Gauge-invariant variables in general-relativistic perturbations
— globalization and zero-mode problem —

Kouji Nakamura*

TAMA Project, Optical and Infrared Astronomy Division,
National Astronomical Observatory of Japan,
Osawa 2-21-1, Mitaka 181-8588, Japan

(Dated: March 31, 2012)

An outline of a proof of the local decomposition of linear metric perturbations into gauge-invariant and gauge-variant parts on an arbitrary background spacetime is briefly explained. We explicitly construct the gauge-invariant and gauge-variant parts of the linear metric perturbations based on some assumptions. We also point out the zero-mode problem is an essential problem to globalize of this decomposition of linear metric perturbations. The resolution of this zero-mode problem implies the possibility of the development of the higher-order gauge-invariant perturbation theory on an arbitrary background spacetime in a global sense.

1. Introduction — Higher-order general-relativistic perturbation theory is one of topical subject in recent general relativity. As well-known, general relativity is based on the concept of general covariance. Due to this general covariance, the “gauge degree of freedom”, which is unphysical degree of freedom of perturbations, arises in general-relativistic perturbations. To obtain physical results, we have to fix this gauge degrees of freedom or to extract some invariant quantities of perturbations. This situation becomes more complicated in higher-order perturbation theory. Therefore, it is worthwhile to investigate higher-order gauge-invariant perturbation theory from a general point of view.

According to this motivation, in Ref. [1], we proposed a procedure to find gauge-invariant variables for higher-order perturbations on an arbitrary background spacetime. This proposal is based on the single assumption (Conjecture .1 in this article). Under this assumption, we summarize some formulae for the second-order perturbations of the curvatures and energy-momentum tensor for matter fields [2, 3]. Confirming this assumption in cosmolog-

* e-mail:kouji.nakamura@nao.ac.jp
Gauge-invariant variables in general-relativistic perturbations

General perturbations, the second-order gauge-invariant cosmological perturbation theory was developed [4, 5]. Through these works, we find that our general framework of higher-order gauge-invariant perturbation theory is well-defined except for the above assumption. Therefore, we proposed the above assumption as a conjecture in Ref. [3]. We also proposed a brief outline of a proof of this conjecture [6, 7].

However, in the outline of a proof in Ref. [7], special modes of perturbations are not included in our considerations. We called these special modes as zero modes in Ref. [7]. Through the above proposal of our outline of a proof, we also pointed out that the zero modes may appear in perturbation theories on an arbitrary background spacetime. We called issues concerning about these zero modes as zero-mode problem in Ref. [7]. At least in the current status, this zero-mode problem is not resolved, yet. However, we expect that the zero modes play important roles in some situations. The purpose of this article is not to resolve this zero-mode problem, but to point out an important role of the zero modes, which is related to the globalization of perturbations.

2. Perturbations in general relativity — The notion of “gauge” in general relativity arise in the theory due to the general covariance. There are two kinds of “gauges” in general relativity. These two “gauges” are called as the first- and the second-kind gauges, respectively. The distinction of these two different notion of “gauges” is an important premise of our arguments. The first-kind gauge is a coordinate system on a single manifold $\mathcal{M}$. The coordinate transformation is also called gauge transformation of the first kind in general relativity. On the other hand, the second-kind gauge appears in perturbation theories in any theory with general covariance. In perturbation theories, we always treat two spacetime manifolds. One is the physical spacetime $\mathcal{M}$ which is our nature itself and we want to clarify the properties of $\mathcal{M}$ through perturbations. Another is the background spacetime $\mathcal{M}_0$ which has nothing to do with our nature but is prepared by hand for perturbative analyses. The gauge choice of the second kind is the point identification map $\mathcal{X} : \mathcal{M}_0 \mapsto \mathcal{M}$. We have to note that the correspondence $\mathcal{X}$ between points on $\mathcal{M}_0$ and $\mathcal{M}$ is not unique in the perturbation theory with general covariance, i.e., we have no guiding principle to choose the identification map $\mathcal{X}$. Actually, as a gauge choice of the second kind, we may choose a different point identification map $\mathcal{Y}$ from $\mathcal{X}$. This implies that there is degree of freedom in the gauge choice of the second kind. This is the gauge degree of freedom of the second kind in general-relativistic perturbations. The gauge transformation of the second kind is
To define perturbations of an arbitrary tensor field $\bar{Q}$, we have to compare $\bar{Q}$ on the physical spacetime $M_\lambda$ with $Q_0$ on the background spacetime $M_0$ through the introduction of the above second-kind gauge choice $X_\lambda : M_0 \to M_\lambda$. The pull-back $X_\lambda^* \bar{Q}$, which is induced by the map $X_\lambda$, maps a tensor field $\bar{Q}$ on $M_\lambda$ to a tensor field $X_\lambda^* \bar{Q}$ on $M_0$. Once the definition of the pull-back of the gauge choice $X_\lambda$ is given, the perturbations of a tensor field $\bar{Q}$ under the gauge choice $X_\lambda$ are simply defined by the evaluation of the Taylor expansion at $M_0$:

$$X \hat{Q} := X_\lambda^* \bar{Q}|_{M_0} = Q_0 + \lambda \frac{(1)Q}{\lambda} + \frac{1}{2} \lambda^2 \frac{(2)Q}{\lambda} + O(\lambda^3),$$

(1)

where $(1)Q$ and $(2)Q$ are the first- and the second-order perturbations of $\bar{Q}$, respectively.

When we have two different gauge choices $X_\lambda$ and $Y_\lambda$, we have two different representations of the perturbative expansion (1). Although these two representations are different from each other, these should be equivalent because of general covariance. This equivalence is guaranteed by the gauge-transformation rules between these two gauge choices. The change of the gauge choice from $X_\lambda$ to $Y_\lambda$ is represented by the diffeomorphism $\Phi_\lambda := (X_\lambda)^{-1} \circ Y_\lambda$. This diffeomorphism $\Phi_\lambda$ is the map $\Phi_\lambda : M_0 \to M_0$ for each value of $\lambda \in \mathbb{R}$ and does change the point identification. The gauge transformation $\Phi_\lambda$ induces a pull-back from the representation $X_\lambda Q_\lambda$ in the gauge choice $X_\lambda$ to the representation $Y_\lambda Q_\lambda$ in the gauge choice $Y_\lambda$ by $Y_\lambda Q_\lambda = \Phi_\lambda^* X_\lambda Q_\lambda$. According to generic arguments concerning the Taylor expansion of the pull-back of tensor fields on the same manifold [5], we obtain the order-by-order gauge-transformation rules for the perturbative variables $(1)Q$ and $(2)Q$ as

$$Y_\lambda (1)Q - X_\lambda (1)Q = \xi_{\xi (1)} Q_0, \quad Y_\lambda (2)Q - X_\lambda (2)Q = 2 \xi_{\xi (2)} (1)Q + \left\{ \xi_{\xi (1)} + \xi^2_{\xi (1)} \right\} Q_0,$$

(2)

where $\xi^a_{\xi (1)}$ and $\xi^a_{\xi (2)}$ are the generators of the gauge transformation $\Phi_\lambda$.

The notion of gauge invariance considered in this article is the order-by-order gauge invariance proposed in Ref. [3]. We call the $k$th-order perturbation $(k)Q$ is gauge invariant iff $(k)Q = (k)Q$ for any gauge choice $X_\lambda$ and $Y_\lambda$. Through this concept of the order-by-order gauge invariance, we can develop the gauge-invariant perturbation theory.

3. Construction of gauge-invariant variables — To construct gauge-invariant variables, we first consider the metric perturbation. The metric $\bar{g}_{ab}$ on $M_\lambda$, which is pulled back to $M_0$ using a gauge choice $X_\lambda$, is expanded as Eq. (1):

$$X_\lambda^* \bar{g}_{ab} = g_{ab} + \lambda h_{ab} + \frac{1}{2} \lambda^2 \bar{h}_{ab} + O^3(\lambda),$$

(3)
where $g_{ab}$ is the metric on $\mathcal{M}_0$. Although this expansion depends entirely on the gauge choice $X_\lambda$, henceforth, we do not explicitly express the index of the gauge choice $X_\lambda$ if there is no possibility of confusion. Through these setup, in Ref. [1], we proposed a procedure to construct gauge-invariant variables for higher-order perturbations. Our starting point is the following conjecture for $h_{ab}$:

**Conjecture 1** If there is a symmetric tensor field $h_{ab}$ of the second rank, whose gauge transformation rule is $\gamma h_{ab} - \chi h_{ab} = L_{\xi(1)} g_{ab}$, then there exist a tensor field $H_{ab}$ and a vector field $X^a$ such that $h_{ab}$ is decomposed as $h_{ab} = H_{ab} + \xi(1)$ under the gauge transformation (2), respectively.

In this conjecture, $H_{ab}$ and $X^a$ are *gauge-invariant* and *gauge-variant* parts of the perturbation $h_{ab}$. In the case of the perturbation theory on an arbitrary background spacetime, this conjecture is a highly non-trivial statement due to the non-trivial curvature of the background spacetime, though its inverse statement is trivial.

Based on Conjecture 1, we can decompose the second-order metric perturbation $l_{ab}$ as [1]

$$l_{ab} = L_{ab} + 2L_X h_{ab} + \left(L_Y - L_X^2\right) g_{ab},$$

(4)

where $L_{ab}$ is gauge-invariant part of the second-order metric perturbation $l_{ab}$ and $Y^a$ is the gauge-variant part of second order whose gauge-transformation rule is given by $\gamma Y^a - \chi Y^a = \xi(2) + [\xi(1), X]^a$. Furthermore, using the first- and second-order gauge-variant parts, $X^a$ and $Y^a$, of the metric perturbations, the gauge-invariant variables for an arbitrary tensor field $Q$ are given by

$$^{(1)}Q := Q - L_X Q_0, \quad ^{(2)}Q := (2)Q - 2L_X^{(1)} Q - \left(L_Y - L_X^2\right) Q_0.$$

(5)

In Ref. [1], we extended this construction to the third-order perturbations and we have already confirmed that this construction is valid in the fourth-order perturbations [8].

4. An outline of a proof of Conjecture 1 — To give an outline of a proof of Conjecture 1 for an arbitrary background spacetime, we assume that background spacetimes admit ADM decomposition. Therefore, the background spacetime $\mathcal{M}_0$ considered here is $n + 1$-dimensional spacetime which is described by the direct product $\mathbb{R} \times \Sigma$. Here, $\mathbb{R}$ is a time direction and $\Sigma$ is the spacelike hypersurface $(\dim \Sigma = n)$ embedded in $\mathcal{M}_0$. This means
that $\mathcal{M}_0$ is foliated by the one-parameter family of spacelike hypersurface $\Sigma(t)$, where $t \in \mathbb{R}$ is a time function. Then, the metric on $\mathcal{M}_0$ is described by the ADM decomposition

$$g_{ab} = -\alpha^2(dt)_a(dt)_b + q_{ij}(dx^i + \beta^i dt)_a(dx^j + \beta^j dt)_b,$$

where $\alpha$ is the lapse function, $\beta^i$ is the shift vector, and $q_{ij}$ is the metric on $\Sigma(t)$.

Since the ADM decomposition (6) is a local one, we may regard that our arguments are restricted to that for a single patch in $\mathcal{M}_0$ which is covered by the metric (6). Therefore, we regard $\Sigma$ as this single patch of a spacelike hypersurface in $\mathcal{M}_0$. Further, we may change the region which is covered by the metric (6) through the choice of the lapse function $\alpha$ and the shift vector $\beta^i$. The choice of $\alpha$ and $\beta^i$ is regarded as the first-kind gauge choice, which have nothing to do with the second-kind gauge. Since we may regard that the representation (6) of the background metric is that on a single patch in $\mathcal{M}_0$, in general situation, each $\Sigma$ may have its boundaries $\partial \Sigma$.

To prove Conjecture 1, we consider the components of the metric $h_{ab}$ as $h_{ab} = h_{tt}(dt)_a(dt)_b + 2h_{ti}(dt)_a(dx^i)_b + h_{ij}(dx^i)_a(dx^j)_b$. The gauge-transformation rules for the components $\{h_{tt}, h_{ti}, h_{ij}\}$ are derived from $\chi h_{ab} - \chi h_{ab} = \mathcal{L}_{\xi(t)}g_{ab}$ with $\xi(t)_a = \xi(t)_a + \xi(t)(dx^i)_a$. Inspecting these gauge-transformation rules, we explicitly construct gauge-invariant and gauge-variant variables.

Our strategy for the proof is as follows [6, 7]: we first assume that the existence of the variables $X^i$ and $X_i$ whose gauge-transformation rules are given by $\gamma X^i - \chi X^i = \xi_i$ and $\gamma X_i - \chi X_i = \xi_i$, respectively. This assumption is confirmed through the explicit construction of the gauge-variant part of the linear-order metric perturbation below. Further, inspecting gauge-transformation rules for the components $\{h_{tt}, h_{ti}, h_{ij}\}$, we define the symmetric tensor field $\hat{H}_{ab}$ whose components are given by

$$\hat{H}_{tt} := h_{tt} + \frac{2}{\alpha} (\partial_t \alpha + \beta^i D_i \alpha - \beta^j \beta^i K_{ij}) X_t + \frac{2}{\alpha^2} \left(\beta^i \beta^j \beta^k K_{kij} - \beta^i \partial_t \alpha + \alpha q^j \partial_i \beta_j + \beta^j \beta^k D^j \beta_k - \beta^i \beta^j D_j \alpha\right),$$

$$\hat{H}_{ti} := h_{ti} + \frac{2}{\alpha} (D_i \alpha - \beta^j K_{ij}) X_t + \frac{2}{\alpha} M_{ij} X_j,$$

$$\hat{H}_{ij} := h_{ij} - \frac{2}{\alpha} K_{ij} X_t + \frac{2}{\alpha} \beta^k K_{ij} X_k,$$

where $M_{ij}$ is defined by $M_{ij} := -\alpha^2 K_{ij} + \beta^j \beta^k K_{ki} - \beta^j D_i \alpha + \alpha D_i \beta^j$. Here, $K_{ij}$ is the components of the extrinsic curvature of $\Sigma$ in $\mathcal{M}_0$ and $D_i$ is the covariant derivative associate
with the metric \( q_{ij} \) \( (D_i q_{jk} = 0) \). The extrinsic curvature \( K_{ij} \) is related to the time derivative of the metric \( q_{ij} \) by \( K_{ij} = -\ (1/2\alpha) \ [\partial_t q_{ij} - D_i \beta_j - D_j \beta_i] \). The gauge transformation rules for the components of \( \hat{H}_{ab} \) are given by

\[
\gamma \hat{H}_{tt} - \chi \hat{H}_{tt} = 2 \partial_t \xi_t, \quad \gamma \hat{H}_{ti} - \chi \hat{H}_{ti} = \partial_t \xi_i + D_i \xi_t, \quad \gamma \hat{H}_{ij} - \chi \hat{H}_{ij} = 2 D_t (\xi_j).
\] (10)

Since the components \( \hat{H}_{tt} \) and \( \hat{H}_{ij} \) are regarded as components of a vector and a symmetric tensor on \( \Sigma \), respectively, we may apply the following decomposition [9] to \( \hat{H}_{tt} \) and \( \hat{H}_{ij} \):

\[
\hat{H}_{ti} := D_i h_{(V) i}, \quad D^i h_{(V)i} = 0,
\] (11)

\[
\hat{H}_{ij} := \frac{1}{n} q_{ij} h_{(L)} + h_{(T)ij}, \quad q^{ij} h_{(T)ij} = 0,
\] (12)

\[
h_{(T)ij} := (Lh_{(TV)})_{ij} + h_{(TT)ij}, \quad D^i h_{(TT)ij} = 0,
\] (13)

where \( (Lh_{(TV)})_{ij} \) is defined by \( (Lh_{(TV)})_{ij} := D_i h_{(TV)j} + D_j h_{(TV)i} - \frac{2}{n} q_{ij} D^k h_{(TV)k} \). Equations (10) give the gauge-transformation rules for the variables \( h_{(V)L}, h_{(V)i}, h_{(L)}, h_{(T)ij}, h_{(TT)ij} \), and \( h_{(TT)ij} \). From these gauge-transformation rules, we easily find the explicit form of the variables \( X_t \) and \( X_i \) as follows [6, 7]:

\[
X_t := h_{(TV)i}, \quad X_i := h_{(VL)} - \Delta^{-1} D^k \partial_t h_{(TV)k}.
\] (14)

This is the most non-trivial part in our outline of a proof of Conjecture 1. Further, we have to note that in the derivation of Eqs. (14), we assume that the existence of the Green functions \( \Delta^{-1} \) and \( \mathcal{D}^{ij} \) of the Laplacian \( \Delta := D^i D_i \) and the elliptic derivative operator \( \mathcal{D}^{ij} \) defined by \( \mathcal{D}^{ij} := q^{ij} \Delta + (1 - \frac{2}{n}) D^i D^j + R^{ij} \). Here, \( R^{ij} \) is the Ricci curvature on \( \Sigma \). In other words, we have ignored the perturbative modes which belong to the kernel of the derivative operators \( \Delta \) and \( \mathcal{D}^{ij} \) [6, 7]. We call these modes as zero modes.

Furthermore, we easily construct gauge-invariant variables for the linear-order metric perturbation \( h_{ab} \). We have two scalar modes \( (\Phi \text{ and } \Psi) \), one transverse vector mode \( \nu_i \), one transverse-traceless tensor mode \( \chi_{ij} \). These gauge-invariant variables are given by

\[
-2\Phi := \hat{H}_{tt} - 2 \partial_t X_t, \quad -2n \Psi := h_{(L)} - 2 D^i X_i,
\]

\[
\nu_i := h_{(V)i} - \partial_t X_i + D_i \Delta^{-1} D^k \partial_t X_k, \quad \chi_{ij} := h_{(TT)ij}.
\] (15)

Moreover, we can derive the expressions of the original components \( \{h_{tt}, h_{ti}, h_{ij}\} \) of the metric perturbation \( h_{ab} \) in terms of these gauge-invariant variables and the variables \( X_t \) and
Then, we conclude that we may identify the components of the gauge-invariant variables \( H_{ab} \) and the gauge-variant variable \( X_a \) so that \( H_{tt} := -2\Phi, \ H_{ti} := \nu_i, \ H_{ij} := -2\Psi q_{ij} + \chi_{ij}, \ X_a := X_t(dt)_a + X_i(dx^i)_a \). These identifications lead to the assertion of Conjecture 1.

5. Zero-mode problem and the globalization of gauge-invariant variables — In the above outline of a proof of Conjecture 1, we concentrate only on a local region \( \Sigma \) in a spacelike hypersurface which is covered by the metric (6). This local region \( \Sigma \) may have its boundaries \( \partial \Sigma \). Furthermore, we assumed the existence of Green functions of the elliptic derivative operators \( \Delta \) or \( D^{ij} \). Since we concentrated only on a local region \( \Sigma \) of the whole spacelike hypersurface in the above outline of a proof, we have to discuss the globalization of our proof to the whole region of the spacelike hypersurface in the background spacetime \( M_0 \) if we insist that Conjecture 1 is true on the whole background spacetime manifold \( M_0 \). In my opinion, the key of this globalization is zero modes.

As mentioned above, we define zero modes as perturbative modes which belongs to the kernel of the elliptic derivative operators \( \Delta \) or \( D^{ij} \). The kernel of \( D^{ij} \) also includes the (conformal) Killing vectors. Therefore, we may say that zero modes are related to the symmetries of the background spacetime. Furthermore, we should emphasize that we have to impose boundary conditions at boundaries \( \partial \Sigma \) for the explicit construction of the Green functions for \( \Delta \) and \( D^{ij} \). Since the operators \( \Delta \) and \( D^{ij} \) are elliptic, the change of the boundary conditions at \( \partial \Sigma \) is adjusted by functions which belong to the kernel of the operators \( \Delta \) and \( D^{ij} \), i.e., zero modes. Thus, we may say that the informations for the boundary conditions for the Green functions \( \Delta^{-1} \) and \( (D^{ij})^{-1} \) are also included in the zero modes. These modes should be separately treated in different manner. We call the issue concerning about treatments of these zero modes as the zero-mode problem. This problem is a remaining problem in our general framework on higher-order general-relativistic gauge-invariant perturbation theory.

Now, we comment on the relation between the globalization of our outline of a proof and the zero-mode problem. If we want to consider the global behaviors of perturbations, we have to consider the globalization of the definition of the gauge-invariant and gauge-variant variables to the whole region of a spacelike hypersurface in \( M_0 \). To do this, we should consider different patches which cover the region outside the local region \( \Sigma \). The same arguments as above is applied to perturbative variables on these different patches. The key problem is how to identify the gauge-invariant and gauge-variant variables for perturbations on these different patches to those on \( \Sigma \). To accomplish this, the behavior of the perturbative
variables at the boundaries $\partial \Sigma$ is important. If we impose some smoothness of the gauge-invariant and gauge-variant variables on the whole region of a spacelike hypersurface, we have to impose an appropriate boundary conditions to the perturbative variables at $\partial \Sigma$ and to match the perturbative variables at $\partial \Sigma$ with those on the region outside $\Sigma$. If we accomplish this matching, we may regard that the gauge-invariant and gauge-variant variables are global variables on the whole region of a spacelike hypersurface on $\Sigma$. As mentioned above, the change of the boundary conditions at $\partial \Sigma$ is adjusted by zero modes. Thus, zero modes is necessary to construct global gauge-invariant and gauge-variant variables.

If we impose the Einstein equation as the field equation, this globalization is related to the construction of a global solution to the perturbative Einstein equation. To construct perturbative solutions to the initial value constraints, we have to consider the perturbative solutions outside the local region $\Sigma$ and to match with the perturbative solutions in $\Sigma$ at $\partial \Sigma$. To accomplish this matching, each solution to the initial value constraints will be required to satisfy some appropriate boundary conditions at the boundaries $\partial \Sigma$. The boundary behavior of the perturbative variables is also adjusted by the zero mode which satisfy the perturbative Einstein equation. If we accomplish this matching of solutions to the initial value constraints smoothly, we can consider the time evolution of the global perturbations following to the evolution equations in the Einstein equation.

We also note that to ignore the zero modes is regarded as to impose boundary conditions for the Green functions $\Delta$ or $D^{ij}$ at the boundary $\partial \Sigma$ in some way. There is no guarantee whether this boundary condition at $\partial \Sigma$ is appropriate to construct global solution to the perturbative Einstein equation or not. The information at $\partial \Sigma$ propagates along the boundary of the domain of dependence of $\Sigma$ through the dynamics of the Einstein equation. If the imposed boundary conditions are not appropriate to construct global solutions, the obtained solution to perturbative Einstein equation cannot be extend to the outside of the domain of dependence of $\Sigma$ in general and loses its physical relevance of the behavior at the boundary of the domain of dependence of $\Sigma$. In this sense, zero modes are the important to construct global perturbative solutions. This is the main point of this article.

6. Summary — We briefly explained our proposal of an outline of a proof Conjecture 1 for an arbitrary background spacetime. Although there will be many approaches to prove Conjecture 1, in this article, we just show an outline a proof. We also note that our arguments do not include zero modes. The existence of zero modes is also related to the
symmetries of the background spacetime. Furthermore, the zero modes are also important to construct global gauge-invariant and gauge-variant variables of perturbations and to derive global solutions to the perturbative Einstein equations. To resolve this zero-mode problem, careful discussions on domains of functions for perturbations and its boundary conditions at $\partial \Sigma$ will be necessary. If we resolved this zero-mode problem, the general framework of the general-relativistic higher-order gauge-invariant perturbation theory will be completed and the wide applications of this general framework will be opened.

[1] K. Nakamura, Prog. Theor. Phys. 110, (2003), 723.
[2] K. Nakamura, Prog. Theor. Phys. 113 (2005), 481.
[3] K. Nakamura, Phys. Rev. D 80 (2009), 124021.
[4] K. Nakamura, Phys. Rev. D 74 (2006), 101301(R); K. Nakamura, Prog. Theor. Phys. 117 (2007), 17; K. Nakamura, Prog. Theor. Phys. 121 (2009), 1321;
[5] K. Nakamura, Advances in Astronomy, 2010 (2010), 576273.
[6] K. Nakamura, Class. Quantum Grav. 28 (2011), 122001.
[7] K. Nakamura, arXiv:1105.4007 [gr-qc].
[8] K. Nakamura, in progress.
[9] J. W. York, Jr. J. Math. Phys. 14 (1973), 456; Ann. Inst. H. Poincaré 21 (1974), 319. S. Deser, Ann. Inst. H. Poincaré 7 (1967), 149.