturbation, we should also ignore terms  $\mathcal{O}(\delta\chi^\mu_1 h_{\alpha\beta})$  and so on. It follows that perturbed AKE governing the evolution of the vector  $\delta\chi^\mu_1$  takes the relatively simple form

$$\square \delta\chi^\mu_1 + R^\mu_\rho \delta\chi^\rho_1 + (1 - \mu) \left[ \nabla^\mu (\nabla_\sigma \delta\chi^\sigma_1) + \frac{1}{2} \nabla^\mu (\xi^\rho \nabla_\rho h) \right] = 0. \quad (3.33)$$

Note also that if we assume the background spacetime to be vacuum with no incoming matter perturbation, then the use of the transverse-traceless gauge would reduce Eq. (3.33) to AKE for the background spacetime  $g_{\mu\nu}$ . The consequences of this equation with or without matter field will be discussed in a subsequent section.

Thus, we have derived the evolution equation for the perturbed GKV field from the perturbed action and also from the perturbed AKE to first order in the metric perturbations. Both of these procedures yield identical equations depicting the internal consistency of our analysis. We have also verified using computer algebra (in particular, the `xAct` package for *Mathematica*) that this consistency holds to second order as well. The AKE is rather complicated in the second-order case, so we do not present the result here; the interested reader can view the *Mathematica* file posted at [23]. In what follows, we will discuss the structure of the Hamiltonian associated with the dynamical equation for the perturbed AKE, leading to an understanding of the stability as well as its hyperbolicity.

#### IV. HAMILTONIAN STABILITY AND HYPERBOLICITY

In this section, we will construct the Hamiltonian out of the Lagrangian, whose variation yields the evolution equation for the perturbed GKV. The stability of the Hamiltonian and its bounded nature will also be examined. In addition, the hyperbolicity of the perturbed AKE will also be explored.

##### A. Hamiltonian for the perturbed AKE

We have derived the evolution equation for the perturbed GKVs in the preceding section in two different ways: first by varying the perturbed action from which the AKE can be derived, and then by direct perturbation of the AKE. In this section, we will discuss the stability and the hyperbolicity of the perturbed AKE, restricting (for simplicity) to first order in the metric perturbations  $h_{\mu\nu}$ . First, we will

construct the Hamiltonian for the perturbed GKV field and discuss its stability. Subsequently, we will discuss the hyperbolicity of the perturbed AKE and its consequences. The starting point for the Hamiltonian analysis is the action for the perturbed GKV field  $\delta\xi^\mu$ ; in particular, the zeroth component of the boundary term in the variation will provide the Hamiltonian. The action for the perturbed GKV simplifies considerably to first order in the metric perturbations. Neglecting all terms of  $\mathcal{O}(h^2)$  and using symmetry properties of the resulting expression, the structure of the action can be simplified to

$$\begin{aligned} \mathcal{A}[\delta\xi] = & \int_{\mathcal{V}} d^4x \sqrt{-g} \left[ \frac{1}{2} \left( 1 + \frac{1}{2}h \right) \{ \bar{\mu} \delta_\alpha^\mu \delta_\beta^\nu - g_{\alpha\beta} g^{\mu\nu} - h_{\alpha\beta} g^{\mu\nu} + g_{\alpha\beta} h^{\mu\nu} - \delta_\alpha^\nu \delta_\beta^\mu \} (\nabla_\mu \delta\xi^\alpha \nabla_\nu \delta\xi^\beta) \right. \\ & \left. + \frac{\bar{\mu}}{2} \{ (\xi^\alpha + \delta\xi^\alpha) \nabla_\alpha h (\nabla_\mu \delta\xi^\mu) \} - (\nabla_\mu \delta\xi^\alpha) \{ h_{\alpha\beta} \nabla^\mu \xi^\beta + h^{\mu\beta} \nabla_\alpha \xi_\beta \} - g_{\nu\alpha} \nabla_\mu \delta\xi^\alpha (\xi^\sigma + \delta\xi^\sigma) \nabla_\sigma h^{\mu\nu} \right]. \end{aligned} \quad (4.1)$$

Note that the Lagrangian density associated with the above action is identical to Eq. (3.15), though written in a different form. Collecting all the total derivative terms that we have thrown away in the derivation of the field equation for  $\delta\xi^\mu$  in the previous section, we obtain (recalling the notation  $\Delta$  for the first-order variation)

$$\begin{aligned} \text{Total derivative terms} = & \sqrt{-g} \Delta(\delta\xi^\alpha) \left[ (\mu \delta_\alpha^\mu \delta_\beta^\nu - g_{\alpha\beta} g^{\mu\nu} - \delta_\alpha^\nu \delta_\beta^\mu) \nabla_\nu \delta\xi^\beta + \frac{1}{2} h (\mu \delta_\alpha^\mu \delta_\beta^\nu - g_{\alpha\beta} g^{\mu\nu} - \delta_\alpha^\nu \delta_\beta^\mu) \nabla_\nu \delta\xi^\beta \right. \\ & + \left( 1 + \frac{1}{2}h \right) \delta\mu \delta_\alpha^\mu \delta_\beta^\nu \nabla_\nu \delta\xi^\beta + (g_{\alpha\beta} h^{\mu\nu} - h_{\alpha\beta} g^{\mu\nu}) \nabla_\nu \delta\xi^\beta + \frac{\bar{\mu}}{2} \{ (\xi^\sigma + \delta\xi^\sigma) \nabla_\sigma h \} \delta_\alpha^\mu \\ & \left. - \{ h_{\alpha\beta} \nabla^\mu \xi^\beta + h^{\mu\beta} \nabla_\alpha \xi_\beta \} - g_{\nu\alpha} (\xi^\sigma + \delta\xi^\sigma) \nabla_\sigma h^{\mu\nu} \right] \equiv \sqrt{-g} \Delta(\delta\xi^\alpha) P^\mu{}_\alpha, \end{aligned} \quad (4.2)$$

where the last equality defines the quantity  $P^\mu{}_\alpha$ , which is the polymomentum conjugate to the perturbed GKV field  $\delta\xi^\alpha$ . In particular, starting from the perturbed action presented in Eq. (4.1), one can immediately verify that  $P^\mu{}_\alpha$  has the following expression in terms of the perturbed Killing vector field  $\delta\xi^\alpha$  and its derivatives:

$$\begin{aligned} P^\mu{}_\alpha & \equiv \frac{1}{\sqrt{-g}} \frac{\Delta \mathcal{A}}{\Delta(\nabla_\mu \delta\xi^\alpha)} \\ & = \left[ \left( 1 + \frac{1}{2}h \right) (\bar{\mu} \delta_\alpha^\mu \delta_\beta^\nu - g_{\alpha\beta} g^{\mu\nu} - \delta_\alpha^\nu \delta_\beta^\mu + g_{\alpha\beta} h^{\mu\nu} - h_{\alpha\beta} g^{\mu\nu}) \nabla_\nu \delta\xi^\beta \right. \\ & \left. + \frac{\bar{\mu}}{2} \{ (\xi^\sigma + \delta\xi^\sigma) \nabla_\sigma h \} \delta_\alpha^\mu - \{ h_{\alpha\beta} \nabla^\mu \xi^\beta + h^{\mu\beta} \nabla_\alpha \xi_\beta \} - g_{\nu\alpha} (\xi^\sigma + \delta\xi^\sigma) \nabla_\sigma h^{\mu\nu} \right]. \end{aligned} \quad (4.3)$$

We write the variation of the action for the perturbed Killing vector field  $\delta\xi^\alpha$  (incorporating the variation of the boundary surface) in the Weiss form [24]:

$$\begin{aligned} \Delta \mathcal{A} = & \int_{\mathcal{V}} d^4x \sqrt{-g} E_\alpha \Delta(\delta\xi^\alpha) \\ & + \int_{\partial\mathcal{V}} d\Sigma_\mu [P^\mu{}_\alpha \Delta(\delta\xi^\alpha) + L \Delta x^\mu], \end{aligned} \quad (4.4)$$

where  $d\Sigma_\alpha = d^3x \sqrt{-g} \nabla_\alpha \phi$  is the volume measure on the boundary surface  $\partial\mathcal{V}$  denoting the  $\phi(x^\mu) = \text{constant}$  hypersurface. If we instead use the unit normal vector  $n_\alpha$ , then the volume measure of the boundary hypersurface  $\partial\mathcal{V}$  becomes  $d\Sigma_\alpha = \epsilon d^3x \sqrt{h} n_\alpha$ , where  $\epsilon = -1(+1)$  for spacelike (time-like) hypersurfaces, respectively. Defining  $d\Sigma = d^3x \sqrt{h}$  and choosing the hypersurface to be  $t = \text{constant}$  and using the (1+3) decomposition for the metric, we find the boundary Hamiltonian to be

$$H = \int d\Sigma \mathcal{H}, \quad \mathcal{H} = \epsilon P^\mu{}_\alpha n_\mu \dot{\delta\xi}^\alpha - NL, \quad (4.5)$$

where  $N$  is the lapse function and the ‘‘overdot’’ denotes derivative with respect to time. For the spacelike hypersurface we are considering, and using orthonormal coordinates, e.g., in the synchronous frame, the Hamiltonian density  $\mathcal{H}$  takes the following form:

$$\mathcal{H} = P^0{}_\alpha \dot{\delta\xi}^\alpha - \mathcal{L}. \quad (4.6)$$

We now consider terms which are quadratic in the time derivative as well as terms which are quadratic in the spatial derivative of the perturbed GKV field  $\delta\xi^\alpha$ . Collecting these terms, the time derivative part of the Hamiltonian density quadratic in the perturbed Killing vector field becomes

$$\begin{aligned}
 \mathcal{H}^{(2)\text{time}} &= \frac{1}{2} \left( 1 + \frac{1}{2} h \right) \{ \bar{\mu} - 2 \} (\dot{\delta\xi}^0 \dot{\delta\xi}^0) \\
 &+ \frac{1}{2} \left( 1 + \frac{1}{2} h \right) \{ g_{ij} + h_{ij} + g_{ij} h^{00} \} (\dot{\delta\xi}^i \dot{\delta\xi}^j) \\
 &+ \left( 1 + \frac{1}{2} h \right) h_{0i} (\dot{\delta\xi}^0 \dot{\delta\xi}^i). \tag{4.7}
 \end{aligned}$$

Here we have performed a (1 + 3) decomposition of the Hamiltonian density and have collected terms quadratic in the time derivative of the perturbed GKV field. In the limit of vanishing perturbation, the above quadratic contribution

$$\begin{aligned}
 \mathcal{H}^{(2)\text{space}} &= \frac{1}{2} \left( 1 + \frac{1}{2} h \right) \{ 1 - \bar{\mu} \} (\partial_i \delta\xi^i \partial_j \delta\xi^j) + \frac{1}{2} \left( 1 + \frac{1}{2} h \right) \{ -g^{ij} + h_{00} g^{ij} + h^{ij} \} (\partial_i \delta\xi^0 \partial_j \delta\xi^0) \\
 &+ \frac{1}{2} \left( 1 + \frac{1}{2} h \right) \{ g_{ab} g^{ij} + h_{ab} g^{ij} - g_{ab} h^{ij} \} (\partial_i \delta\xi^a \partial_j \delta\xi^b) + \left( 1 + \frac{1}{2} h \right) h_{0a} g^{ij} (\partial_i \delta\xi^0 \partial_j \delta\xi^a). \tag{4.8}
 \end{aligned}$$

In the above expression involving spatial derivatives of the perturbed GKV, the first two terms can provide a negative contribution to the Hamiltonian, thereby making it unbounded from below. In the second term, even though the metric perturbations try to keep this term positive, the background metric drives it to negative values. Similarly, if we want the theory to be ghost-free, the first term will turn negative, leading to an unbounded Hamiltonian. As argued in [12], in the unperturbed case, an unbounded Hamiltonian is potentially dangerous, as it can result in runaway behavior that drives the GKVs far from the Killing condition. Thus, the problems associated with the Hamiltonian density for the AKE in the exact case remain for the perturbed GKV  $\delta\xi^\alpha$  as well.

### B. Hyperbolicity

We now turn our attention to the hyperbolicity of the perturbed AKE. For this purpose, we employ the methods of hyperbolicity analysis for second-order systems, particularly that presented in [25,26] (see also [12]). In this approach, we compute the principal symbol for the system

to the Hamiltonian coincides with that presented in [12]. Though the kinetic term of the zeroth component of the perturbed Killing vector field, i.e.,  $\delta\xi^0$  in the Hamiltonian density harbors a negative sign in the unperturbed space-time for  $\mu < 2$  (see [12]), we see that in Eq. (4.7), the corresponding kinetic term in perturbed GKV has positive sign for  $\mu < 2 < \bar{\mu}$ . Thus, ghost modes can be avoided. The other quadratic terms in the time derivative have positive sign. Similarly, terms quadratic in the space derivatives of the perturbed Killing vector field yield the following expression for the Hamiltonian density:

of equations; if the principal symbol has real eigenvalues, the system is weakly hyperbolic, and if the principal symbol has a complete set of eigenvectors, the system is strongly hyperbolic. Collecting all the terms involving double derivatives of  $\delta\xi^\alpha$ , we obtain from Eq. (3.29),

$$(g^{\alpha\beta} - h^{\alpha\beta}) \partial_\alpha \partial_\beta \delta\xi^\mu + (1 - \mu - \delta\mu) (g^{\mu\alpha} - h^{\mu\alpha}) \partial_\alpha \partial_\sigma \delta\xi^\sigma \approx 0, \tag{4.9}$$

where the symbol  $\approx$  denotes equality up to terms not included in the principal part. In order to express the above equation in the desired form, we can decompose the metric as  $g^{\alpha\beta} = q^{\alpha\beta} - n^\alpha n^\beta + s^\alpha s^\beta$ , where  $n^\alpha$  is a timelike unit vector and  $s^\alpha$  is a spacelike unit vector. Further, defining  $n^\alpha \partial_\alpha = \partial_n$ ,  $s^\alpha \partial_\alpha = \partial_s$ , and  $q^{\alpha\beta} = q^{AB} \delta_A^\alpha \delta_B^\beta$ , we can rewrite the above equation into three separate equations; we obtain one by contraction of Eq. (4.9) with  $n_\mu$ , another by contraction of Eq. (4.9) with  $s_\mu$ , and the last by the projection of Eq. (4.9) along transverse directions. Keeping only the principal parts of these equations, we obtain

$$\begin{aligned}
 (2 - \mu - \delta\mu) (1 + h^{nn}) \partial_n^2 \delta\xi^n &\approx (1 - h^{ss}) \partial_s^2 \delta\xi^n + (3 - \mu - \delta\mu) h^{ns} \partial_n \partial_s \delta\xi^n + (1 - \mu - \delta\mu) (1 + h^{nn}) \partial_n \partial_s \delta\xi^s \\
 &- (1 - \mu - \delta\mu) h^{ns} \partial_s^2 \delta\xi^s. \tag{4.10}
 \end{aligned}$$

Note that since  $n^\mu$  is the timelike unit vector,  $\partial_n^2 \delta\xi^n$  corresponds to a double time derivative of the time component of the perturbed GKV. A similar analysis yields the following equation for  $\partial_n^2 \delta\xi^s$ , i.e., for the double time derivative of the spatial component of the perturbed GKV,

$$\begin{aligned}
 (1 + h^{nn}) \partial_n^2 \delta\xi^s &\approx (2 - \mu - \delta\mu) (1 - h^{ss}) \partial_s^2 \delta\xi^s + (3 - \mu - \delta\mu) h^{ns} \partial_n \partial_s \delta\xi^s - (1 - \mu - \delta\mu) h^{sn} \partial_n^2 \delta\xi^n \\
 &- (1 - \mu - \delta\mu) (1 - h^{ss}) \partial_s \partial_n \delta\xi^n. \tag{4.11}
 \end{aligned}$$

Finally, the double time derivative of the transverse component of the perturbed GKV yields

$$(1 + h^{nn})\partial_n^2 \delta \xi^A \approx (1 - h^{ss})\partial_s^2 \delta \xi^A + 2h^{ns}\partial_n \partial_s \delta \xi^A + (1 - \mu - \delta\mu)h^{As}\partial_s \partial_n \delta \xi^n - (1 - \mu - \delta\mu)h^{An}\partial_n^2 \delta \xi^n + (1 - \mu - \delta\mu)h^{An}\partial_n \partial_s \delta \xi^s - (1 - \mu - \delta\mu)h^{As}\partial_s^2 \delta \xi^s. \quad (4.12)$$

Therefore, we can read off the principal symbol  $P^s$  for this system of second-order differential equations, which takes the following form:

$$P^s = \begin{bmatrix} O_{4 \times 4} & I_{4 \times 4} \\ \mathbb{A} & \mathbb{B} \end{bmatrix}, \quad (4.13)$$

where  $O_{4 \times 4}$  and  $I_{4 \times 4}$  are the  $(4 \times 4)$  null and unit matrix, respectively. The entries  $\mathbb{A}$  and  $\mathbb{B}$  are also  $(4 \times 4)$  matrices with the following expressions:

$$\mathbb{A} = \begin{bmatrix} \mathcal{A} & \mathcal{B} & 0 \\ \mathcal{E} & \mathcal{F} & 0 \\ \mathcal{I}^A & \mathcal{J}^A & \mathcal{K}I_{2 \times 2} \end{bmatrix}, \quad \mathbb{B} = \begin{bmatrix} \mathcal{C} & \mathcal{D} & 0 \\ \mathcal{G} & \mathcal{H} & 0 \\ \mathcal{L}^A & \mathcal{M}^A & \mathcal{N}I_{2 \times 2} \end{bmatrix}, \quad (4.14)$$

where the unknown quantities appearing in the matrices  $\mathbb{A}$  and  $\mathbb{B}$  have the following expressions,

$$\mathcal{A} = \frac{(1 - h^{ss})}{(2 - \mu - \delta\mu)(1 + h^{nn})}, \quad \mathcal{B} = -\frac{(1 - \mu - \delta\mu)h^{ns}}{(2 - \mu - \delta\mu)(1 + h^{nn})}, \quad (4.15)$$

$$\mathcal{C} = \frac{(3 - \mu - \delta\mu)h^{ns}}{(2 - \mu - \delta\mu)(1 + h^{nn})}, \quad \mathcal{D} = \frac{(1 - \mu - \delta\mu)}{(2 - \mu - \delta\mu)}, \quad (4.16)$$

$$\mathcal{E} = -\frac{(1 - \mu - \delta\mu)h^{sn}}{(2 - \mu - \delta\mu)(1 + h^{nn})^2}, \quad \mathcal{F} = \frac{(2 - \mu - \delta\mu)(1 - h^{ss})}{(1 + h^{nn})}, \quad (4.17)$$

$$\mathcal{G} = -\frac{(1 - \mu - \delta\mu)(1 - h^{ss})}{(1 + h^{nn})}, \quad \mathcal{H} = \frac{(3 - \mu - \delta\mu)h^{ns}}{(1 + h^{nn})} - \frac{(1 - \mu - \delta\mu)^2 h^{sn}}{(2 - \mu - \delta\mu)(1 + h^{nn})}, \quad (4.18)$$

$$\mathcal{I} = -\frac{(1 - \mu - \delta\mu)h^{An}}{(2 - \mu - \delta\mu)(1 + h^{nn})^2}, \quad \mathcal{J} = -\frac{(1 - \mu - \delta\mu)h^{As}}{(1 + h^{nn})}, \quad (4.19)$$

$$\mathcal{K} = \frac{(1 - h^{ss})}{(1 + h^{nn})}, \quad \mathcal{L} = \frac{(1 - \mu - \delta\mu)h^{As}}{(1 + h^{nn})}, \quad (4.20)$$

$$\mathcal{M} = -\frac{(1 - \mu - \delta\mu)^2 h^{An}}{(2 - \mu - \delta\mu)(1 + h^{nn})} + \frac{(1 - \mu - \delta\mu)h^{An}}{(1 + h^{nn})}, \quad \mathcal{N} = 2\frac{h^{ns}}{(1 + h^{nn})}. \quad (4.21)$$

Note that as the perturbations vanish, we obtain  $\mathcal{A} = 1/(2 - \mu)$ ,  $\mathcal{B} = 0 = \mathcal{C}$ ,  $\mathcal{D} = (1 - \mu)/(2 - \mu)$ ,  $\mathcal{E} = 0$ ,  $\mathcal{F} = 2 - \mu$ ,  $\mathcal{G} = -(1 - \mu)$ ,  $\mathcal{H} = 0 = \mathcal{I} = \mathcal{J}$ ,  $\mathcal{K} = 1$ , and  $\mathcal{L} = 0 = \mathcal{M} = \mathcal{N}$ . Thus, our result agrees with that derived in [12]. If one chooses  $\mu + \delta\mu = 1$ , one has a complete set of eigenvectors, and if one chooses  $\mu + \delta\mu = 2$ , the principal symbol becomes singular. Setting  $\mu = 1$  and expanding the eigenvalues to first order in metric perturbations and second order in  $\delta\mu$ , one obtains the following expression for the eigenvalues of the principal symbol<sup>4</sup>:

$$\left\{ \frac{1}{2}(h^{nn} - 2h^{ns} + h^{ss} - 2), -\frac{1}{2}(h^{nn} + 2h^{ns} + h^{ss} - 2), \frac{1}{4}(-\delta\mu^2 h^{ns} + (\delta\mu^2 + 4)h^{ns} - 2h^{nn} - 2(h^{ss} - 2)), \frac{1}{4}(-\delta\mu^2 h^{ns} + (\delta\mu^2 + 4)h^{ns} - 2h^{nn} - 2(h^{ss} - 2)), \frac{1}{4}(-\delta\mu^2 h^{ns} + (\delta\mu^2 + 4)h^{ns} + 2h^{nn} + 2(h^{ss} - 2)), \frac{1}{4}(-\delta\mu^2 h^{ns} + (\delta\mu^2 + 4)h^{ns} + 2h^{nn} + 2(h^{ss} - 2)) \right\}. \quad (4.22)$$

Since the eigenvalues presented above are real, this implies that to first order in the metric perturbation and second order in  $\delta\mu$ , the perturbed AKE is weakly hyperbolic. This is consistent with the result in [12], in which the AKE for

<sup>4</sup>These calculations were performed using the package `xact` in *Mathematica*.

the background spacetime was found to be weakly hyperbolic for general  $\mu \neq 1, 2$ .

## V. INDUCED PERTURBATIONS: IMPLICATIONS FOR THE ALMOST KILLING EQUATION

So far, we had considered the metric perturbation and the perturbation of the GKV to be independent. However, in most of the physical scenarios of interest, e.g., perturbation of the black hole spacetime due to matter fields entering the horizon, the metric perturbation sources the perturbation of the almost Killing equation. Thus, we may consider the perturbation of the GKV to be induced by the metric perturbation. In what follows, we will consider such induced perturbation of the GKV and hence determine their evolution equations order by order.

### A. First-order perturbation

As emphasized before, we will imagine a class of nontrivial perturbative solutions to the AKE which are induced by the metric perturbations. This may be quantified by assuming that  $\delta\xi^\alpha$  and  $h_{\alpha\beta}$  are implicitly proportional to the same expansion parameter  $\epsilon$  (see footnote 2), and then solving the perturbed AKE order by order in  $\epsilon$ . In particular, for the GKV  $\chi^\alpha$  we write

$$\chi^\alpha = \xi^\alpha + \delta\chi_1^\alpha + \delta\chi_2^\alpha + \mathcal{O}(\epsilon^3), \quad (5.1)$$

where we have assumed  $\delta\chi_i^\alpha \propto \epsilon^i$ . As before, the background spacetime is assumed to admit a Killing vector field  $\xi^\alpha$ , which satisfies the AKE presented in Eq. (2.2) exactly. Further, imposing the Lorenz gauge condition,  $\nabla^\beta h_{\alpha\beta} = \nabla_\alpha h/2$ , and choosing the background value of the parameter  $\mu$  to be constant, the expansion of the perturbed AKE presented in Eq. (3.29) to  $\mathcal{O}(\epsilon)$  reduces to the following equation for  $\delta\chi_1^\alpha$ :

$$\square\delta\chi_1^\alpha + R^\alpha{}_\beta\delta\chi_1^\beta + (1-\mu)\nabla^\alpha[\nabla \cdot \delta\chi_1 + (\xi^\beta\nabla_\beta h)/2] = 0. \quad (5.2)$$

Note that this reduces to the background AKE when  $\mu = 1$  or, when  $\xi^\beta\nabla_\beta h$  is a constant; in those cases, one must consider higher-order corrections to the AKE. In the cases where  $\xi^\beta\nabla_\beta h$  is nontrivial, for instance, if the perturbations are the result of an energy-momentum tensor with nontrivial trace, one can apply the expression Eq. (2.4) for the Noether current defined with respect to the background derivatives directly to  $\delta\chi_1^\alpha$  and substitute Eq. (5.2) to obtain the following current:

$$J_1^\alpha = 2R^\alpha{}_\beta\delta\chi_1^\beta + \nabla^\alpha[(2-\mu)\nabla \cdot \delta\chi_1 + (1-\mu)\xi^\alpha\nabla_\alpha h/2], \quad (5.3)$$

which satisfies the identity  $\nabla \cdot J_1 = 0$ . If the background spacetime is vacuum, setting  $R^\alpha{}_\beta = 0$ , this identity yields

$$\square\Phi_1 = 0, \quad (5.4)$$

where

$$\Phi_1 := (2-\mu)\nabla \cdot \delta\chi_1 + \frac{1}{2}(1-\mu)(\xi^\alpha\nabla_\alpha h). \quad (5.5)$$

It follows that if the initial data satisfy the constraint  $\Phi_1 = 0$  and its first time derivative also vanishes, then the constraint  $\Phi_1 = 0$  is preserved by the evolution of Eq. (5.4). Under this constraint, Eq. (5.2) may be rewritten as

$$\square\delta\chi_1^\alpha - \nabla^\alpha(\nabla \cdot \delta\chi_1) = 0, \quad (5.6)$$

which is equivalent to the AKE for  $\mu = 2$  in vacuum spacetime. It was demonstrated in [12] that the  $\mu = 2$  parameter choice avoids ghosts and is dynamically equivalent to the Maxwell theory, thus avoiding the problems arising from an unbounded Hamiltonian. Furthermore, recall that to first order in the gravitational perturbation, imposing the Lorenz gauge condition, the trace  $h$  of the gravitational perturbation satisfies the following evolution equation on a vacuum background:

$$\square h = 16\pi\delta_1 T, \quad (5.7)$$

where  $\delta_1 T$  is the first-order perturbation of the trace of the matter energy-momentum tensor. If one assumes  $\delta_1 T = 0$ , i.e., the matter field perturbing the spacetime is dilute radiation and the background is vacuum, one can impose the condition that  $h = 0$ , in which case, from Eq. (5.5), it follows that the constraint  $\Phi_1 = 0$  implies  $\nabla \cdot \delta\chi_1 = 0$ . Subsequently, substituting this result in Eq. (5.6) simplifies to  $\square\delta\chi_1^\alpha = 0$ , which is satisfied by the background Killing vector field  $\xi^\alpha$ . Thus, one may expect that in this case, the perturbed spacetime will respect the symmetries of the background spacetime. On the other hand, if  $\delta_1 T \neq 0$ , then the background Killing vector field does not, in general, satisfy the constraint  $\Phi_1 = 0$ , and the resulting solutions for Eq. (5.6) differ nontrivially from the background Killing vector.

We have therefore derived that if the spacetime perturbation  $h_{\alpha\beta}$  is being sourced by an energy-momentum tensor with a nonvanishing trace, then the evolution equation for the first-order perturbation  $\delta\chi_1^\alpha$  must differ from the background Killing vector field in a nontrivial manner. Furthermore, we find that for the weakly hyperbolic parameter choices, i.e.,  $\mu \neq 1$  and  $\mu \neq 2$ , and initial data satisfying the constraint  $\Phi_1 = 0$  and  $\partial_t\Phi_1 = 0$ , the AKE propagates the constraint  $\Phi_1 = 0$  and avoids the dynamical instabilities associated with ghosts and unboundedness in

the Hamiltonian. First-order perturbations of the AKE are therefore different from the background Killing vector field and are well suited for describing the perturbations of the vacuum spacetimes induced by dilute matter in which  $\delta_1 T \neq 0$ .

### B. Second-order perturbations

As the previous section demonstrates, perturbations of the AKE to  $\mathcal{O}(\epsilon)$  are insensitive to the metric perturbations in scenarios with  $\delta_1 T = 0$ . This is the case if the perturbations consist of weak gravitational radiation, or dilute

matter source encompassing radiation or null dust. In such situations, one should consider the evolution equation for the second-order correction  $\delta\chi_2^\nu$ . Assuming constant  $\mu$  and a vacuum background spacetime, the homogeneous part of the equation is the same as that of the first-order case

$$\square\delta\chi_2^\nu + (1 - \mu)\nabla^\nu(\nabla \cdot \delta\chi_2) = -(j_2^\nu + k_2^\nu + l_2^\nu + m_2^\nu), \quad (5.8)$$

and the inhomogeneous part  $j_2^\nu$  takes the explicit form

$$\begin{aligned} j_2^\nu &= 2\xi^\alpha h^{\beta\sigma} \{h^{\nu\tau} R_{\alpha\beta\tau\sigma} - h_\beta{}^\tau R^\nu{}_{\sigma\alpha\tau} - \nabla_\sigma \nabla_\alpha h^\nu{}_\beta\} - \xi^\alpha \nabla_\alpha h_{\beta\sigma} \nabla^\nu h^{\beta\sigma} + 2\nabla^\nu (\xi^\alpha h^{\beta\sigma} \nabla_\alpha h_{\beta\sigma}) - 2h^{\beta\sigma} \nabla^\nu \xi^\alpha \nabla_\sigma h_{\alpha\beta} \\ &\quad + 2h_\alpha{}^\sigma \nabla^\beta \xi^\alpha \{ \nabla_\beta h^\nu{}_\sigma - \nabla^\nu h_{\beta\sigma} + \nabla_\sigma h^\nu{}_\beta \}, \\ k_2^\nu &= 2h^\nu{}_\beta \nabla^\beta \nabla_\alpha \delta\chi_1^\alpha - 2h_{\alpha\beta} \nabla^\beta \nabla^\alpha \delta\chi_1^\nu + 2\nabla^\beta \delta\chi_1^\alpha \{ \nabla_\alpha h^\nu{}_\beta + \nabla_\beta h^\nu{}_\alpha - \nabla^\nu h_{\alpha\beta} \} - 2\delta\chi_1^\alpha h^{\beta\sigma} R^\nu{}_{\beta\alpha\sigma}, \\ l_2^\nu &= h^{\nu\beta} \nabla_\beta \xi^\alpha \nabla_\alpha h - \nabla^\nu [\delta\chi_1^\alpha \nabla_\alpha h + (2\nabla_\alpha \delta\chi_1^\alpha + \xi^\alpha \nabla_\alpha h) \delta\mu] + (\mu - 2) \{ h^{\nu\beta} \nabla_\beta (\xi^\alpha \nabla_\alpha h) - \nabla^\nu (\delta\chi_1^\alpha \nabla_\alpha h) \}, \\ m_2^\nu &= (\mu - 2) \{ \nabla^\nu (\xi^\alpha h^{\beta\sigma} \nabla_\alpha h_{\beta\sigma}) + 2h^\nu{}_\beta \nabla^\beta \nabla_\alpha \delta\chi_1^\alpha \}. \end{aligned} \quad (5.9)$$

The terms in Eq. (5.9) are organized so that  $k_2^\nu = 0$ , if  $\delta\chi_1^\mu = 0$  or, if  $\delta\chi_1^\mu$  satisfies the Killing equation, and  $l_2^\nu = 0$  if  $h = 0 = \delta\mu$ . As pointed out in the first-order case, for  $h = 0$ , the relevant solutions for  $\delta\chi_1^\alpha$  are essentially those of the background Killing vector; in that case, one may choose  $\delta\chi_1^\mu = 0$ , as one can absorb it into the background Killing vector field.

Further understanding of the second-order perturbation can be achieved by computing the second-order perturbation of the Noether current assuming constant  $\mu$  and a vacuum background  $R_{\mu\nu} = 0$ :

$$\begin{aligned} J_2^\nu &= 4\delta\chi_1^\alpha \delta_1 R^\nu{}_\alpha + 2\xi^\alpha (\delta_2 R^\nu{}_\alpha - 2h^\nu{}_\beta \delta_1 R_\alpha{}^\beta) \\ &\quad + (\mu - 2) [\nabla_\alpha h_{\sigma\tau} (\xi^\alpha \nabla^\nu h^{\sigma\tau} + h^{\sigma\tau} \nabla^\nu \xi^\alpha) \\ &\quad + \xi^\alpha h^{\sigma\tau} \nabla^\nu \nabla_\alpha h_{\sigma\tau} + 2h^\nu{}_\beta \nabla^\beta (\nabla \cdot \delta\chi_1) - \nabla^\nu (\nabla \cdot \delta\chi_2) \\ &\quad - \{ \nabla^\nu (\delta\chi_1^\alpha \nabla_\alpha h) - h^{\nu\beta} \nabla_\beta (\xi^\alpha \nabla_\alpha h) \}], \end{aligned} \quad (5.10)$$

where  $\delta_1 R_{\alpha\beta}$  and  $\delta_2 R_{\alpha\beta}$  are the respective first- and second-order perturbations of the lowered index Ricci tensor. Note that the quantity within the curly brackets vanishes when  $h = 0$ . We further see that when  $\mu = 2$ , most of the terms in the Noether current, except for the first two terms, identically vanish. Therefore, for  $\mu = 2$ , the Noether current depends on the perturbations of the Ricci tensor alone, as one might expect.

We then turn to the second-order perturbations of the identity presented in Eq. (2.7). Assuming vacuum spacetime along with transverse-traceless gauge  $h = 0$ , and setting  $\mu$  to be a constant, the perturbation of the Komar identity takes the following form:

$$(\mu - 2)\Psi_L = \Psi_R, \quad (5.11)$$

where

$$\begin{aligned} \Psi_L &\equiv \square[\nabla \cdot \delta\chi_2 + \delta\chi_1^\alpha \nabla_\alpha h - h^{\sigma\tau} \nabla_\alpha (\xi^\alpha h_{\sigma\tau})] \\ &\quad + 2h^{\sigma\tau} (\nabla_\nu \nabla_\sigma \nabla_\tau \delta\chi_1^\nu + R_{\alpha\sigma\tau\beta} \nabla^\beta \delta\chi_1^\alpha) \\ &\quad + \nabla_\beta [\nabla_\alpha (h^{\beta\alpha} \xi^\sigma \nabla_\sigma h)] + \nabla^\beta h \nabla_\beta (\xi^\alpha \nabla_\alpha h), \end{aligned} \quad (5.12)$$

$$\begin{aligned} \Psi_R &\equiv \xi^\alpha (\nabla_\alpha \delta_2 R - 2h^{\sigma\tau} \nabla_\alpha \delta_1 R_{\sigma\tau}) - 4\delta_1 R_{\alpha\sigma} h_\beta{}^\sigma \nabla^\beta \xi^\alpha \\ &\quad + 2(\delta\chi_1^\alpha \nabla_\alpha \delta_1 R + 2\delta_1 R_{\alpha\beta} \nabla^\beta \delta\chi_1^\alpha). \end{aligned} \quad (5.13)$$

Note that the quantity  $\Psi_R$  is independent of the second-order perturbation of the GKV  $\delta\chi_2^\mu$ , and when  $\mu = 2$ , it follows from Eq. (5.11) that  $\Psi_R = 0$ . For vacuum spacetime, in the transverse-traceless gauge, at first order, one may choose  $\delta\chi_1^\alpha$  to satisfy the background Killing equation, and hence, the terms dependent on  $\delta\chi_1^\alpha$  disappear from  $\Psi_R$ , so that  $\Psi_R$  depends only on the background quantities and the perturbation  $h_{\mu\nu}$ . It follows that to second order,  $\Psi_R$  must be zero for vacuum perturbations (in which  $\bar{R}_{\mu\nu} = 0$ ) or any metric perturbation  $h_{\mu\nu}$  which satisfies  $h = \delta_1 R = 0$  and permits a solution to the  $\mu = 2$  AKE.

If  $\Psi_R = 0$ , the second-order perturbations should satisfy the following:

$$\square\Phi_2 = 0, \quad (5.14)$$

where

$$\Phi_2 := (\mu - 2)\{\nabla \cdot \delta\chi_2 - h^{\sigma\tau}\nabla_\alpha(\xi^\alpha h_{\sigma\tau})\}. \quad (5.15)$$

It follows that when  $\Psi_R = 0$ , the field equations propagate the constraint  $\Phi_2 = 0$  for the second-order perturbations (assuming initial data satisfying  $\partial_t\Phi_2 = 0$ ). On a vacuum background, one may (assuming  $\mu \neq 2$ ) use the constraint  $\Phi_2 = 0$  to rewrite Eq. (5.8)

$$\begin{aligned} \square\delta\chi_2^\nu - \nabla^\nu(\nabla \cdot \delta\chi_2) \\ = -j_2^\nu - m_2^\nu - (2 - \mu)\nabla^\nu\{h^{\sigma\tau}\nabla_\alpha(\xi^\alpha h_{\sigma\tau})\}, \end{aligned} \quad (5.16)$$

so that the equation for  $\delta\chi_2^\nu$  resembles the  $\mu = 2$  equation with a source. For second-order perturbations, the question of whether the perturbations suffer from dynamical instabilities due to the unboundedness of the Hamiltonian depends on the behavior of the rhs of Eq. (5.16).

The claim that  $\Psi_R = 0$  for transverse-traceless metric perturbations of vacuum spacetimes that admit Killing vectors suggests an identity for such perturbations. However, the arguments we have presented so far do not yet constitute a proof of such an identity, as they depend on the existence of solutions for the exact  $\mu = 2$  AKE. Though the AKE fails to admit a well-posed initial value problem for the  $\mu = 2$  parameter choice, there is some reason to expect that the failure is primarily due to nonuniqueness, rather than existence. For instance, it is straightforward to show that on vacuum spacetimes, the exact AKE becomes an identity for the gradient of an arbitrary function. Furthermore, one can show that in locally flat coordinates, the time derivatives for the time component of  $\chi^\mu$  disappear, so that the AKE is an underdetermined dynamical system. One might therefore expect the existence of solutions to the exact  $\mu = 2$  AKE (and consequently,  $\Phi_2 = 0$ ) to hold for a

large class of transverse-traceless metric perturbations of vacuum spacetimes which admit Killing vectors.

## VI. PERTURBATIONS OF THE NOETHER CHARGE AND ITS THERMODYNAMICAL INTERPRETATION

In this section, we will discuss how the Noether charge associated with the GKV  $\chi^\alpha$  associated with the AKE is affected by the perturbation of the metric. As we will demonstrate, the perturbed Noether charge will have an interesting thermodynamical interpretation. Applying Eq. (2.4) to the solutions of the AKE, which correspond to the GKV field  $\chi^\mu$ , the Noether current for the GKVs in the perturbed spacetime takes the following form:

$$\bar{J}^\mu = 2\bar{R}^\mu{}_\nu\chi^\nu + (2 - \bar{\mu})\bar{\nabla}^\mu[\bar{\nabla}_\sigma\chi^\sigma] + (\bar{\nabla}_\sigma\chi^\sigma)\bar{g}^{\mu\alpha}\nabla_\alpha(2 - \bar{\mu}), \quad (6.1)$$

where the AKE presented in Eq. (2.2) has been used. Let us now use the fact, following Eq. (3.1), that the spacetime metric can be expressed as  $\bar{g}_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu}$ , where  $h_{\mu\nu}$  is the perturbation, possibly due to some matter field entering the background spacetime geometry. As a consequence, we also have  $\xi^\sigma \rightarrow \xi^\sigma + \delta\xi^\sigma = \chi^\sigma$ , where  $\chi^\sigma$  is the GKV, with the associated Noether current being given by Eq. (6.1).

Thus, the Noether current associated with the GKV field can be decomposed into the Komar current for the background Killing vector field  $\xi^\alpha$  and a part containing additional corrections arising out of the gravitational perturbation  $h_{\mu\nu}$  and the perturbation of the Killing vector field  $\delta\xi^\alpha$ . We then obtain the following expression for the Noether current associated with the GKV field  $\chi^\alpha$ :

$$\begin{aligned} \bar{J}^\nu = 2R^\nu{}_\sigma\chi^\sigma + (-\square h_\sigma^\nu + R^{\nu\mu}h_{\mu\sigma} - R_{\sigma\mu}h^{\mu\nu} - 2R^\nu{}_{\mu\sigma\rho}h^{\mu\rho})\chi^\sigma + (2 - \bar{\mu})(g^{\nu\alpha} - h^{\nu\alpha})\nabla_\alpha\left(\nabla_\sigma\chi^\sigma + \frac{1}{2}\chi^\sigma\nabla_\sigma h\right) \\ + (g^{\nu\alpha} - h^{\nu\alpha})\left(\nabla_\sigma\chi^\sigma + \frac{1}{2}\chi^\sigma\nabla_\sigma h\right)\nabla_\alpha(2 - \bar{\mu}), \end{aligned} \quad (6.2)$$

where we have used the Lorenz gauge condition to simplify the expression further. The above provides the expansion of the Noether current in terms of the gravitational perturbation; further expressing  $\chi^\sigma = \xi^\sigma + \delta\xi^\sigma$ , we obtain the following expression for the change in the Noether current:

$$\begin{aligned} \delta J^\nu = 2R^\nu{}_\sigma\delta\xi^\sigma + (-\square h_\sigma^\nu + R^{\nu\mu}h_{\mu\sigma} - R_{\sigma\mu}h^{\mu\nu} - 2R^\nu{}_{\mu\sigma\rho}h^{\mu\rho})(\xi^\sigma + \delta\xi^\sigma) + (2 - \bar{\mu})(g^{\nu\alpha} - h^{\nu\alpha})\nabla_\alpha(\nabla_\sigma\delta\xi^\sigma) \\ + \frac{1}{2}(2 - \bar{\mu})g^{\nu\alpha}\nabla_\alpha[(\xi^\sigma + \delta\xi^\sigma)\nabla_\sigma h] - \left[g^{\nu\alpha}\left(\nabla_\sigma\delta\xi^\sigma + \frac{1}{2}(\xi^\sigma + \delta\xi^\sigma)\nabla_\sigma h\right) - h^{\nu\alpha}(\nabla_\sigma\delta\xi^\sigma)\right](\nabla_\alpha\delta\mu). \end{aligned} \quad (6.3)$$

Here we have assumed that the background spacetime inherits Killing symmetry, and hence,  $\xi^\sigma$  is a Killing vector field, such that  $\nabla_\sigma\xi^\sigma = 0$ , which we have used in deriving the above expression. Using the expressions for the perturbations of the Ricci tensor and the Einstein tensor in the Lorenz gauge, the above change in the Noether current may be expressed in several ways, among which we quote the expression involving the Einstein tensor below

$$\begin{aligned} \delta J^\nu &= 2R^\nu{}_\sigma \delta \xi^\sigma + 2\delta G^\nu{}_\sigma (\xi^\sigma + \delta \xi^\sigma) - \left( \frac{1}{2} \square h + R_{\mu\rho} h^{\mu\rho} \right) (\xi^\nu + \delta \xi^\nu) + (2 - \bar{\mu})(g^{\nu\alpha} - h^{\nu\alpha}) \nabla_\alpha (\nabla_\sigma \delta \xi^\sigma) \\ &+ \frac{1}{2} (2 - \bar{\mu}) g^{\nu\alpha} \nabla_\alpha [(\xi^\sigma + \delta \xi^\sigma) \nabla_\sigma h] - \left[ g^{\nu\alpha} \left( \nabla_\sigma \delta \xi^\sigma + \frac{1}{2} (\xi^\sigma + \delta \xi^\sigma) \nabla_\sigma h \right) - h^{\nu\alpha} (\nabla_\sigma \delta \xi^\sigma) \right] (\nabla_\alpha \delta \mu). \end{aligned} \quad (6.4)$$

Observe that, using the perturbed Einstein equations, one can replace the perturbation of the Einstein tensor appearing in the above expression with the perturbation of the matter energy-momentum tensor. In that case, the object  $(\delta G^\nu{}_\sigma \xi^\sigma)$  will correspond to the flux of the matter energy-momentum tensor through the Killing horizon, to which the Killing vector field is orthogonal. We will come back to this point later in this section, but we first discuss a couple of interesting limits:

- (i) Even though we have treated the gravitational perturbation and the perturbation of the Killing vector field separately, the perturbation  $\delta \xi^\alpha$  we are interested in is induced by the perturbation of the background spacetime. Thus, it is natural to decompose the perturbed Killing vector field as  $\delta \xi^\mu = \delta \chi_1^\mu + \delta \chi_2^\mu$  [cf. Eq. (5.1)], where  $\delta \chi_1^\mu$  is linear in the gravitational perturbation and  $\delta \chi_2^\mu$  is quadratic in the gravitational perturbation. Then to the linear order in the gravitational perturbation, the perturbed Noether current from Eq. (6.4) takes the following form<sup>5</sup>:

$$\begin{aligned} \delta J^\nu &= 2R^\nu{}_\sigma \delta \chi_1^\sigma + 2\delta_1 R^\nu{}_\sigma \xi^\sigma + (2 - \mu) \nabla^\nu (\nabla \cdot \delta \chi_1) \\ &+ \frac{1}{2} (2 - \mu) \nabla_\nu (\xi^\sigma \nabla_\sigma h) \quad (6.5) \\ &= 2R^\nu{}_\sigma \delta \chi_1^\sigma + 2\delta_1 G^\nu{}_\sigma \xi^\sigma - \left( \frac{1}{2} \square h + R_{\mu\rho} h^{\mu\rho} \right) \xi^\nu \\ &+ \frac{1}{2} \nabla_\nu (\xi^\sigma \nabla_\sigma h), \quad (6.6) \end{aligned}$$

where the Lorenz gauge condition has been used. In the first line of the above expression, we have expressed the change in the Noether current in terms of the change in the Ricci tensor. In the second line, we express the change in Noether current in terms of the change in the Einstein tensor, and apply the constraint  $\Phi_1 = 0$  [with  $\Phi_1$  given in Eq. (5.5)], which follows from the fact that  $\delta \chi_1^\mu$  satisfies Eq. (5.2), which as we showed earlier propagates the constraint  $\Phi_1 = 0$  for an appropriate choice of initial data.

- (ii) For spacetimes which may contain a dilute amount of nongravitational radiation on vacuum backgrounds, i.e., with  $R_{\mu\nu} = 0$ , one can use the

<sup>5</sup>One should be careful not to confuse this expression with Eq. (5.3), which satisfies the identity  $\nabla \cdot J_1 = 0$ .  $\delta J^\alpha$ , on the other hand, satisfies  $\nabla \cdot (J + \delta J) = 0$ .

transverse-traceless gauge (effectively setting  $h=0$ ), and hence, the above expression for the change in the Noether current simplifies considerably:

$$\delta J^\nu = 2\delta_1 G^\nu{}_\sigma \xi^\sigma. \quad (6.7)$$

Thus, using the perturbed Einstein equations, we find that given the constraint  $\Phi_1 = 0$ , the change in the Noether current to first order is simply equal to  $16\pi(\delta T^\nu{}_\sigma \xi^\sigma)$ , which corresponds to the matter field flowing into the Killing horizon. This will have thermodynamical interpretation, as we compute the associated change in the Noether charge.

- (iii) For spacetimes which contain a dilute amount of matter on vacuum backgrounds, one can instead employ the  $\mu = 2$  parameter choice, in which case, the Noether current also simplifies to

$$\delta J^\nu = 2\delta_1 R^\nu{}_\sigma \xi^\sigma = (2\delta_1 G^\nu{}_\sigma + \delta_\sigma^\nu \delta_1 R) \xi^\sigma. \quad (6.8)$$

We see that for matter fields, the Noether current does not measure energy and momentum in the sense of the energy-momentum tensor, due to the term containing  $\delta_1 R$ . However, one may nonetheless still regard the Noether current and its associated charge as a measure of the matter content, and an analysis [27] comparing Komar integrals for radiation with that of matter in cylindrical symmetry suggests that charges constructed from Noether currents measure the effective gravitating mass.

Thus, we have discussed situations of physical interest and how the change in the Noether current can be simplified and interpreted in these scenarios. We will now proceed to compute the change in the Noether charge due to the gravitational perturbation and the perturbation of the Killing vector field.

In order to determine the Noether charge, we have to integrate the Noether current over a three-surface. One may define such a hypersurface as a level surface of some scalar function  $\Phi = \Phi(x)$ ; defined in this way, the surface is held fixed in the manifold and does not change under metric perturbations. However, the measure on the surface will change. In particular, for a  $\Phi = \text{constant}$  surface, the integration measure is  $d\Sigma_\alpha = d^3x \sqrt{h} n_\alpha$ , where  $n_\alpha$  is the normalized normal on this surface, and  $h$  is the determinant of the induced metric on the  $\Phi = \text{constant}$  surface. It is also possible to express the integration measure on a  $\Phi = \text{constant}$  surface as  $d\Sigma_\alpha = d^3x \sqrt{-g} \nabla_\alpha \Phi$ , where  $\nabla_\alpha \Phi$



is the unnormalized normal to the  $\Phi = \text{constant}$  surface. Under metric perturbations, the measure changes as  $\sqrt{-g} \rightarrow \sqrt{-\bar{g}} = \sqrt{-g}\{1 + (1/2)h\}$ . Thus, the Noether charge associated with the GKV field  $\chi^\alpha = \xi^\alpha + \delta\xi^\alpha$  becomes

$$\begin{aligned} \bar{Q}[\xi + \delta\xi] &= \int d\bar{\Sigma}_\alpha \bar{J}^\alpha = \int d\Sigma_\alpha \left(1 + \frac{1}{2}h\right) (J^\alpha + \delta J^\alpha) \\ &= Q[\xi] + \int d\Sigma_\alpha \left(\frac{1}{2}h\right) J^\alpha + \int d\Sigma_\alpha \delta J^\alpha \\ &\quad + \int d\Sigma_\alpha \left(\frac{1}{2}h\right) \delta J^\alpha, \end{aligned} \quad (6.9)$$

where  $Q[\xi]$  denotes the Komar charge of the background Killing vector field. It is straightforward to read off the change in the Noether charge, which takes the following form:

$$\begin{aligned} \delta Q &\equiv \bar{Q}[\xi + \delta\xi] - Q[\xi] \\ &= \int d\Sigma_\alpha \left(\frac{1}{2}h\right) J^\alpha + \int d\Sigma_\alpha \delta J^\alpha + \int d\Sigma_\alpha \left(\frac{1}{2}h\right) \delta J^\alpha. \end{aligned} \quad (6.10)$$

Note that the first term is due to the perturbation of the integration measure, while the second and third terms result from a change in the Noether current. As in the case of the Noether current, this expression can be simplified further if we assume that the perturbations correspond to dilute radiation in vacuum background spacetime. The transverse-traceless gauge condition can then be imposed, and thus, the trace of the perturbation  $h$  can be set to zero. Therefore, the above expression for the change in the Noether charge becomes

$$\delta Q = 2 \int d\Sigma_\alpha \delta G^\alpha{}_\beta \xi^\beta. \quad (6.11)$$

As we will discuss below, this expression can be understood from a physical as well as thermodynamical perspective.

We consider the case where the metric  $g_{\mu\nu}$  is a vacuum black hole spacetime, with a Killing vector field  $\xi^\alpha$  defining the Killing horizon, to which the Killing vector is orthogonal. Since the Killing horizon is a null surface, it follows that we can consider the surface on which the Noether charge is computed to be the Killing horizon, which is bound to have a thermodynamic interpretation. In particular, for a generic null surface, the Noether charge associated with a Killing-like vector corresponds to  $Q = 16\pi \mathcal{T}S$ , where  $\mathcal{T}$  is the temperature associated with the null surface, and  $S$  is the associated entropy, which for general relativity is simply (area/4) [21,28]. As evident from the Vaidya solution considered in Sec. II and it follows that to first

order, under the assumptions considered here, the temperature does not change, and hence, the above change in the Noether charge may be interpreted as

$$\mathcal{T} \delta S = \int d\Sigma_\alpha \delta T^\alpha{}_\beta \xi^\beta, \quad (6.12)$$

which is the Clausius relation. In brief, this suggests that due to radiation falling into the Killing horizon, the spacetime is perturbed, and the Killing vector also ceases to be Killing, rather it becomes a GKV. This perturbation under the appropriate limit yields the Clausius relation. Thus, the formalism developed here for the case of radiation provides a close correspondence with the thermodynamic nature of gravity. One may make a similar argument for the case of nonradiative matter with the  $\mu = 2$  parameter choice for null surfaces, provided that  $d\Sigma_\alpha \xi^\alpha = 0$ , or by interpreting the Noether charge in terms of gravitating mass, as suggested in [27].

## VII. FINAL REMARKS

Killing vectors play a central role in characterizing spacetime symmetries, which are crucial in determining the conserved quantities that can be constructed in a given spacetime geometry. Systems of astrophysical interest are often symmetric only in a first approximation, and the spacetime geometries for such systems are often more accurately described in terms of a symmetric spacetime background with perturbations (the latter due to gravitational radiation or inhomogeneities and highly dynamical behavior in matter fields) which explicitly break the symmetry of the background. This motivates the perturbative study of the AKE, the solutions of which provide the generalizations of Killing vectors (which we refer to as GKVs) appropriate for the perturbed spacetime.

In this article, we have examined in detail the construction and behavior of these GKVs as perturbative solutions to the AKE associated with the metric perturbations of vacuum and nonvacuum spacetimes, which admit a Killing vector field. This has been achieved in two steps: (a) by considering the perturbation of the action yielding the AKE and then varying the same with respect to the perturbed GKV and (b) by perturbing the AKE and hence determining the evolution equation of the perturbed GKV. The matching of both of these equations explicitly demonstrates the internal consistency of these results.

Additionally, it turns out that the hyperbolicity and Hamiltonian stability of the perturbed GKV remains unchanged compared to its unperturbed counterpart if the GKV and the metric perturbations are kept independent. However, we have found that in the case where the perturbations of the GKV are sourced by the metric perturbation, the problem of an unbounded Hamiltonian can be avoided at first order, and at second order, the problem may also be avoided if the metric perturbations are

transverse and traceless (assuming the perturbations remain well behaved), and perturbative solutions to the  $\mu = 2$  AKE exist to second order. We have found that the first-order equations trivialize (they reduce to the background AKE) for traceless metric perturbations; for dilute radiation, the second-order case is necessary. We have also examined the first-order behavior of the Noether current constructed from a GKV and its associated charge. Intriguingly, it turns out that the conservation of the Noether current introduces additional constraints in the theory, which helps significantly to simplify the evolution equation for the perturbed GKV. In particular, if the perturbed matter energy-momentum tensor is traceless, i.e., the perturbation is due to null matter field, it follows that the first-order perturbation of the GKV can be absorbed within the background Killing vector field. To second order, we find that the second-order perturbation always yields nontrivial modifications to the background Killing vector field.

Finally, the perturbation of a background spacetime respecting Killing symmetry also has interesting thermodynamic implications. In particular, as we have demonstrated, the perturbation of the Noether charge to first order can be expressed as  $\mathcal{T}\delta S$ . This is because to first order in the perturbation, under these assumptions, the surface gravity does not change. This is also apparent from the example of Vaidya spacetime considered in this work, which further corroborates our claims regarding the thermodynamic interpretation for the perturbed Noether charge and currents associated with the generalized Killing vector fields.

## ACKNOWLEDGMENTS

We are grateful to Edgar Gasperin and David Hilditch for helpful discussions. Some of the calculations were performed using the `xAct` package [29] for *Mathematica*. J. C. F. acknowledges support from FCT Grants No. PTDC/MAT-APL/30043/2017 and No. UIDB/00099/2020. Research of S. C. is funded by the INSPIRE Faculty fellowship from DST, Government of India (Reg. No. DST/INSPIRE/04/2018/000893) and by the Start-Up Research Grant from SERB, DST, Government of India (Reg. No. SRG/2020/000409).

## APPENDIX: VARIATIONAL PRINCIPLE FOR AKE

Here, we review the derivation of the AKE from a variational principle, following the notation and conventions of Sec. III A. We rewrite here the action given in Eq. (2.3) in expanded form:

$$\mathcal{A}[\chi] = \int_{\mathcal{V}} d^4x \sqrt{-\bar{g}} \left[ -\frac{1}{4} (\bar{\nabla}_{\mu} \chi_{\nu} + \bar{\nabla}_{\nu} \chi_{\mu}) (\bar{\nabla}^{\mu} \chi^{\nu} + \bar{\nabla}^{\nu} \chi^{\mu}) + \frac{\bar{\mu}}{2} (\bar{\nabla}_{\mu} \chi^{\mu})^2 \right], \quad (\text{A1})$$

where  $\bar{g}_{\alpha\beta}$  is the metric of the perturbed spacetime, and  $\bar{\mu} \equiv \bar{\mu}(x)$  is an arbitrary function of the spacetime coordinates. Varying the above action with respect to arbitrary variations of  $\chi^{\mu}$ , including endpoint contributions, we obtain

$$\begin{aligned} \Delta \mathcal{A} = & \int_{\mathcal{V}} d^4x \sqrt{-\bar{g}} [-\bar{g}_{\mu\alpha} \bar{g}_{\nu\beta} (\bar{\nabla}^{\mu} \chi^{\nu} + \bar{\nabla}^{\nu} \chi^{\mu}) (\bar{\nabla}^{\alpha} \Delta \chi^{\beta}) + \bar{\mu} (\bar{\nabla}_{\mu} \chi^{\mu}) (\bar{\nabla}_{\alpha} \Delta \chi^{\alpha})] \\ & + \int_{\partial \mathcal{V}} d^3x \sqrt{-\bar{g}} \left[ -\frac{1}{4} (\bar{\nabla}_{\mu} \chi_{\nu} + \bar{\nabla}_{\nu} \chi_{\mu}) (\bar{\nabla}^{\mu} \chi^{\nu} + \bar{\nabla}^{\nu} \chi^{\mu}) + \frac{\bar{\mu}}{2} (\bar{\nabla}_{\mu} \chi^{\mu})^2 \right] \Delta x^{\alpha} \bar{\nabla}_{\alpha} \phi, \end{aligned} \quad (\text{A2})$$

where  $\partial \mathcal{V}$  is the boundary surface of the full spacetime volume  $\mathcal{V}$  described by some arbitrary scalar function,  $\phi(x) = \text{constant}$ . By performing integration by parts, the above expression for the variation of the action can be further simplified and it will yield several boundary terms. Since these boundary terms will not play any significant role immediately, we will neglect all the boundary contributions, and hence, the above variation of the action functional can be expressed in the following manner:

$$\begin{aligned} \Delta \mathcal{A} = & \int_{\mathcal{V}} d^4x \sqrt{-\bar{g}} [\Delta \chi^{\beta} \bar{g}_{\nu\beta} (\bar{\square} \chi^{\nu} + [\bar{\nabla}_{\mu}, \bar{\nabla}^{\nu}] \chi^{\mu} + \bar{\nabla}^{\nu} \bar{\nabla}_{\mu} \chi^{\mu}) - \Delta \chi^{\beta} \bar{\nabla}_{\beta} (\bar{\mu} \bar{\nabla}_{\alpha} \chi^{\alpha})] \\ = & \int_{\mathcal{V}} d^4x \sqrt{-\bar{g}} \bar{g}_{\nu\beta} \Delta \chi^{\beta} [\bar{\square} \chi^{\nu} + \bar{R}^{\nu}_{\mu} \chi^{\mu} + \bar{\nabla}^{\nu} \{(1 - \bar{\mu}) \bar{\nabla}_{\alpha} \chi^{\alpha}\}]. \end{aligned} \quad (\text{A3})$$

Here we have used the fact that the commutator of covariant derivatives acting on a vector is given by the Riemann tensor. Thus, setting the variation of the action functional  $\Delta \mathcal{A}$  to be zero, for arbitrary variations of the GKV field  $\chi^{\beta}$ , we obtain

$$\sqrt{-\bar{g}} \bar{g}_{\nu\beta} [\bar{\square} \chi^{\nu} + \bar{R}^{\nu}_{\mu} \chi^{\mu} + \bar{\nabla}^{\nu} \{(1 - \bar{\mu}) \bar{\nabla}_{\alpha} \chi^{\alpha}\}] = 0. \quad (\text{A4})$$

Since we are interested in nondegenerate spacetime, i.e., spacetimes with a metric  $\bar{g}_{\alpha\beta}$  with nonzero determinant and nontrivial inverse, the above equation can be casted in the following form:

$$\bar{\square}\chi^\nu + \bar{R}^\nu{}_\mu\chi^\mu + \bar{\nabla}^\nu\{(1-\bar{\mu})\bar{\nabla}_\sigma\chi^\sigma\} = 0. \quad (\text{A5})$$

The above equation corresponds to the AKE satisfied by the GKV  $\chi^\mu$  in the exact spacetime, with metric  $\bar{g}_{\mu\nu}$ . Note that, in the above expression we have kept  $\bar{\mu}$  inside the derivative terms since it is a function of the spacetime coordinates.

- 
- [1] R. A. Matzner, *J. Math. Phys. (N.Y.)* **9**, 1657 (1968); C. Beetle, arXiv:0808.1745; C. Beetle and S. Wilder, *Classical Quantum Gravity* **31**, 075009 (2014).
- [2] J. Brunekreef and M. Reitz, arXiv:2012.14518.
- [3] D. Garfinkle and C. Gundlach, *Classical Quantum Gravity* **16**, 4111 (1999).
- [4] A. I. Harte, *Classical Quantum Gravity* **25**, 205008 (2008).
- [5] C. H. Taubes, *J. Math. Phys. (N.Y.)* **19**, 1515 (1978).
- [6] C. Bona, J. Carot, and C. Palenzuela-Luque, *Phys. Rev. D* **72**, 124010 (2005).
- [7] A. Komar, *Phys. Rev.* **113**, 934 (1959).
- [8] A. Komar, *Phys. Rev.* **127**, 1411 (1962).
- [9] M. Ruiz, C. Palenzuela, and C. Bona, *Phys. Rev. D* **89**, 025011 (2014).
- [10] J. C. Feng, *Phys. Rev. D* **98**, 104035 (2018).
- [11] J.-J. Peng, *Commun. Theor. Phys.* **72**, 065402 (2020); J.-J. Peng and C.-L. Zou, arXiv:1910.00339.
- [12] J. C. Feng, E. Gasperín, and J. L. Williams, *Phys. Rev. D* **100**, 124034 (2019).
- [13] J. Katz, *Classical Quantum Gravity* **2**, 423 (1985); D. Bak, D. Cangemi, and R. Jackiw, *Phys. Rev. D* **49**, 5173 (1994); J. Katz, J. Bičák, and D. Lynden-Bell, *Phys. Rev. D* **55**, 5957 (1997); Y. N. Obukhov and G. F. Rubilar, *Phys. Rev. D* **73**, 124017 (2006); **74**, 064002 (2006).
- [14] N. Deruelle, J. Katz, and S. Ogushi, *Classical Quantum Gravity* **21**, 1971 (2004); Y. N. Obukhov, G. F. Rubilar, and J. G. Pereira, *Phys. Rev. D* **74**, 104007 (2006); Y. N. Obukhov and D. Puetzfeld, *Phys. Rev. D* **87**, 081502 (2013); **90**, 024004 (2014); Y. N. Obukhov, F. Portales-Oliva, D. Puetzfeld, and G. F. Rubilar, *Phys. Rev. D* **92**, 104010 (2015); J. Schmidt and J. Bičák, *J. Math. Phys. (N.Y.)* **59**, 042501 (2018).
- [15] V. Iyer and R. M. Wald, *Phys. Rev. D* **50**, 846 (1994).
- [16] T. Jacobson and A. Mohd, *Phys. Rev. D* **92**, 124010 (2015).
- [17] K. Prabhu, *Classical Quantum Gravity* **34**, 035011 (2017).
- [18] S. Chakraborty and R. Dey, *Phys. Lett. B* **786**, 432 (2018).
- [19] P. Aneesh, S. Chakraborty, S. J. Hoque, and A. Virmani, *Classical Quantum Gravity* **37**, 205014 (2020).
- [20] T. Padmanabhan, *Gen. Relativ. Gravit.* **46**, 1673 (2014).
- [21] S. Chakraborty and T. Padmanabhan, *Phys. Rev. D* **92**, 104011 (2015).
- [22] S. Chakraborty, *J. High Energy Phys.* 08 (2015) 029.
- [23] <https://github.com/justincfeng/mathematica-files/blob/a7f627be270e2de69ea9e82edbbe6a1e3e640d8b/2020/AKEPert.nb>.
- [24] J. C. Feng and R. A. Matzner, *Gen. Relativ. Gravit.* **50**, 99 (2018); E. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Krieger, Melbourne, FL, 1983); R. A. Matzner and L. C. Shepley, *Classical Mechanics* (Prentice-Hall, Englewood Cliffs, NJ, 1991); P. Weiss, *Proc. R. Soc. A* **156**, 192 (1936).
- [25] C. Gundlach and J. M. Martín-García, *Classical Quantum Gravity* **23**, S387 (2006).
- [26] D. Hilditch, *Int. J. Mod. Phys. A* **28**, 1340015 (2013).
- [27] D. Lynden-Bell and J. Bičák, *Phys. Rev. D* **96**, 104053 (2017).
- [28] S. Chakraborty and T. Padmanabhan, *Phys. Rev. D* **101**, 064023 (2020).
- [29] J. M. Martín-García, xAct: tensor computer algebra, 2020, <http://www.xact.es/>.