An analysis of characteristics in nonlinear massive gravity

Keisuke Izumi[1] **and Yen Chin Ong**[1,2]

[1] Leung Center for Cosmology and Particle Astrophysics, National Taiwan University, Taipei, Taiwan 10617
[2] Graduate Institute of Astrophysics, National Taiwan University, Taipei, Taiwan 10617

E-mail: izumi@phys.ntu.edu.tw and ongyenchin@member.ams.org

Abstract

We study the Cauchy problem in a special case of nonlinear massive gravity.
Despite being ghost free, it has recently been argued that the theory is
inherently problematic due to the existence of superluminal shock waves.
Furthermore, it is claimed that an acausal characteristic can arise for *any*
choice of background. In order to further understand the causal structure
of the theory, we carefully perform a detailed analysis of the characteristic
equations and show that the theory does admit a well-posed Cauchy problem,
i.e., there exists hypersurfaces that are *not* a characteristic hypersurface. Puzzles
remain regarding the existence of a superluminal propagating mode in both the
minimal ghost-free theory that we analyzed, as well as in the full nonlinear
massive gravity. That is, our result should not be taken as any indication
of the healthiness of the theory. We also give a detailed review of Cauchy–
Kovalevskaya theorem and its application in the appendix, which should be
useful for investigating causal structures of other theories of gravity.

PACS number: 04.50.Kd

1. Introduction

The question as to whether graviton can have a mass is an old one. A recent review of the literature can be found in [1]. Here we give a very brief review of the main progress and major difficulties encountered in constructing a healthy massive gravity theory.

It is well known that Fierz and Pauli constructed a theory of a free massive spin-2 particle back in 1939 [2], in which the quadratic form of the action can be uniquely fixed as linearized general relativity (GR) with the Fierz–Pauli mass term, by imposing the tachyon-free and ghost-free conditions [3]. However, it was discovered in 1970 that this linear massive spin-2 theory suffers from the van Dam–Veltman–Zakharov discontinuity [4, 5], with a massless limit that does not recover that of linearized GR. In particular, the prediction of light bending near massive objects was off by 25% and is therefore completely ruled out by observations.

The *Vainshtein mechanism* was proposed in 1972, in which a nonlinear effect is introduced to force the massless limit to recover GR [6, 7]. However, in the same year it was shown that the exact nonlinearity introduced gives rise to the *Boulware–Deser ghost* or BD ghost for short [8], where the mass of the ghost mode is typically the same as that of the graviton, which means the ghost mode cannot be ignored (e.g., by pushing to the Planck scale to be dealt with by a quantum theory of gravity [9]) since its mass is expected to be *small*. The BD ghost is essentially due to an extra degree of freedom reinstated by the nonlinearity, beyond the linearized five degrees of freedom. The interest in massive gravity was revived recently due to the construction of a presumably consistent nonlinear massive gravity without the BD ghost [10, 11]. The absence of the BD ghost was checked at a low-order perturbative analysis level in the metric perturbation approach [12, 13], and subsequently even at fully nonlinear orders [14–17]. It is interesting to note that, as pointed out in [18], Wess–Zumino already constructed a theory without the sixth degree of freedom and thus free of the BD ghost in 1970 [19]. However, it should be pointed out that this is a realization by hindsight: Wess–Zumino had simply presented a number of massive gravity models, the linear Fierz–Pauli theory, without calculating the number of degrees of freedom, and without considering the issue of ghosts. We shall henceforth refer to such presumably consistent theories as nonlinear massive gravity in general. In nonlinear massive gravity, in addition to the dynamical metric $g_{\mu\nu}$, there is also a fixed background with a non-dynamical metric $\bar{g}_{\mu\nu}$, also known as the *fiducial metric*.

Recently, the characteristic equations of a special case of nonlinear massive gravity, essentially one of the original model of Wess–Zumino with five degrees of freedom, were analyzed by Deser and Waldron [18] (see also the newer work [20]). It was argued that the model admits superluminal shock wave solutions and therefore is acausal, and in addition, the superluminal shocks are only absent in the special case where the contortion term vanishes. Furthermore, ironically this originates from the very constraint that removes the BD ghost. In the same work, it is mentioned that this fatal problem arises in *any* choice of background. This is an intriguing claim, which if true, is a rather strong one. After all, we do know that at least some physics *does* depend on the choice of background. For example, a Friedmann–Lemaître–Robertson–Walker (FLRW) background does not admit the flat or closed cosmology[3] [21], however more general fiducial metrics do allow FLRW universes of all signs of spatial curvature [22]. Motivated by the interesting work of [18], we perform a careful analysis of the characteristic equations to further understand the causal structure of nonlinear massive gravity.

In section 2, we introduce the necessary setup. Section 3 contains the main analysis of the characteristic equations, while section 4 gives a concrete example to show a well-posed Cauchy problem in the theory. We conclude with a discussion on the existence of superluminal propagation in nonlinear massive gravity in section 5. The method of characteristic employed in our work is reviewed in detail in appendix.

2. Nonlinear massive gravity theory

As mentioned in the introduction, in nonlinear massive gravity, we need to first choose and fix a background with a non-dynamical reference metric $\bar{g}_{\mu\nu}$, in addition to the physical metric $g_{\mu\nu}$. For convenience of our analysis, we will work in the tetrad formalism [23–27], following the work of [18] [4]. We denote the background tetrad by $f^a{}_\mu$, and the dynamical tetrad by $e^a{}_\mu$. Here, Greek letters correspond to spacetime indices, while Latin letters label tetrad vectors. The theory is also referred to as the two-metric '*f-g*' theory in [18]. However, since there is

---

[3] This is also caused by the very constraint that removes the BD ghost.
[4] The massive gravity theory in the tetrad formalism is, at least in classical physics, the same as the nonlinear massive gravity proposed in [12–14, 17].

a risk of confusion with the bimetric theory, in which both metrics are dynamical, we will refer to this model as the *minimal ghost-free model* instead, since this is the simplest of the ghost-free models in the formulation of [29].

The Levi-Civita connection $\omega(e)_\mu{}^m{}_n$ that corresponds to the dynamical tetrad is torsionless. However, the same connection employed as the connection for the background tetrad has nonzero torsion in general. The difference of the two connections is the contortion tensor[5]:

$$K_\mu{}^m{}_n := \omega(e)_\mu{}^m{}_n - \omega(f)_\mu{}^m{}_n. \tag{1}$$

The background metric $\bar{g}_{\mu\nu}$ and the dynamical metric $g_{\mu\nu}$ are related to the background tetrad and the dynamical tetrad, respectively, by

$$f^a{}_\mu f^b{}_\nu \eta_{ab} = \bar{g}_{\mu\nu} \qquad \text{and} \qquad e^a{}_\mu e^b{}_\nu \eta_{ab} = g_{\mu\nu}. \tag{2}$$

Raising and lowering of the Greek indices are performed via the dynamical metric, i.e.,

$$V^\mu = V_\nu g^{\nu\mu}, \quad \text{and} \quad V_\mu = V^\nu g_{\nu\mu}. \tag{3}$$

The action of the nonlinear massive gravity theory is defined as [29]

$$S = \frac{1}{2\kappa} \int d^4x \, e \left(R + 2 \sum_{n=0}^{4} \alpha_n \mathcal{L}_n \right), \tag{4}$$

in the unit $c = 1$, where α_n are constant parameters of the theory and the exact expressions of $\mathcal{L}_n$ are

$$\mathcal{L}_0 = 1, \tag{5}$$

$$\mathcal{L}_1 = f, \tag{6}$$

$$\mathcal{L}_2 = f^2 - f^\mu{}_\nu f^\nu{}_\mu, \tag{7}$$

$$\mathcal{L}_3 = f^3 - 3f f^\mu{}_\nu f^\nu{}_\mu + 2f^\mu{}_\nu f^\nu{}_\lambda f^\lambda{}_\mu, \tag{8}$$

$$\mathcal{L}_4 = f^4 - 6f^2 f^\mu{}_\nu f^\nu{}_\mu + 8f f^\mu{}_\nu f^\nu{}_\lambda f^\lambda{}_\mu + 3\left(f^\mu{}_\nu f^\nu{}_\mu\right)^2 - 6f^\mu{}_\nu f^\nu{}_\lambda f^\lambda{}_\delta f^\delta{}_\mu. \tag{9}$$

Here, $f_{\mu\nu} = e^a{}_\mu f^b{}_\nu \eta_{ab}$ and $f = f^\mu{}_\mu$.

For simplicity, we consider the case where $\alpha_2 = \alpha_3 = \alpha_4 = 0$. The equation of motion can be obtained by the variation with respect to the dynamical tetrad $e^a{}_\mu$:

$$\mathcal{G}_{\mu\nu} := G_{\mu\nu} - \alpha_0 g_{\mu\nu} + \alpha_1 (f_{\nu\mu} - f g_{\mu\nu}) = 0, \tag{10}$$

where $G_{\mu\nu}$ is the Einstein tensor. Here, in order to have a simpler form of equation of motion, we have acted upon the expression obtained after taking variation with respect to $e^a{}_\mu$ by $-e_{a\mu} g_{\alpha\nu}$. We emphasize that $f_{\mu\nu}$ has all of the information of the dynamical tetrad $e^a{}_\mu$, and thus we analyze the dynamics of $f_{\mu\nu}$ instead of that of $e^a{}_\mu$. From the antisymmetric components of equation (10), we find that $f_{\mu\nu}$ is a symmetric tensor and thus has ten degrees of freedom.

Equation (10) includes not only dynamical equations but also five constraint equations, which we shall now derive. The first two terms in equation (10) are divergence free, and thus, taking the divergence of equation (10) gives four constraint equations [30]:

$$\mathcal{C}_\mu := \nabla^\nu f_{\mu\nu} + \nabla_\mu f = 0. \tag{11}$$

Another constraint equation [16, 17] stems from the nontrivial combination of the trace of equations (10) and (11); it is

$$\mathcal{C} := 2\nabla_\mu (l^{\mu\nu} \mathcal{C}_\nu) - \mathcal{G}^\mu{}_\mu = 0, \tag{12}$$

---

[5] We would like to remark that the proper term is *contortion* instead of 'contorsion'. See e.g., [28].

where $l^{\mu\nu} = l^\mu{}_a e^{a\nu}$ and $l^\mu{}_a$ is the inverse of the background tetrad $f^a{}_\mu$. From the symmetry of the indices of $f_{\mu\nu}$, we can infer that the indices of $l^{\mu\nu}$, which is simply the inverse of $f_{\mu\nu}$, are also symmetric. In the combination (12), the terms involving second-order derivatives of $f_{\mu\nu}$ are canceled out, and it becomes a constraint equation. To sum up, the ten symmetric components of equation (10) include the five constraint equations given by equations (11) and (12), and five dynamical components containing the second-order derivatives of $f_{\mu\nu}$.

3. Characteristic in nonlinear massive gravity

Suppose we have a hypersurface Σ and ξ^μ is the unit normal vector to Σ. Here, we consider the case where ξ^μ is timelike. We use Gaussian normal coordinate near the hypersurface Σ,

$$ds^2 = -(dx^0)^2 + h_{ij}\, dx^i\, dx^j, \quad \text{with } \left(\frac{\partial}{\partial x^0}\right)^\mu = \xi^\mu. \tag{13}$$

In equation (10), all second derivatives of $f_{\mu\nu}$ are of the form $\partial_\mu \partial_{[\nu} f_{\alpha]\beta}$, where the bracket $[\ldots]$ represents antisymmetrization, i.e.,

$$T_{\alpha\cdots[\beta\gamma]\cdots} := \tfrac{1}{2}(T_{\alpha\cdots\beta\gamma\cdots} - T_{\alpha\cdots\gamma\beta\cdots}). \tag{14}$$

Having in mind the Cauchy–Kovalevskaya theorem [31, 32] (see appendix for a detailed discussion), we shall introduce new variables

$$M_{\alpha\beta\gamma} := \partial_{[\alpha} f_{\beta]\gamma}. \tag{15}$$

Since $M_{\alpha\beta\gamma}$ is antisymmetric with respect to α and β, we naively expect that $M_{\alpha\beta\gamma}$ has 24 degrees of freedom. There are, however, four constraints for $M_{\alpha\beta\gamma}$, namely,

$$\epsilon^{\mu\alpha\beta\gamma} M_{\alpha\beta\gamma} = 0, \tag{16}$$

where $\epsilon^{\mu\alpha\beta\gamma}$ is the Levi-Civita symbol. This fact can be deduced from definition (15) of $M_{\alpha\beta\gamma}$. Therefore, the actual number of degrees of freedom of $M_{\alpha\beta\gamma}$ is only 20. Under $3 + 1$ decomposition of spacetime based on the metric (13), all non-vanishing components can be written as

$$M_{0i0} = -M_{i00} = \tfrac{1}{2}(\partial_0 f_{i0} - \partial_i f_{00}), \tag{17}$$

$$M_{0ij} = -M_{i0j} = \tfrac{1}{2}\left(\partial_0 f_{ij} - \partial_i f_{0j}\right), \tag{18}$$

$$M_{ij0} = -M_{ji0} = \tfrac{1}{2}(\partial_i f_{j0} - \partial_j f_{i0}), \tag{19}$$

$$M_{ijk} = -M_{jik} = \tfrac{1}{2}(\partial_i f_{jk} - \partial_j f_{ik}). \tag{20}$$

Using the four constraints (16) for $M_{\alpha\beta\gamma}$, we can express three of M_{0ij} and one of M_{ijk} by the other $M_{\alpha\beta\gamma}$. Thus, after careful comparison it can be deduced that M_{0i0}, M_{0ij}, M_{ij0} and M_{ijk} have three, six, three and eight degrees of freedom, respectively. In addition to the four constraints (16), we can derive other relations among the $M_{\alpha\beta\gamma}$. They are:

$$\partial_0 M_{ij0} = \partial_i M_{0j0} - \partial_j M_{0i0}, \tag{21}$$

$$\partial_0 M_{ijk} = \partial_i M_{0jk} - \partial_j M_{0ik}. \tag{22}$$

Equations (17), (18), (21) and (22) can be regarded as the time-evolution equations for f_{i0}, f_{ij}, M_{ij0} and M_{ijk}, respectively. All coefficients of the terms with a x^0-derivative in these equations are obviously nonzero. Therefore, the hypersurface Σ is *not* a characteristic surface for f_{i0}, f_{ij}, M_{ij0} and M_{ijk}.

Now, in order for the time evolution to be well defined and solvable, we need to find appropriate equations to fix the time derivatives of the ten variables: f_{00}, M_{0i0} and M_{0ij}. We have the ten components of the equation of motion (10), and so, should expect that the time derivatives of these ten variables can be obtained from them.

Having introduced $M_{\alpha\beta\gamma}$, all second-order derivatives of $f_{\mu\nu}$ in the equation of motion (10) can be written in terms of the first-order derivatives of $M_{\alpha\beta\gamma}$. Then, the equation of motion (10) becomes

$$
\begin{aligned}
2\big(g^{\alpha\gamma}l^\delta{}_{(\mu}\delta^\beta{}_{\nu)} &+ l^{\beta\delta}\delta^\alpha{}_{(\mu}\delta^\gamma{}_{\nu)} - g^{\alpha\gamma}l^{\beta\delta}g_{\mu\nu}\big)\partial_\alpha M_{\beta\gamma\delta} + \bar{F}_{\mu\nu} \\
&+ l^\gamma{}_b\big(\partial_\delta f^b{}_\lambda\big)\big(l^{\lambda\beta}\delta^\alpha{}_{(\mu}\delta^\delta{}_{\nu)} - l^{\lambda\delta}\delta^\alpha{}_{(\mu}\delta^\beta{}_{\nu)} + g^{\alpha\delta}l^\lambda{}_{(\mu}\delta^\beta{}_{\nu)} \\
&\quad - g^{\alpha\beta}l^\lambda{}_{(\mu}\delta^\delta{}_{\nu)} - g^{\alpha\delta}l^{\lambda\beta}g_{\mu\nu} + g^{\alpha\beta}l^{\lambda\delta}g_{\mu\nu}\big)\partial_\alpha f_{\beta\gamma} \\
&+ N_{\delta\epsilon\lambda}\big(2g^{\alpha\gamma}l^{\beta\delta}l^\lambda{}_{(\mu}\delta^\epsilon{}_{\nu)} - 2l^{\beta\epsilon}l^{\lambda\delta}\delta^\alpha{}_{(\mu}\delta^\gamma{}_{\nu)} + 2l^{\lambda\gamma}l^{\beta\epsilon}\delta^\delta{}_{(\mu}\delta^\alpha{}_{\nu)} + 2g^{\delta\gamma}l^{\beta\epsilon}l^\lambda{}_{(\mu}\delta^\alpha{}_{\nu)} \\
&\quad - \{2g^{\alpha\gamma}l^{\beta\delta}l^{\lambda\epsilon} + g^{\delta\alpha}l^{\lambda\gamma}l^{\beta\epsilon} + g^{\delta\gamma}l^{\beta\epsilon}l^{\lambda\alpha}\}g_{\mu\nu}\big)\partial_\alpha f_{\beta\gamma} = 0,
\end{aligned}
\tag{23}
$$

where

$$
N_{\delta\epsilon\lambda} := M_{\delta\epsilon\lambda} + e_{a[\delta}\partial_{\epsilon]}f^a{}_\lambda,
\tag{24}
$$

and the brackets $(\ldots)$ denote symmetrization, i.e.,

$$
T_{\alpha\cdots(\beta\gamma)\cdots} := \tfrac{1}{2}\big(T_{\alpha\cdots\beta\gamma\cdots} + T_{\alpha\cdots\gamma\beta\cdots}\big),
\tag{25}
$$

and $\bar{F}_{\mu\nu}$ is a function of $f_{\alpha\beta}$, $M_{\mu\nu\lambda}$, $f^a{}_\alpha$, $\partial_\mu f^a{}_\alpha$ and $\partial_\mu\partial_\nu f^a{}_\alpha$. Note that derivatives of the dynamical variables $f_{\alpha\beta}$ and $M_{\mu\nu\lambda}$ do not appear in $\bar{F}_{\mu\nu}$ and that derivatives of $f^a{}_\alpha$ are not dynamical but fixed variables because $f^a{}_\alpha$ is the given background tetrad. The constraint equations (11) and (12) can be written without derivatives of the dynamical variables $f_{\alpha\beta}$ and $M_{\mu\nu\lambda}$ as

$$
l^\nu{}_\mu\mathcal{C}_\nu = 2l^{\alpha\beta}N_{\mu\alpha\beta} + 2l^\beta{}_\mu e_a{}^\alpha\partial_{[\alpha}f^a{}_{\beta]} = 0,
\tag{26}
$$

$$
\begin{aligned}
\mathcal{B} :=&\ \mathcal{C} - \mathcal{C}_\mu\bar{g}^{\mu\nu}\mathcal{C}_\nu \\
=&\ 4\alpha_0 + 3\alpha_1 f - g^{\mu\alpha}(2l^{\beta\lambda}l^{\gamma\nu} + g^{\beta\nu}\bar{g}^{\lambda\gamma})N_{\mu\nu\lambda}N_{\alpha\beta\gamma} \\
&+ e^\mu{}_a e^\nu{}_b R(f)_{\mu\nu}{}^{ab} - 2N_{\nu\rho\gamma}(2l^{\gamma\mu}l^{\beta\nu}e^{b\rho}\mathcal{F}_{\mu\beta b} + e^{b\gamma}l^{\beta\nu}l^{\rho\lambda}\mathcal{F}_{\lambda\beta b}) \\
&+ \big(2\eta^{ab}l^{\beta\nu}l^{\mu\alpha} + \bar{g}^{\nu\alpha}f^a{}_\lambda e^{b\lambda}l^{\beta\mu}\big)\mathcal{F}_{\mu\nu a}\mathcal{F}_{\alpha\beta b},
\end{aligned}
\tag{27}
$$

with

$$
\mathcal{F}_{\alpha\beta}{}^c := \partial_{[\alpha}f^c{}_{\beta]},
\tag{28}
$$

where, in accord with [18], we use the combination (27) of the constraint equations instead of $\mathcal{C}$. Here, $R(f)_{\mu\nu}{}^{ab}$ is the Riemann tensor corresponding to the background vierbein $f^a{}_\mu$. Acted upon by the differential operator ∂_0, the two constraint equations (26) and (27) now become time-evolution equations.

Now, we have equations (23), $\partial_0 l^\alpha{}_\beta\mathcal{C}_\alpha$ and $\partial_0\mathcal{B}$ which are expected to describe the time evolution of f_{00}, M_{0i0} and M_{0ij}. These equations are quasilinear[6] for the first-order derivatives of $f_{\mu\nu}$ and $M_{\mu\nu\lambda}$. Using the time-evolution equations for f_{0i}, f_{ij}, M_{ij0} and M_{ijk}, i.e., equations (17), (18), (21) and (22), we can eliminate their time derivatives from equation (23), $\partial_0 l^\alpha{}_\beta\mathcal{C}_\alpha$ and $\partial_0\mathcal{B}$, and consequently only the time derivatives of f_{00}, M_{0i0} and M_{0ij} appear in these equations. The forms of the equations become

$$
\begin{aligned}
E_{\mu\nu} :=&\ 2\big(l^{00}\delta^0{}_{(\mu}\delta^i{}_{\nu)} - g^{00}l^0{}_{(\mu}\delta^i{}_{\nu)} - l^{0i}\delta^0{}_{(\mu}\delta^0{}_{\nu)} + g^{00}l^{0i}g_{\mu\nu}\big)\partial_0 M_{0i0} \\
&+ 2\big(l^{0j}\delta^0{}_{(\mu}\delta^i{}_{\nu)} - g^{00}l^j{}_{(\mu}\delta^i{}_{\nu)} - l^{ij}\delta^0{}_{(\mu}\delta^0{}_{\nu)} + g^{00}l^{ij}g_{\mu\nu}\big)\partial_0 M_{0ij}
\end{aligned}
$$

---

[6] Quasilinear means the coefficient of the highest order derivatives only depend on derivatives of *strictly* lower order.

$$+ \left[2N_{\alpha\beta\gamma} \left(-l^{0\alpha} l^{0\gamma} \delta^{\beta}{}_{(\mu} \delta^{0}_{\nu)} + g^{00} l^{0\alpha} \delta^{\beta}{}_{(\mu} l^{\gamma}{}_{\nu)} + g^{0\alpha} l^{0\beta} l^{\gamma}{}_{(\mu} \delta^{0}{}_{\nu)} \right. \right.$$

$$- l^{\alpha\gamma} l^{0\beta} \delta^{0}{}_{\mu} \delta^{0}_{\nu} - g^{0\alpha} l^{0\beta} l^{0\gamma} g_{\mu\nu} + g^{00} l^{\alpha\gamma} l^{0\beta} g_{\mu\nu} \Big)$$

$$+ l^{0}{}_{b} \left(\partial_{\delta} f^{b}{}_{\lambda} \right) \left(l^{\lambda 0} \delta^{0}{}_{(\mu} \delta^{\delta}{}_{\nu)} - l^{\lambda\delta} \delta^{0}{}_{\mu} \delta^{0}_{\nu} + g^{0\delta} l^{\lambda}{}_{(\mu} \delta^{0}{}_{\nu)} \right.$$

$$\left. \left. - g^{00} l^{\lambda}{}_{(\mu} \delta^{\delta}{}_{\nu)} - g^{0\delta} l^{\lambda 0} g_{\mu\nu} + g^{00} l^{\lambda\delta} g_{\mu\nu} \right) \right] \partial_0 f_{00}$$

$$= F^{i\alpha\beta}_{\mu\nu} \partial_i f_{\alpha\beta} + F^{i\alpha\beta\gamma}_{\mu\nu} \partial_i M_{\alpha\beta\gamma} + F_{\mu\nu}, \tag{29}$$

$$E_{\mu} := 2 \left(l^{0i} \delta^{0}{}_{\mu} - l^{00} \delta^{i}{}_{\mu} \right) \partial_0 M_{0i0} + 2 \left(l^{ij} \delta^{0}{}_{\mu} - l^{0j} \delta^{i}{}_{\mu} \right) \partial_0 M_{0ij}$$

$$+ 2 \left(-l^{0\gamma} l^{0\delta} N_{\mu\gamma\delta} + \delta^{0}{}_{\mu} l^{0}{}_{\lambda} l^{\delta\lambda} e_a{}^{\gamma} \partial_{[\gamma} f^a{}_{\delta]} - l^{0\gamma} e_a{}^{0} l^{0}{}_{\mu} \partial_{[\gamma} f^a{}_{\delta]} + l^{0}{}_{a} l^{\gamma\delta} \delta^{0}{}_{[\mu} \partial_{\gamma]} f^a{}_{\delta} \right) \partial_0 f_{00}$$

$$= G^{i\alpha\beta}_{\mu} \partial_i f_{\alpha\beta} + G^{i\alpha\beta\gamma}_{\mu} \partial_i M_{\alpha\beta\gamma} + G_{\mu}, \tag{30}$$

$$E := 4 \left(g^{i\alpha} l^{0\beta} l^{0\gamma} - g^{0\alpha} l^{0\beta} l^{i\gamma} - g^{0\alpha} g^{i\beta} \bar{g}^{0\gamma} \right) N_{\alpha\beta\gamma} \partial_0 M_{0i0}$$

$$+ 4 \left(g^{i\alpha} l^{j\beta} l^{0\gamma} - g^{0\alpha} l^{j\beta} l^{i\gamma} - g^{0\alpha} g^{i\beta} \bar{g}^{j\gamma} \right) N_{\alpha\beta\gamma} \partial_0 M_{0ij} + \Xi \partial_0 f_{00}$$

$$= H^{i\alpha\beta} \partial_i f_{\alpha\beta} + H^{i\alpha\beta\gamma} \partial_i M_{\alpha\beta\gamma} + H, \tag{31}$$

with

$$\Xi := -3\alpha_1 g^{00} - 2l^{\mu 0} e_a{}^{0} e_b{}^{\nu} R_{\mu\nu}{}^{ab} + 4N_{\mu\nu\lambda} N_{\alpha\beta\gamma} \left(g^{0\mu} l^{0\alpha} l^{\nu\gamma} l^{\lambda\beta} + g^{\mu\alpha} l^{0\nu} l^{\lambda\beta} l^{0\gamma} + g^{\mu\alpha} g^{0\nu} l^{0\beta} \bar{g}^{\gamma\lambda} \right)$$

$$- 2 \left(g^{\mu 0} l^{\beta\lambda} l^{\gamma\nu} + g^{\mu 0} g^{\beta\nu} \bar{g}^{\lambda\gamma} - g^{\mu\beta} l^{0\lambda} l^{\gamma\nu} \right) N_{\mu\nu\lambda} l^{0}{}_{a} \partial_{\beta} f^a{}_{\gamma} - 4l^{0}{}_{a} \partial_{\rho} f^a{}_{\gamma} \left(l^{\gamma\mu} l^{\beta 0} e^{b\rho} \right.$$

$$+ e^{b\gamma} l^{\beta 0} l^{\rho\mu} - l^{\gamma\mu} l^{\beta\rho} e^{b0} \Big) \mathcal{F}_{\mu\beta b} + 2N_{\nu\rho\gamma} \left(2l^{\gamma 0} l^{0\mu} l^{\beta\nu} e^{b\rho} + 2l^{\gamma\mu} l^{\beta 0} l^{0\nu} e^{b\rho} \right.$$

$$+ l^{\gamma\mu} l^{\beta\nu} e^{b0} l^{0\rho} + 2l^{\beta 0} l^{0\nu} l^{\rho\mu} e^{b\gamma} \Big) \mathcal{F}_{\mu\beta b} - \left(2\eta^{ab} l^{\beta 0} l^{\nu 0} l^{\mu\alpha} + 2\eta^{ab} l^{\beta\nu} l^{\mu 0} l^{0\alpha} \right.$$

$$+ \bar{g}^{\nu\alpha} e^{a0} e^{b0} l^{\beta\mu} + \bar{g}^{\nu\alpha} e^{a\lambda} f^b{}_{\lambda} l^{\beta 0} l^{\mu 0} \Big) \mathcal{F}_{\mu\nu a} \mathcal{F}_{\alpha\beta b}, \tag{32}$$

where $F^{i\alpha\beta}_{\mu\nu}$, $F^{i\alpha\beta\gamma}_{\mu\nu}$, $F_{\mu\nu}$, $G^{i\alpha\beta}_{\mu}$, $G^{i\alpha\beta\gamma}_{\mu}$, G_{μ}, $H^{i\alpha\beta}$, $H^{i\alpha\beta\gamma}$ and H are functions of $f_{\alpha\beta}$, $M_{\mu\nu\lambda}$, $f^a{}_{\alpha}$, $\partial_{\mu} f^a{}_{\alpha}$ and $\partial_{\mu} \partial_{\nu} f^a{}_{\alpha}$, and contain no derivative terms of $f_{\alpha\beta}$ and $M_{\mu\nu\lambda}$. Although equations (29)–(31) seem to have fifteen components in total, only ten of the fifteen are actually independent time-evolution equations. This can be seen as follows: for $\mu = 0$ in equation (29), the left-hand side becomes zero, and thus it does not give any time evolution. Moreover, the trace of the left-hand side of equation (29) is proportional to the left-hand side of equation (30) with $\mu = 0$, which means that the trace does not give any independent information for the time evolution. This is consistent with the discussion in the previous section, i.e., five of ten degrees of freedom of equation (10) should become constraints as in equations (11) and (12).

To conclude, we now have ten time-evolution equations. We can finally write down the characteristic equations. They are:

$$\frac{1}{4} \Xi \tilde{f}_{00} + \sum_i \left(g^{i\alpha} l^{0\beta} l^{0\gamma} - g^{0\alpha} l^{0\beta} l^{i\gamma} - g^{0\alpha} g^{i\beta} \bar{g}^{0\gamma} \right) N_{\alpha\beta\gamma} \tilde{M}_{0i0}$$

$$+ \sum_i \left(g^{i\alpha} l^{i\beta} l^{0\gamma} - g^{0\alpha} l^{i\beta} l^{i\gamma} - g^{0\alpha} g^{i\beta} \bar{g}^{i\gamma} \right) N_{\alpha\beta\gamma} \tilde{M}_{0ii}$$

$$+ \sum_{i>j} 2 \left(g^{i\alpha} l^{j\beta} l^{0\gamma} - g^{0\alpha} l^{j\beta} l^{i\gamma} - g^{0\alpha} g^{i\beta} \bar{g}^{j\gamma} \right) N_{\alpha\beta\gamma} \tilde{M}_{0(ij)} = 0, \tag{33}$$

$$l^{00} \tilde{M}_{0i0} + l^{0i} \tilde{M}_{0ii} + \sum_{i \neq j} l^{0j} \tilde{M}_{0(ij)} + \left(l^{0\alpha} l^{0\beta} N_{i\alpha\beta} - l^{0\gamma} e_a{}^{0} l^{\delta}{}_{i} \partial_{[\gamma} f^a{}_{\delta]} + \frac{1}{2} l^{0}{}_{a} l^{0\delta} \partial_i f^a{}_{\delta} \right) \tilde{f}_{00} = 0, \tag{34}$$

$$\sum_k \left(l^0{}_{(i}\delta^k{}_{j)} - l^{0k}g_{ij} \right)\tilde{M}_{0k0} + \sum_k \left(l^k{}_{(i}\delta^k{}_{j)} - l^{kk}g_{ij} \right)\tilde{M}_{0kk} + 2\sum_{k>l} \left(l^k{}_{(i}\delta^l{}_{j)} - l^{lk}g_{ij} \right)\tilde{M}_{0(kl)}$$

$$+ \left[N_{\mu\nu\lambda}\left(-l^{0\mu}\delta^\nu{}_{(i}l^\lambda{}_{j)} - g^{0\mu}l^{0\nu}l^{0\lambda}g_{ij} - l^{\mu\lambda}l^{0\nu}g_{ij} \right) \right.$$

$$\left. + \frac{1}{2}l^0{}_b\left(\partial_\delta f^b{}_\lambda \right)\left(l^\lambda{}_{(i}\delta^\delta{}_{j)} - g^{0\delta}l^{\lambda 0}g_{ij} - l^{\lambda\delta}g_{ij} \right) \right]\tilde{f}_{00} = 0, \tag{35}$$

where we explicitly show the summation with respect to spatial coordinates because some of the indices do not obey Einstein's summation rule. Here, we decompose M_{0ij} into $M_{0(ij)}$ and $M_{0[ij]}$, but $M_{0[ij]}$ does not appear in the characteristic equations because from equation (16) we see that $M_{0[ij]}$ can be expressed in terms of M_{ij0}. Here, the notation $\tilde{M}_{\mu\nu\lambda}$ and $\tilde{f}_{00}$ means they are not the values of $M_{\mu\nu\lambda}$ and f_{00}, but represent the '(finite) difference' of the fields in the x^0-direction. The term 'difference' is related to the x_0-derivative in the case when the derivative is well defined. In the context of calculus of single variable $y = y(x)$, the 'difference' is the quantity $\Delta y/\Delta x$, related to the derivative dy/dx by the well-known definition $dy/dx = \lim_{\Delta x \to 0} \Delta y/\Delta x$. Here we do not use the x_0-derivative but 'difference' because, in characteristic analysis, sometimes we consider shock waves on which derivatives are not well defined (see the discussion in appendix).

4. Characteristic equations with non-vanishing contortion

In [18], it was argued that, if the contortion does not vanish, an arbitrary hypersurface can be a characteristic hypersurface. If so, a generic solution always has an acausal structure, and thus the theory is very problematic. With our characteristic equations, we revisit this problem.

We first consider the simplest example with the Minkowski background tetrad. Suppose the dynamical tetrad completely coincides with the background tetrad, i.e., $f^a{}_\mu = e^a{}_\mu$, and let us examine the characteristic determinant corresponding to a flat spatial hypersurface. Under this situation, the expressions of $g_{\mu\nu}$, $\bar{g}_{\mu\nu}$, $f_{\mu\nu}$, $l_{\mu\nu}$, $M_{\mu\nu\lambda}$ and $N_{\mu\nu\lambda}$ become

$$g_{\mu\nu} = \bar{g}_{\mu\nu} = f_{\mu\nu} = l_{\mu\nu} = \eta_{\mu\nu}, \quad \text{and} \quad M_{\mu\nu\lambda} = N_{\mu\nu\lambda} = 0. \tag{36}$$

Then the characteristic equations simplify significantly and reduce to

$$-\tilde{M}_{011} - \tilde{M}_{022} = 0, \qquad 2\tilde{M}_{0(12)} = 0, \qquad -\tilde{M}_{010} = 0,$$

$$-\tilde{M}_{022} - \tilde{M}_{033} = 0, \qquad 2\tilde{M}_{0(23)} = 0, \qquad -\tilde{M}_{020} = 0,$$

$$-\tilde{M}_{033} - \tilde{M}_{011} = 0, \qquad 2\tilde{M}_{0(31)} = 0, \qquad -\tilde{M}_{030} = 0, \qquad -\tfrac{3}{4}\alpha_1\tilde{f}_{00} = 0. \tag{37}$$

The characteristic determinant of this system is obviously generically nonzero and its eigenvalues are of order $O(1)$ and $O(\alpha_1)$, which are finite constants.

Now we consider a tiny perturbation on the Minkowski space and consider the changes in the characteristic determinant. Since the characteristic determinant on the Minkowski space is finite and nonzero, by continuity, with a sufficiently small perturbation, the characteristic determinant must still be nonzero *even if the contortion does not vanish*. This means that the hypersurface is *not* a characteristic hypersurface (although for some values of the fields, the characteristic matrix could become zero and thus a superluminal mode could potentially arise. We comment on this further in section 5). This is in conflict with the conclusion of [18]. There is thus a need to point out why our analysis gives a result that is different from that of [18].

In our analysis, the first-order derivative of f_{00} joins in the characteristic equations, while in the analysis of [18] it does not. The reason why it does not appear in their analysis is that the authors assume all discontinuities across a shock surface only show up in the second-order derivative of the dynamical metric. However, as we will show in appendix, in the discussion

of characteristic we should define the characteristic matrix such that every dynamical field appears in the general solutions. Indeed, except for the contribution from $\partial_0 f_{00}$, our result is otherwise consistent with that of [18]. However, since their characteristic matrix does not have the contribution from $\partial_0 f_{00}$, its determinant always becomes zero. If we go back to the original viewpoint of the Cauchy–Kovalevskaya theorem and the discussion about shock waves, it is reasonable for $\partial_0 f_{00}$ to appear in the characteristic matrix. This is further discussed in appendix.

5. Discussion: Does nonlinear massive gravity admit superluminal propagation?

We have carefully examined the characteristic equations of a special case of nonlinear massive gravity and showed that Cauchy problem can be well posed, in the sense that not every hypersurface can be a characteristic hypersurface given a physical metric. In the viewpoint on the fixed fiducial metric, of course, on an arbitrary hypersurface there probably exists specific field values, i.e., the specific physical metrics, which give vanishing characteristic determinant. In this sense, any hypersurface potentially *can* be characteristic by tuning values of a physical metric. However, the situation here is a bit more subtle than the result in [18], in which the authors argued that, any hypersurface can be characteristic on a given physical metric. The difference is due to the contribution of the first derivative term of f_{00}, which should appear in the characteristic equation in view of the discussion in appendix.

We emphasize again that our result does *not* guarantee that the theory is free of problems. Indeed, we have merely shown that there exist good non-characteristic hypersurfaces, *not* the *absence* of superluminal characteristics. One should next examine the characteristic matrix carefully to determine if its determinant admits any real root, the presence of which would imply the existence of superluminal propagations. The existence of such real root will directly affect the discussion of causality. It reduces the Cauchy region and is sometimes responsible for giving rise to acausality. In order to have a deeper understanding of the causal structure of the theory, we will need an elaborate inspection of the characteristics. We shall leave this issue for future work. Nevertheless, we would like to emphasize that superluminality does *not* always imply acausality [34–37]. In addition, one should check if occurrence of superluminal modes simply signal that the theory as an *effective* theory is breaking down, *à la* [38].

It is probably a good guess that nonlinear massive gravity does suffer from superluminal propagation in general and that this might result in acausality. Indeed, we have seen tantalizing hints of superluminal propagation on the self-accelerating solution [33], and in fact, *instantaneous* propagation may arise in the full theory with all α_n being nonzero [39]. In particular, energy can probably be emitted with infinite speed on the self-accelerating background by the helicity-0 mode of the massive graviton. In addition to superluminal mode, nonlinear massive gravity on the self-accelerating solution has the property that although it generically has five degrees of freedom [14], in the second-order action on an open FLRW background, perturbative analysis only reveals two tensors degrees of freedom [22]. That is, the full theory has different number of degrees of freedom compared to some of the low-order limits. This could be dangerous as the excitations of the extra degrees of freedom may be accompanied by an anomalous characteristic, as discussed in the context of $f(T)$ gravity [40] and Poincaré gauge theory [41]. Solutions with such anomalous characteristic are inappropriate to describe our Universe. Furthermore, nonlinear instability has also been pointed out in the self-accelerating solution [42].

There remain a few puzzles to be resolved. Firstly, a further careful analysis is required to check if nonlinear massive gravity with $\{\alpha_n\}_{n=2,3,4} = 0$ does admit superluminal propagation as argued in [18], which might well be the case. Secondly, it would be interesting to

investigate how the superluminal helicity-0 mode arises in the full theory and its corresponding characteristic analysis. Solving them, we can assess the appropriateness of each solution, and it might give us further constraints on the theory.

Acknowledgments

Appendix. Review of characteristic analysis *à la* Cauchy and Kovalevskaya

The analysis of the characteristic is related to the Cauchy–Kovalevskaya theorem, from which the propagation of shock wave front can be obtained. Our discussion here is based on the equations that appear in the usual proof of the Cauchy–Kovalevskaya theorem (see, e.g., the seminal work of Courant and Hilbert [43]). Since the naive extension of the result in the case involving only a single field to that of a multi-field case is invalid, we revisit the original arguments used in the proof of the Cauchy–Kovalevskaya theorem and derive the exact characteristic equations in the multi-field case.

A.1. Single field

Suppose we have an nth order differential equation which is quasilinear for the nth order derivative terms, i.e.

$$A^{\alpha_1\cdots\alpha_n}\partial_{\alpha_1\cdots\alpha_n}\phi + A = 0, \tag{A.1}$$

where ϕ is a field and $A^{\alpha_1\cdots\alpha_n}$ and A are functions of x^μ and derivatives of ϕ up to the $(n-1)$th order. We now introduce a hypersurface Σ which is orthogonal to ξ^μ. If

$$A^{\alpha_1\cdots\alpha_n}\xi_{\alpha_1}\cdots\xi_{\alpha_n} = 0, \tag{A.2}$$

then the hypersurface Σ is called a *characteristic hypersurface*. On the other hand, if

$$A^{\alpha_1\cdots\alpha_n}\xi_{\alpha_1}\cdots\xi_{\alpha_n} \neq 0, \tag{A.3}$$

the hypersurface Σ is called a *non-characteristic hypersurface*. We have the following important result [31, 32].

Cauchy–Kovalevskaya theorem. *If all the coefficients $A^{\alpha_1\cdots\alpha_n}$ and A are analytic and if the hypersurface Σ is a non-characteristic hypersurface, then there exists a unique local analytic solution ϕ in the neighbourhood of Σ.*

This theorem implies that the uniqueness of time evolution from a characteristic hypersurface is not guaranteed. Moreover, there is an important fact about characteristics, namely:

Remark. A characteristic hypersurface can be a shock wave front.

This remark means that there could be nonlinear propagation on the characteristic hypersurface. It can be intuitively understood as follows: since we have the nth order differential equation of motion for the single field ϕ, we can introduce, on the initial hypersurface Σ_0, n

numbers of initial conditions for $(\partial_t)^{n-1}\phi, \ldots, \partial_t\phi$ and ϕ, where $\xi^\mu := (\partial/\partial t)^\mu$. If the initial hypersurface Σ_0 is not characteristic, we can derive $(\partial_t)^n\phi$ from the equation of motion, and then the values $(\partial_t)^{n-1}\phi, \ldots, \partial_t\phi$ and ϕ after a small time evolution from the initial hypersurface Σ_0 can be uniquely fixed as

$$(\partial_t)^{k-1}\phi(t_0 + \Delta t) = (\partial_t)^{k-1}\phi(t_0) + (\partial_t)^k\phi(t_0)\Delta t, \qquad (1 \leqslant k \leqslant n), \quad \text{(A.4)}$$

where $t = t_0$ on the initial hypersurface Σ_0. However, if the hypersurface Σ_0 is characteristic, $(\partial_t)^n\phi$ cannot be fixed uniquely, which allows a discontinuity in $(\partial_t)^n\phi$ and thus non-unique time evolution of $(\partial_t)^{n-1}\phi$.

In order to make the discussion in the next subsection clearer, we comment here that, in the usual proof of Cauchy–Kovalevskaya theorem [43], a quasilinear nth order differential equation for the single field ϕ is reduced to a system of quasilinear first-order differential equations for the multi-field $\mathbf{u}$, where

$$\mathbf{u} := \left(\phi, \frac{\partial\phi}{\partial x^0}, \ldots, \frac{\partial\phi}{\partial x^d}, \frac{\partial^2\phi}{\partial x^0 \partial x^0}, \ldots \right), \tag{A.5}$$

and $\mathbf{u}$ includes all partial derivatives of ϕ up to the $(n-1)$th order. Let m denote the number of components of $\mathbf{u}$. Then, equation (A.2) can be exactly represented by

$$B^t [\mathbf{u}, x] \, \partial_t\mathbf{u} = B^i [\mathbf{u}, x] \, \partial_i\mathbf{u} + B [\mathbf{u}, x], \tag{A.6}$$

where $B^t [\mathbf{u}, x]$, $B^i [\mathbf{u}, x]$ and $B [\mathbf{u}, x]$ depend on $\mathbf{u}$ and x. The first two are $m \times m$ matrices and the last one is a m-vector. Time evolution of every component in $\mathbf{u}$ can be obtained if $\det B^t \neq 0$. While most of the eigenvalues of $B^t [\mathbf{u}, x]$ are always nonzero, the only nontrivial eigenvalue is the same as $A^{t\cdots t}$ in equation (A.2). In the proof of the Cauchy–Kovalevskaya theorem, after transforming into equation (A.5), the uniqueness is proved under the condition $\det B^t \neq 0$ (of course we also need the analyticity condition).

To conclude, in the Cauchy–Kovalevskaya theorem and in the discussion of shock waves, the only important term is $A^{t\cdots t}$. If $A^{t\cdots t} = 0$, it is outside the range of applicability of the Cauchy–Kovalevskaya theorem to ensure a well-posed Cauchy problem, and a shock wave can indeed propagate on the characteristic hypersurface.

A.2. Multi-field

Here, we extend the discussion in the previous subsection to a multi-field case. While this can sometimes be extended straightforwardly, it is in general, quite nontrivial. We will give examples of both cases in this subsection.

A.2.1. The simple case. We consider equations for two fields ϕ and ψ,

$$S^{\mu\nu}\partial_\mu\partial_\nu\phi + T^{\mu\nu}\partial_\mu\partial_\nu\psi + S = 0, \tag{A.7}$$

$$U^{\mu\nu}\partial_\mu\partial_\nu\phi + V^{\mu\nu}\partial_\mu\partial_\nu\psi + U = 0, \tag{A.8}$$

where $S^{\mu\nu}$, $T^{\mu\nu}$, S, $U^{\mu\nu}$, $V^{\mu\nu}$ and U are all functions of $\partial_\alpha\phi$, ϕ, $\partial_\beta\psi$, ψ and x. Picking up the higher order derivative terms with respect to t, we have

$$\mathbf{M} \begin{pmatrix} \tilde\phi \\ \tilde\psi \end{pmatrix} = 0, \qquad \text{with} \qquad \mathbf{M} := \begin{pmatrix} S^{tt} & T^{tt} \\ U^{tt} & V^{tt} \end{pmatrix}, \tag{A.9}$$

where $\tilde\phi$ and $\tilde\psi$ mean they are not the values of ϕ and ψ, but represent the 'difference' of the fields in the t-direction. If $\det\mathbf{M} = 0$, we cannot derive one of the linear combinations of $\tilde\phi$ and $\tilde\psi$, and the hypersurface is a characteristic hypersurface, according to the discussion in the previous subsection.

A.2.2. The nontrivial case.　　Here, we consider the equations for two fields ζ and σ,

$$I^{\mu\nu}\partial_\mu\partial_\nu\zeta + J^\mu\partial_\mu\sigma + I = 0, \tag{A.10}$$

$$L = 0, \tag{A.11}$$

where $I^{\mu\nu}$, $J^{\mu\nu}$, I and L are all functions of $\partial_\alpha\zeta$, ζ, σ and x. Actually, the equations in the nonlinear massive gravity theory are very similar to these equations. To obtain the second-order differential equation, we operate with ∂_μ in equation (A.11) and obtain

$$\frac{\partial L}{\partial(\partial_\alpha\zeta)}\partial_\mu\partial_\alpha\zeta + \frac{\partial L}{\partial\zeta}\partial_\mu\zeta + \frac{\partial L}{\partial\sigma}\partial_\mu\sigma + \frac{\partial L}{\partial x^\mu} = 0. \tag{A.12}$$

If we naively pick up the higher order derivative terms, the characteristic equations are

$$\bar{\mathbf{N}}\begin{pmatrix}\tilde{\zeta}\\\tilde{\sigma}\end{pmatrix} = 0, \qquad \text{with} \qquad \bar{\mathbf{N}} := \begin{pmatrix} I^{tt} & 0 \\ \dfrac{\partial L}{\partial(\partial_t\zeta)} & 0 \end{pmatrix}, \tag{A.13}$$

where $\tilde{\zeta}$ and $\tilde{\sigma}$ mean they are not the values of ζ and σ, but represent the 'difference' of the fields in the t-direction. Then, det $\bar{\mathbf{N}}$ is trivially zero. Does it mean that every hypersurface is a characteristic hypersurface?

In order to investigate this question, first we consider a simple example

$$L = P + Q\sigma, \tag{A.14}$$

where P is a function of $\partial_\alpha\zeta$ and ζ, while Q is a nonzero constant. Then, equation (A.11) can be regarded as a constraint equation for fixing the auxiliary field σ and we can solve it for σ. Substituting the solution into equation (A.10), we can obtain an equation for ζ:

$$\left(QI^{\mu\nu} + J^\mu\frac{\partial P}{\partial(\partial_\nu\zeta)}\right)\partial_\mu\partial_\nu\zeta + \bar{I} = 0, \tag{A.15}$$

where $\bar{I}$ is a function of $\partial_\alpha\zeta$, ζ and x. Now, equation (A.15) is the equation for the single field ζ and the characteristic equation for ζ is obviously

$$\left(QI^{tt} + J^t\frac{\partial P}{\partial(\partial_t\zeta)}\right)\tilde{\zeta} = 0, \tag{A.16}$$

which is different from equation (A.13). From the discussion in the previous subsection, ζ is uniquely fixed in the neighborhood if

$$QI^{tt} + J^t\frac{\partial P}{\partial(\partial_t\zeta)} \neq 0, \tag{A.17}$$

and then, σ can be also determined uniquely from equations (A.11) and (A.14). Therefore, we find that equation (A.16) is the exact characteristic equation instead of equation (A.13).

How do we derive the characteristic equation in a general multi-field case? Going back to equations (A.10) and (A.11), it can be seen that in both equations, the order of derivative operators of σ are one less than those of ζ. This means that discontinuity of $(\partial_t)^k\zeta$ can be balanced with that of $(\partial_t)^{k-1}\sigma$. Therefore, not only the second-order derivative of ζ but also the *first*-order derivative of σ should appear in the characteristic equation. Then, the characteristic equation can be written as

$$\mathbf{N}\begin{pmatrix}\tilde{\zeta}\\\tilde{\sigma}\end{pmatrix} = 0, \qquad \text{with} \qquad \mathbf{N} := \begin{pmatrix} I^{tt} & J^t \\ \dfrac{\partial L}{\partial(\partial_t\zeta)} & \dfrac{\partial L}{\partial\sigma} \end{pmatrix}, \tag{A.18}$$

and the condition of the characteristic is consistent with equation (A.16).

This statement can be justified further by writing the equations in the form of equation (A.6). Let us introduce new variables

$$v_i := \partial_i \zeta,$$ (A.19)

$$u := \partial_t \zeta,$$ (A.20)

which are related by the commutativity of partial derivatives:

$$\partial_t v_i = \partial_i u.$$ (A.21)

Equations (A.10) and (A.12) can be represented by

$$I^{tt} \partial_t u + J^t \partial_t \sigma = -(I^{ti} + I^{it}) \partial_i u - I^{ij} \partial_i v_j - J^i \partial_i \sigma - I,$$ (A.22)

$$\frac{\partial L}{\partial (\partial_t \zeta)} \partial_t u + \frac{\partial L}{\partial \sigma} \partial_t \sigma = -\frac{\partial L}{\partial (\partial_i \zeta)} \partial_i u - \frac{\partial L}{\partial \zeta} \partial_i \zeta - \frac{\partial L}{\partial x^t},$$ (A.23)

where all coefficients $I^{\mu\nu}$, J^μ, I, $\partial L/\partial (\partial_\mu \zeta)$, $\partial L/\partial \zeta$, $\partial L/\partial \sigma$ and $\partial L/\partial x^t$ are written in non-derivative variables ζ, σ, u and v_i. The combination of equations (A.20)–(A.23) is exactly of the form of equation (A.6). We can then apply the Cauchy–Kovalevskaya theorem and also have the same discussion about shock waves. Again, we can deduce that $\det \mathbf{N} = 0$ is the condition for the hypersurface to be the characteristic hypersurface.

Finally, we discuss in brief the method for deriving characteristic equations with general multi-fields. We begin by choosing one of the fields and define the dimension of it as zero. The dimension, denoted by [·], of the differential operator is 1, i.e. $[\partial] = 1$. We determine the dimensions of other fields such that each field necessarily appears in one of the characteristic equations. In the case of equations (A.10) and (A.11), for instance, we define the dimension of ζ as 0. If the dimension of σ is larger than one, only $\partial_\mu \sigma$ can appear in the characteristic equations. On the other hand, if the dimension of σ is smaller than one, only $\partial_\mu \partial_\nu \zeta$ can appear in the characteristic equations. Therefore, only in the case where the dimension of σ is equal to one, both can appear in the characteristic equations, in which case we have the situation discussed above.

References

[1] Hinterbichler K 2012 Theoretical aspects of massive gravity *Rev. Mod. Phys.* **84** 671 (arXiv:1105.3735v2 [hep-th])
[2] Fierz M and Pauli W 1939 On relativistic wave equations for particles of arbitrary spin in an electromagnetic field *Proc. R. Soc. Lond.* A **173** 211
[3] Van Nieuwenhuizen P 1973 On ghost-free tensor Lagrangians and linearized gravitation *Nucl. Phys.* B **60** 478
[4] van Dam H and Veltman M J G 1970 Massive and massless Yang–Mills and gravitational field *Nucl. Phys.* B **22** 397
[5] Zakharov V I 1970 Linearized gravitation theory and the graviton mass *JETP Lett.* **12** 312
[6] Vainshtein A I 1972 To the problem of nonvanishing graviton mass *Phys. Lett.* B **39** 393
[7] Babichev E, Deffayet C and Ziour R 2010 The recovery of general relativity in massive gravity via the Vainshtein mechanism *Phys. Rev.* D **82** 104008 (arXiv:1007.4506v1[gr-qc])
[8] Boulware D and Deser S 1972 Can gravitation have a finite range? *Phys. Rev.* D **6** 3368
[9] Chen P 2010 Gauge theory of gravity with de Sitter symmetry as a solution to the cosmological constant problem and the dark enegry puzzle *Mod. Phys. Lett.* A **25** 2795 (arXiv:1002.4275v2 [gr-qc])
[10] de Rham C and Gabadadze G 2010 Generalization of the Fierz–Pauli action *Phys. Rev.* D **82** 044020 (arXiv:1007.0443v2 [hep-th])
[11] de Rham C, Gabadadze G and Tolley A J 2011 Resummation of massive gravity *Phys. Rev. Lett.* **106** 231101 (arXiv:1011.1232v2 [hep-th])
[12] de Rham C, Gabadadze G and Tolley A J 2012 Ghost free massive gravity in the Stückelberg language *Phys. Lett.* B **711** 190 (arXiv:1107.3820v1 [hep-th])
[13] de Rham C, Gabadadze G and Tolley A J 2011 Helicity decomposition of Ghost-free massive gravity *J. High Energy Phys.* JHEP11(2011)093 (arXiv:1108.4521v2 [hep-th])

[14] Hassan S F and Rosen R A 2012 Resolving the ghost problem in non-linear massive gravity *Phys. Rev. Lett.* **108** 041101 (arXiv:1106.3344v3 [hep-th])

[15] Hassan S F, Rosen R A and Schmidt-May A 2012 Ghost-free massive gravity with a general reference metric *J. High Energy Phys.* JHEP02(2012)026 (arXiv:1109.3230v2 [hep-th])

[16] Hassan S F and Rosen R A 2012 Confirmation of the secondary constraint and absence of ghost in massive gravity and bimetric gravity *J. High Energy Phys.* JHEP04(2012)123 (arXiv:1111.2070v1 [hep-th])

[17] Hassan S, Schmidt-May A and von Strauss M 2012 Proof of consistency of nonlinear massive gravity in the Stúckelberg formulation *Phys. Lett.* B **715** 335 (arXiv:1203.5283v2 [hep-th])

[18] Deser S and Waldron A 2013 Acausality of massive gravity *Phys. Rev. Lett.* **110** 111101 (arXiv:1212.5835v3 [hep-th])

[19] Zumino B 1970 *Effective Lagrangians and Broken Symmetries (Brandeis University Lectures on Elementary Particles and Quantum Field Theory* vol 2*)* ed S Deser, M Grisaru and H Pendleton (Cambridge, MA: MIT Press) p 437

[20] Deser S, Sandora M and Waldron A 2013 Nonlinear partially massless from massive gravity? *Phys. Rev.* D **87** 101501 (arXiv:1301.5621 [hep-th])

[21] D'Amico G, de Rham C, Dubovsky S, Gabadadze G, Pirtskhalava D and Tolley A J 2011 Massive cosmologies *Phys. Rev.* D **84** 124046 (arXiv:1108.5231v1 [hep-th])

[22] Gumrukcuoglu A E, Lin C and Mukohyama S 2012 Cosmological perturbations of self-accelerating universe in nonlinear massive gravity *J. Cosmol. Astropart. Phys.* JCAP03(2012)006 (arXiv:1111.4107v2 [hep-th])

[23] Chamseddine A H and Mukhanov V 2011 Massive gravity simplified: a quadratic action *J. High Energy Phys.* JHEP08(2011)091 (arXiv:1106.5868 [hep-th])

[24] Chamseddine A H and Volkov M S 2011 Cosmological solutions with massive gravitons *Phys. Lett.* B **704** 652 (arXiv:1107.5504 [hep-th])

[25] Hinterbichler K and Rosen R A 2012 Interacting spin-2 fields *J. High Energy Phys.* JHEP07(2012)047 (arXiv:1203.5783 [hep-th])

[26] Hassan S F, Schmidt-May A and von Strauss M 2012 Metric formulation of ghost-free multivielbein theory arXiv:1204.5202 [hep-th]

[27] Nomura K and Soda J 2012 When is multimetric gravity ghost-free? *Phys. Rev.* D **86** 084052 (arXiv:1207.3637 [hep-th])

[28] Hehl F W, von der Heyde P, Kerlick G D and Nester J M 1976 General relativity with spin and torsion: foundations and prospects *Rev. Mod. Phys.* **48** 393

[29] Hassan S F and Rosen R A 2011 On non-linear actions for massive gravity *J. High Energy Phys.* JHEP07(2011)009 (arXiv:1103.6055v3 [hep-th])

[30] Deffayet C, Mourad J and Zahariade G 2013 Covariant constraints in ghost free massive gravity *J. Cosmol. Astropart. Phys.* JCAP01(2013)031 (arXiv:1207.6338 [hep-th])

[31] Cauchy A 1842 Compt. Rend. 15 (reprinted in Oeuvres Completes, serie 1, VII, pp 17–58)

[32] Kowalevskaya S 1875 Zur Theorie der partiellen Differentialgleichung *J. Reine Angew. Math.* **80** 1

[33] Gumrukcuoglu A E, Lin C S and Mukohyama S 2012 Anisotropic Friedmann–Robertson–Walker universe from nonlinear massive gravity *Phys. Lett.* B **717** 295 (arXiv:1206.2723v2 [hep-th])

[34] Bruneton J-P 2007 On causality and superluminal behavior in classical field theories: applications to k-essence theories and MOND-like theories of gravity *Phys. Rev.* D **75** 085013 (arXiv:gr-qc/0607055v2)

[35] Afshordi N, Chung D J H and Geshnizjani G 2007 Cuscuton: a causal field theory with an infinite speed of sound *Phys. Rev.* D **75** 083513 (arXiv:hep-th/0609150v2)

[36] Geroch R 2010 Faster than light? arXiv:1005.1614v1 [gr-qc]

[37] Berezhiani L, Chkareuli G and Gabadadze G 2013 Restricted Galileons arXiv:1302.0549v2 [hep-th]

[38] Burrage C, de Rham C, Heisenberg L and Tolley A J 2012 Chronology protection in Galileon models and massive gravity *J. Cosmol. Astropart. Phys.* JCAP07(2012)004 (arXiv:1111.5549v1 [hep-th])

[39] Chiang C-I, Izumi K and Chen P 2012 Spherically symmetric analysis on open FLRW solution in nonlinear massive gravity *J. Cosmol. Astropart. Phys.* JCAP12(2012)025 (arXiv:1208.1222v2 [hep-th])

[40] Ong Y C, Izumi K, Nester J M and Chen P 2013 Problems with propagation and time evolution in f(T) gravity arXiv:1303.0993v1 [gr-qc]

[41] Chen H, Nester J M and Yo H-J 1998 Acausal PGT modes and the nonlinear constraint effect *Acta. Phys. Pol.* B **29** 961

[42] De Felice A, Gumrukcuoglu A E and Mukohyama S 2012 Massive gravity: nonlinear instability of the homogeneous and isotropic universe *Phys. Rev. Lett.* **109** 171101 (arXiv:1206.2080v2 [hep-th])

[43] Courant R and Hilbert D 1962 *Methods of Mathematical Physics* vol 2 (New York: Wiley-Interscience) p 590
Courant R and Hilbert D 1962 *Methods of Mathematical Physics* vol 2 (New York: Wiley-Interscience) p 596