

## Dynamics of the generalized unimodular gravity theory

A. O. Barvinsky,<sup>1,\*</sup> N. Kolganov,<sup>2,3,†</sup> A. Kurov,<sup>1,‡</sup> and D. Nesterov<sup>1,§</sup>

<sup>1</sup>Theory Department, Lebedev Physics Institute, Leninsky Prospekt 53, Moscow 119991, Russia

<sup>2</sup>Moscow Institute of Physics and Technology, Institutskii per. 9, Dolgoprudny 141700, Russia

<sup>3</sup>Bogoliubov Laboratory of Theoretical Physics, JINR, Joliot-Curie 6, Dubna 141980, Russia



(Received 9 April 2019; published 29 July 2019)

The Hamiltonian formalism of the generalized unimodular gravity theory, which was recently suggested as a model of dark energy, is shown to be a complicated example of constrained dynamical system. The set of its canonical constraints has a bifurcation—splitting of the theory into two branches differing by the number and type of these constraints, one of the branches effectively describing a gravitating perfect fluid with the time-dependent equation of state, which can potentially play the role of dark energy in cosmology. The first class constraints in this branch generate local gauge symmetries of the Lagrangian action—two spatial diffeomorphisms—and rule out the temporal diffeomorphism which does not have a realization in the form of the canonical transformation on phase space of the theory and turns out to be either nonlocal in time or violating boundary conditions at spatial infinity. As a consequence, the Hamiltonian reduction of the model enlarges its physical sector from two general relativistic modes to three degrees of freedom including the scalar graviton. This scalar mode is free from ghost and gradient instabilities on the Friedmann background in a wide class of models subject to a certain restriction on time-dependent parameter  $w$  of the dark fluid equation of state,  $p = w\varepsilon$ . For a special family of models this scalar mode can be ruled out even below the phantom divide line  $w = -1$ , but this line cannot be crossed in the course of the cosmological expansion. This is likely to disable the generalized unimodular gravity as a model of the phenomenologically consistent dark energy scenario, but opens the prospects in inflation theory with a scalar graviton playing the role of inflaton.

DOI: [10.1103/PhysRevD.100.023542](https://doi.org/10.1103/PhysRevD.100.023542)

### I. INTRODUCTION

Recently suggested theory of the generalized unimodular gravity (GUMG) [1] was motivated by the necessity to build a model of dark energy with a variable in time equation of state that could fit cosmological acceleration data. As a candidate for dark energy, this model oversteps the limitations of the simplest models—Einstein general relativity (GR) with a fundamental cosmological constant and the unimodular gravity (UMG) theory [2]. In GR the global degree of freedom (d.o.f.) responsible for cosmological acceleration exists in the form of the fundamental cosmological constant  $\Lambda$ , while the UMG theory is more flexible because  $\Lambda$  arises as an integration constant of equations of motion and should be fixed by the choice of

initial conditions. This flexibility is not sufficient, however, to unleash dark energy evolution,  $\dot{\Lambda} \neq 0$ . In particular, it does not allow the system to cross the phantom divide line  $w = -1$  of the equation of state parameter [3,4], which is likely to be indicated by observations [5,6].

Quite interestingly, this dilemma of obtaining nonclustering (that is global or homogeneous in space) but evolving in time physical mode can be solved by generalizing the unimodular gravity theory—replacing the UMG condition of a unit determinant of the spacetime metric  $g_{\mu\nu}$  by the following condition

$$(-g^{00})^{-1/2} = N(\gamma), \quad \gamma = \det \gamma_{ij}, \quad (1.1)$$

with some rather generic function  $N(\gamma)$  of the determinant of the 3-dimensional metric  $\gamma_{ij}$ . Using this kinematical restriction on metric coefficients in the Einstein-Hilbert action  $S_{\text{EH}}[g_{\mu\nu}]$  as the definition of the GUMG action,

$$S_{\text{GUMG}}[g_{ij}, g_{0i}] = S_{\text{EH}}[g_{\mu\nu}]|_{(-g^{00})^{-1/2}=N(\gamma)}, \quad (1.2)$$

one obtains the theory whose equations of motion effectively coincide with Einstein equations in the presence of a perfect fluid which has a barotropic equation of state

\*barvin@td.lpi.ru

†nikita.kolganov@phystech.edu

‡kurov.aleksandr@physics.msu.ru

§nesterov@td.lpi.ru

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.

$p = w\varepsilon$ , its parameter  $w = w(\gamma)$  being determined by the function  $N(\gamma)$  and therefore evolving in time [1]. An important feature of the model is that this dark fluid is built of purely gravitational degrees of freedom—dark energy constituent of the theory is not achieved by adding extra fields, like for example quintessence [7], Chaplygin gas [8], dilaton [9], or khronon [10], but composed of the original spacetime metric variables. As a modification of Einstein gravity, it also does not fall into the category of higher-derivative [11] or nonlocal gravity [12] models, because it is based on an ultralocal kinematical restriction on the lapse function of the theory.

An obvious advantage of this model is that it should not actually be required to pass typical tests on correspondence principle with GR phenomenology, because every solution of Einstein theory solves GUMG equations of motion—as one can show, nine GUMG equations of motion are just the projections of ten Einstein equations onto the subspace of metric coefficients satisfying the condition (1.1). The reversed statement is not correct—there are nontrivial solutions in GUMG theory, which are absent in general relativity. Equivalently, GUMG theory generically has additional physical degrees of freedom missing in GR. Their dynamics might lead to classical and quantum instabilities including as a particular case clustering phenomenon which is not acceptable for dark energy constituents. The attempt of addressing this problem was recently undertaken in [13], but it is hard to agree with various points and conclusions of this work. Therefore, the main goal of our paper is to make a systematic analysis of the physical degrees of freedom in GUMG theory and find out how robust is this theory against potential ghost and gradient instabilities.

The dynamical content of the physical sector strongly depends on local gauge symmetries and their realization as canonical transformations in the constrained Hamiltonian formalism of the theory. Therefore, we develop here the canonical formalism of GUMG model and immediately find out that it is very interesting from the viewpoint of the theory of constrained dynamics [14]. It turns out to be essentially more complicated than that of the Einstein GR or UMG. It has several generations of constraints which in terminology of [14] bifurcate, that is split the theory into two branches differing both by the number of constraints and their type—of the first and of the second class. Moreover, the first class constraints in the physically interesting branch, associated with the presence of dark fluid, do not separately belong to definite generations (primary, secondary, etc.) but represent nontrivial linear combinations of those. This requires us to modify the known Henneaux-Teitelboim procedure of recovering the Lagrangian gauge symmetries from the canonical transformations generated by first class constraints [14]. We develop such a modification and find that the number of local gauge symmetries of the Lagrangian action is actually

smaller than their number anticipated in [1], so that GUMG theory seems to violate the so-called Dirac conjecture of constrained dynamics—equality of the number of Lagrangian gauge symmetries and the number of primary first class constraints. It turns out, however, that the Dirac conjecture still remains true, because one of the three spacetime diffeomorphisms of GUMG theory considered in [1] is actually nonlocal in time (or violates boundary conditions at spatial infinity) and, therefore, does not participate in the Hamiltonian reduction to the physical sector. Ultimately this leads to the third local degree of freedom—the scalar “graviton” which is absent in GR and UMG theory.

We find the range of perturbative stability of the theory on the cosmological Friedmann background—the class of functions  $w(\gamma)$  for which this scalar graviton is free from ghost and gradient instabilities. With a dynamical scalar graviton this range implies  $w > -1$ . Moreover, there is a special family of models in which the scalar graviton is not dynamical, so that stability criteria essentially relax and this range extends below the phantom divide line  $w = -1$ . In any case, however, this line is not crossed in the course of the cosmological expansion, which makes GUMG model hardly feasible as a candidate for dark energy scenario. On the other hand, this model turns out to be interesting as a possible source of inflation, the scalar graviton playing the role of inflaton having a nontrivial speed of sound, which is briefly discussed in the conclusions of the paper.

## II. GAUGE INVARIANCE IN THE GENERALIZED UNIMODULAR GRAVITY

The GUMG theory action (1.2) can equivalently be rewritten in terms of ten independent metric coefficients with the kinematic relation (1.1) enforced via the Lagrange multiplier  $\lambda$  [1],<sup>1</sup>

$$S_{\text{GUMG}}[g_{\mu\nu}, \lambda] = \int_{[t_-, t_+] \times \Sigma} d^4x [g^{1/2} R - \lambda((-g^{00})^{-1/2} - N(\gamma))] + S_B. \quad (2.1)$$

Here  $S_B = S_{\perp} + S_{\parallel}$  is the boundary term of the Gibbons-Hawking type which effectively removes due to integration by parts second-order spacetime derivatives from the Einstein-Hilbert Lagrangian. It consists of the surface terms  $S_{\perp}$  at the future and past spacelike boundaries  $\Sigma_{\pm}$  at  $t_{\pm}$  and the surface term  $S_{\parallel}$  at the “side” timelike boundary of the topology  $[t_-, t_+] \times \partial\Sigma$ ,  $\partial\Sigma$  being the boundary of spatial slices  $\Sigma$ —their structure will be discussed in more detail below.

<sup>1</sup>For simplicity of the formalism we use the units in which the gravitational coupling constant is dimensionless and equals  $G = 1/16\pi$ .

Addition of these surface terms guarantees consistency of the variational procedure for this action subject to fixed induced metric of the boundary of the spacetime domain  $[t_-, t_+] \times \Sigma$ . As a result the action is stationary on the configurations satisfying the Einstein equations,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{1}{2}T_{\mu\nu}, \quad (2.2)$$

$$T_{\mu\nu} = \frac{2}{\sqrt{g}} \frac{\delta}{\delta g^{\mu\nu}} \int d^4x \lambda ((-g^{00})^{-1/2} - N(\gamma)) \\ = (\varepsilon + p)u_\mu u_\nu + pg_{\mu\nu}, \quad (2.3)$$

with the stress tensor of the effective perfect fluid which has a 4-velocity  $u_\mu = -\delta_\mu^0 N$  and the equation of state with a variable barotropic parameter  $w$ ,

$$p = w\varepsilon, \quad w = 2 \frac{d \ln N(\gamma)}{d \ln \gamma}, \quad \varepsilon = \frac{\lambda}{\sqrt{\gamma}}. \quad (2.4)$$

Nonconstant nature of  $w$  is what distinguishes this model from unimodular gravity (corresponding to  $w = -1$ ) and serves as a main motivation to consider GUMG as a candidate for dark energy which has the phenomenological equation of state with  $w$  essentially depending on the red shift parameter.

The energy density  $\varepsilon$  here is in fact composed of the metric,

$$\varepsilon = 2u^\mu u^\nu G_{\mu\nu}, \quad G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R, \quad (2.5)$$

as it follows from the contraction of (2.2) with  $u^\mu u^\nu$ . Therefore, the ten equations (2.2) are not independent, but represent the projection of the vacuum Einstein equations on the set of nine independent equations

$$P_{\mu\nu}{}^{\rho\sigma} G_{\rho\sigma} = 0, \\ P_{\mu\nu}{}^{\rho\sigma} \equiv \delta^\rho_{(\mu} \delta^\sigma_{\nu)} - [u_\mu u_\nu + w(u_\mu u_\nu + g_{\mu\nu})] u^\rho u^\sigma, \quad (2.6)$$

as it should be for nine independent metric coefficients  $g_{ij}$  and  $g_{0i}$ . Here the projector  $P_{\mu\nu}{}^{\rho\sigma}$  has left and right zero eigenvalue eigenvectors,  $u^\mu u^\nu P_{\mu\nu}{}^{\rho\sigma} = 0$ ,  $P_{\mu\nu}{}^{\rho\sigma} [u_\rho u_\sigma + w(u_\rho u_\sigma + g_{\rho\sigma})] = 0$ .<sup>2</sup>

The canonical formalism of the theory is usually described in terms of the ADM (3 + 1)-decomposition of the spacetime metric into the lapse function  $N$ , shift functions  $N_i$  and the spatial metric  $\gamma_{ij}$  of spacelike slices of a constant time  $x^0 = t$  [15]

$$N = (-g^{00})^{-1/2}, \quad N_i = g_{0i}, \quad N^i = \gamma^{ij} N_j, \\ \gamma_{ij} = g_{ij}, \quad \gamma^{ij} = (\gamma_{ij})^{-1} = g^{ij} + \frac{N^i N^j}{N^2}. \quad (2.7)$$

In terms of these variables subtraction of Gibbons-Hawking boundary terms  $S_\pm$  at the future and past spatial surfaces  $\Sigma_\pm$  leads to the action

$$S_{\text{GUMG}}[\gamma_{ij}, N^i] \\ = \int_{t_-}^{t_+} dt \int_\Sigma d^3x N \sqrt{\gamma} ({}^3R + K_{ij}^2 - K^2)|_{N=N(\gamma)} + S_\pm, \quad (2.8)$$

where the bulk term is the ADM action in terms the extrinsic curvature of the spatial slices

$$K_{ij} = \frac{1}{2N} (\nabla_i N_j + \nabla_j N_i - \dot{\gamma}_{ij}), \quad K = \gamma^{ij} K_{ij}, \quad (2.9)$$

including spatial covariant derivatives  $\nabla_i$  with respect to the 3-metric  $\gamma_{ij}$  and the 3-dimensional scalar curvature  ${}^3R$  of this metric. The lapse function is explicitly expressed as a function of  $\gamma$ , which turns out to be more convenient than introducing this dependence via the Lagrange multiplier.

We will not give here a concrete form of the ‘‘side’’ surface term  $S_\pm$ . It certainly identically vanishes for spatially closed cosmology, and we will also explicitly describe its effect for the linearized theory on the asymptotically flat spacetime, where  $S_-$  is responsible for the ADM energy [15], and asymptotically Friedmann Universe. As we will see, boundary conditions critically influence gauge invariance properties of the GUMG theory, and the mechanisms of this influence are rather different in these two topologically closed and open cases.

Because of the GUMG kinematical restriction on metric coefficients (1.1) the theory is not invariant under local spacetime diffeomorphisms with a generic vector field parameter  $\xi^\mu(t, x)$

$$\delta^\xi g_{\mu\nu} = D_\mu \xi_\nu + D_\nu \xi_\mu, \quad \xi_\mu = g_{\mu\nu} \xi^\nu, \quad (2.10)$$

( $D_\mu$  is the covariant derivative with respect to the metric  $g_{\mu\nu}$ ). As it was assumed in [1], because of general coordinate invariance of the Einstein action the GUMG action (1.2) at least naively is invariant under the subset of diffeomorphisms which preserve the condition (1.1). Rewritten in the basis of ADM variables (2.7) the diffeomorphisms with a generic  $\xi^\mu = (\xi^0, \xi^i)$  read

$$\delta^\xi N = N \dot{\xi}^0 + \dot{N} \xi^0 - N N^l \partial_l \xi^0 + \xi^l \partial_l N, \quad (2.11)$$

$$\delta^\xi N^i = N^i \dot{\xi}^0 + \dot{N}^i \xi^0 - (N^2 \gamma^{ij} + N^i N^j) \partial_j \xi^0 \\ + \dot{\xi}^i + \xi^l \partial_l N^i - N^l \partial_l \xi^i, \quad (2.12)$$

$$\delta^\xi \gamma_{ij} = \dot{\gamma}_{ij} \xi^0 + 2N^l \gamma_{l(i} \partial_{j)} \xi^0 + \xi^l \partial_l \gamma_{ij} + 2\gamma_{l(i} \partial_{j)} \xi^l, \quad (2.13)$$

<sup>2</sup>This explains why every solution of vacuum Einstein theory is also a solution of the GUMG model, as mentioned in Introduction. The same property also holds for Einstein and GUMG theories with matter sources.

so that the preservation of the condition  $N - N(\gamma) = 0$  reduces to the differential equation on  $\xi^\mu$

$$\delta^5[N - N(\gamma)] = N[\partial_t \xi^0 - (1+w)N^k \partial_k \xi^0 - w \partial_t \xi^l] = 0. \quad (2.14)$$

From the viewpoint of the dynamical content of the theory, the invariance transformations of the action can be considered as gauge ones (that is not changing the physical state of the system) only if they are local in time, which means that they should be labelled by independent gauge parameters and their time derivatives to some finite order. Then the invariance transformations can be endowed with a finite support within the full time range  $[t_-, t_+]$  and should vanish together with all their time derivatives at  $t_\pm$ , thus changing fields only inside the spacetime sandwich. This means that the solution of (2.14) for  $\xi^\mu(x)$  should not contain nonlocality in time. At least naively, this does not apply to nonlocal properties of this solution in space—spatially nonlocal but local in time system can be considered as a legitimate gauge constrained system subject to canonical quantization [14]. This means that Eq. (2.14) should be solved for the spatial components  $\xi^i$  of  $\xi^\mu$  in terms of its temporal one  $\xi^0$  and its time derivative, but not vice versa. There are two solutions of this type, which were suggested in [1]—one is given by a purely spatial 3D transversal vector

$$\xi^\mu = \begin{bmatrix} 0 \\ \xi_\perp^i \end{bmatrix}, \quad \partial_k \xi_\perp^k = 0, \quad (2.15)$$

and the second one has a 3D longitudinal component parametrized by a nonvanishing  $\xi^0$  and  $\dot{\xi}^0$

$$\xi^\mu = \begin{bmatrix} \xi^0 \\ \partial^i \frac{1}{\Delta} \frac{\mathcal{D}_t \xi^0}{w} \end{bmatrix}, \quad \mathcal{D}_t = \partial_t - (1+w)N^k \partial_k, \quad (2.16)$$

where  $1/\Delta$  is the nonlocal operation—the inverse of the spatial Laplacian  $\Delta = \partial^i \partial_i$ ,  $\partial^i \equiv \delta^{ij} \partial_j$ . Here an important subtlety arises which distinguishes the case of spatially closed models without a boundary from the case of asymptotically flat spacetimes or Friedmann cosmologies with asymptotically flat spatial slices.

### A. Spatially closed models

Closed GUMG models are important for the theory of cosmological perturbations on the background of the Friedman metric with the  $S^3$ -topology of the spatial slices  $\Sigma$ . To begin with, we will modify the formulation of GUMG theories—incorporate a kind of 3-dimensional bimetric covariance by introducing in the condition (1.1) the dependence on auxiliary spatial metric  $\sigma_{ij}$ ,

$$N(\gamma) \rightarrow N(\gamma/\sigma), \quad \sigma = \det \sigma_{ij}. \quad (2.17)$$

Then this condition becomes a scalar with respect to *simultaneous* coordinate transformations of two spatial metrics  $\gamma_{ij}$  and  $\sigma_{ij}$  and allows one to consider the model in an arbitrary coordinate system on  $\Sigma$ .<sup>3</sup> The auxiliary metric can be taken time independent, but generally has curvature and involves spatial coordinates,  $\sigma_{ij} = \sigma_{ij}(x)$ ,  $x = x^i$ . Then Eq. (2.14) gets modified by a simple replacement of all 3-dimensional partial derivatives with covariant derivatives for the metric  $\sigma_{ij}$ ,

$$\partial_i \rightarrow \bar{\nabla}_i, \quad \bar{\nabla}_k \sigma_{ij} = 0, \quad \bar{\nabla}^i = \sigma^{ij} \bar{\nabla}_j. \quad (2.18)$$

Note that the transformations (2.11)–(2.13) can also be rewritten by this replacement in the bimetric covariant form which involves nonvanishing  $\bar{\nabla}$ -derivatives of the dynamical metric  $\gamma_{ij}$ .<sup>4</sup>

The same rule applies to the solutions (2.15) and (2.16), but there is a problem with the latter. For the nonlocal longitudinal vector

$$\xi_\parallel^i = \bar{\nabla}^i \frac{1}{\Delta} \frac{\mathcal{D}_t \xi^0}{w} \quad (2.19)$$

to be well defined the function  $\mathcal{D}_t \xi^0/w$  should not contain a constant zero mode of the covariant Laplacian  $\bar{\Delta} = \sigma^{ij} \bar{\nabla}_i \bar{\nabla}_j$ . On closed compact  $\Sigma$  this scalar Laplacian has a discrete spectrum with a single constant mode, and this condition reduces to

$$\int_\Sigma d^3x \sqrt{\sigma} \frac{\dot{\xi}^0 - (1+w)N^k \partial_k \xi^0}{w} = 0. \quad (2.20)$$

This is in complete agreement with the fact that the covariant version of the Eq. (2.14) implies on compact space without a boundary that

$$\int_\Sigma d^3x \sqrt{\sigma} \frac{\dot{\xi}^0 - (1+w)N^k \partial_k \xi^0}{w} = \int_\Sigma d^3x \sqrt{\sigma} \bar{\nabla}_i \xi^i \equiv 0. \quad (2.21)$$

<sup>3</sup>This also allows one to avoid coordinate singularities, which in the absence of this additional metric would necessarily mar the formalism. Say, in the case of  $\Sigma = S^3$  in the natural spherical coordinate system the determinant  $\gamma$  vanishes at the poles of the 3-sphere and becomes not invertible, whereas with the choice of  $\sigma_{ij}$  as a metric of the unit 3-sphere the ratio  $\gamma/\sigma$  becomes regular throughout the whole  $S^3$ . We thank A. Kulyabin for this observation.

<sup>4</sup>It should be emphasized that the *bimetric* covariance does not bring in the theory local gauge invariance because an auxiliary metric is not dynamical—it plays in the action the role of external parameter which is not subject to variations in the variational principle.

Therefore,  $\xi^0(t, x)$  should satisfy a nontrivial differential in time equation, so that it cannot have a compact support within the time range and cannot be treated as a local gauge parameter. Thus, on closed compact space  $\xi^0$ -transformation is not a gauge transformation of the theory—it nontrivially changes field configurations at  $t_{\pm}$  and corresponds to a change in the physical state of the system. Only the 3-dimensional transverse vectors (2.15) generate local gauge transformations on a compact closed space<sup>5</sup>—the property that was not envisaged in [1]. Below we will confirm this property both in the canonical formalism of the full nonlinear GUMG theory and by explicit calculations in the linearized theory on the homogeneous Friedmann background.

### B. Asymptotically flat models

In asymptotically flat models breakdown of gauge invariance under  $\xi^0$ -transformations has a different mechanism. In this case the most natural choice of the auxiliary metric is  $\sigma_{ij} = \delta_{ij}$ , the flat space Laplacian has a continuous spectrum and the kernel of its Green's function

$$\frac{1}{\Delta} \delta(x - y) = -\frac{1}{4\pi|x - y|} \sim \frac{1}{|x|}, \quad |x| \rightarrow \infty, \quad (2.22)$$

generates at the spatial infinity the multipole expansion for  $\xi_{\parallel}^i$  in (2.19), beginning with

$$\xi_{\parallel}^i(x) \sim \frac{1}{|x|^2}. \quad (2.23)$$

In contrast to the closed model case no equations imposed on  $\xi^0$  are needed for the existence of this expansion, except the requirement of the integration convergence in the convolution of the Green's function kernel and  $\mathcal{D}_t \xi^0(y)$ . This can be attained by imposing the falloff condition at infinity  $\mathcal{D}_t \xi^0(y) \sim 1/|y|^3$  which is obviously guaranteed for  $\xi^0(x)$  with a compact support. In contrast to  $\xi^0(x)$ , however,  $\xi_{\parallel}^i(x)$  slowly falls off at infinity, so that the surface integral over the remote spacetime boundary turns out to be finite and nonvanishing,

$$\int dt \int_{|x| \rightarrow \infty} d^2 \Sigma_i(x) \xi_{\parallel}^i(x) \neq 0. \quad (2.24)$$

As we show below in the linearized theory, exactly this integral contributes to the gauge transformation of the GUMG action and breaks its invariance. Thus,

<sup>5</sup>Transverse 3-dimensional vectors  $\xi_{\perp}^i$  can be rendered local compact support in the overcomplete basis of dual two-forms by choosing the representation  $\xi_{\perp}^i = \epsilon^{ijk} \bar{\nabla}_j \lambda_k$  in terms of an arbitrary covector field  $\lambda_k$  with a compact support.

$\xi^0$ -transformation also does not belong to local gauge symmetries of the GUMG theory.

### III. BIFURCATION OF CANONICAL CONSTRAINTS

Canonical formalism for the Lagrangian action (2.8) begins with the definition of momenta conjugated to phase space coordinates  $N^i$  and  $\gamma_{ij}$

$$P_i = 0, \quad (3.1)$$

$$\pi^{ij} = -\sqrt{\gamma}(K^{ij} - \gamma^{ij}K), \quad (3.2)$$

and the Legendre transform with respect to phase space velocities from the Lagrangian to the Hamiltonian  $\mathcal{H}$ . Thus we get the primary constraint (3.1), and the Hamiltonian takes the form of the linear combination of what is called in Einstein theory the Hamiltonian  $H_{\perp}$  and momenta  $H_i$  constraints,

$$\mathcal{H} = \int_{\Sigma} d^3x (NH_{\perp} + N^i H_i) + \mathcal{H}_{\perp}, \quad (3.3)$$

$$H_{\perp} = \frac{\gamma_{in} \gamma_{jm} \pi^{ij} \pi^{mn} - \frac{1}{2} \pi^2}{\sqrt{\gamma}} - \sqrt{\gamma}^3 R, \quad \pi = \gamma_{ij} \pi^{ij}, \quad (3.4)$$

$$H_i = -2\gamma_{ij} \nabla_k \pi^{jk}. \quad (3.5)$$

The Hamiltonian also includes a spatial surface term  $\mathcal{H}_{\perp}$  inherited from integrations by parts and the surface term in the action (2.8), but here we will not go into its details and disregard it, though of course it plays an important role by making the variational procedure consistent under fixed boundary conditions at  $\partial\Sigma$ .

As a result, the theory can be equivalently reformulated as a variational principle for the total canonical action including the primary constraints with the Lagrange multipliers  $u^i$

$$S_T[\gamma_{ij}, \pi^{ij}, N^i, P_i, u^i] = \int dt d^3x (\pi^{ij} \dot{\gamma}_{ij} + P_i \dot{N}^i - NH_{\perp} - N^i H_i - u^i P_i). \quad (3.6)$$

In the conventional terminology [14] this is the *total* action of the constrained system. For all phase space functions  $\Phi$  it generates their canonical evolution via the Poisson bracket  $\{\dots, \dots\}$  defined on the space of canonical coordinates and momenta,

$$\dot{\Phi} = \left\{ \Phi, \mathcal{H} + \int d^3y u_1^k(y) P_k(y) \right\}, \quad (3.7)$$

and yields the primary constraints (3.1) by varying their Lagrange multipliers.

Unlike in Einstein theory,  $H_\perp$  here is no longer a constraint because the lapse function is not an independent variable, but the momenta constraints

$$H_i = 0 \quad (3.8)$$

arise as secondary constraints from the requirement of conservation of the primary ones,  $\dot{P}_i = 0$ , because  $\{P_i, \mathcal{H} + \int d^3y u_1^k(y) P_k(y)\} = -H_i$ . In its turn, their conservation leads in view of the Poisson bracket commutator  $\{H_i, \mathcal{H}\} = \partial_k(N^k H_i) + H_k \partial_i N^k + \partial_i(w N H_\perp)$  to the new tertiary constraints

$$T_i = 0, \quad T_i = \partial_i T, \quad T \equiv w N H_\perp, \quad (3.9)$$

where  $w = w(\gamma)$  is given by (2.4). Conservation of  $T_i$  results in the following relation

$$\{T_i, \mathcal{H}\} = \partial_i(w \partial_k(N^2 \gamma^{kl} H_l) + N^k T_k + TS) = 0, \quad (3.10)$$

where

$$S = \Omega \partial_k N^k - \frac{d \ln w \pi N}{d \ln \gamma \sqrt{\gamma}}, \quad (3.11)$$

$$\Omega = 1 + w + 2 \frac{d \ln w}{d \ln \gamma}. \quad (3.12)$$

Omitting in (3.10) the terms which are already proportional to the existing constraints, one can rewrite (3.10) as  $T \partial_i S = 0$ . This immediately leads to two possible quaternary constraints:

$$T = 0 \quad (3.13)$$

(which is equivalent to general relativistic constraint  $H_\perp = 0$ ) or

$$\partial_i S = 0. \quad (3.14)$$

In the terminology of [14] this is a bifurcation of the system of constraints—the theory has two dynamically different branches corresponding to these cases. No further constraints appear in both of these branches. Indeed, in the first branch the conservation of  $T$

$$\{T, \mathcal{H}\} = w \partial_k(N^2 \gamma^{kl} H_l) + N^k T_k + TS = 0 \quad (3.15)$$

is proportional to the existing constraints, while in the second branch the conservation of  $S_i$  is just an equation on the Lagrange multiplier  $u_i^j$ .

All secondary and higher order constraints are the corollaries of dynamical equations. To see this as a by-product and get the Lagrangian expressions for constraints we note that in view of (2.2) the effective stress tensor of the

perfect fluid is covariantly conserved,  $\nabla^\mu T_{\mu\nu} = 0$ . With  $u^\mu = -g^{0\mu} N$  the spatial component of this conservation law reads

$$\nabla^\mu T_{\mu i} = \frac{1}{N \sqrt{\gamma}} \partial_i(\sqrt{\gamma} \epsilon w N) = 0 \quad (3.16)$$

But from (2.4) and (2.5) it follows that  $\epsilon \sqrt{\gamma} = -H_\perp$ , which immediately brings us to the tertiary constraint (3.9). On the other hand, the temporal component of Bianchi identities gives

$$u^\nu \nabla^\mu T_{\mu\nu} = \frac{\epsilon}{N} [(1+w) \partial_k N^k - (\partial_t - N^k \partial_k) \ln \sqrt{\gamma} \epsilon N] = 0, \quad (3.17)$$

where we took into account that  $u^0 = 1/N$ ,  $u^i = -N^i/N$ ,  $p = w\epsilon$  and the expression for  $w$ ,  $w = 2d \ln N / d \ln \gamma$ . If  $\epsilon = 0$ , then we get the  $T = -w\epsilon \sqrt{\gamma} N = 0$  branch of the theory. If  $T \neq 0$ , then one can express the time derivative of  $\sqrt{\gamma} \epsilon N$  from the equation above and use it in the conservation law  $\dot{T}_i = -\partial_i \partial_t(\sqrt{\gamma} \epsilon w N) = 0$ , which gives  $\dot{T}_i = T S_i = 0$ , where

$$S_i = \partial_i [(1+w) \partial_k N^k + (\partial_t - N^k \partial_k) \ln w] \quad (3.18)$$

coincides with the Lagrangian form of the constraint  $S_i$ —this can be directly verified by substituting into (3.11) the Lagrangian expressions for momenta. Therefore, with  $T \neq 0$  we recover the  $S_i = 0$  branch.

The last comment of this section concerns the compact space modification (2.17). Transition to this case in the above formulas with the *time-independent* metric  $\sigma_{ij}$  consists, as was mentioned above, of the replacement of all partial derivatives by covariant derivatives with the Riemannian  $\sigma$ -metric connection, acting on relevant tensors or tensor densities. In particular, since  $H_\perp$  is a scalar density, the covariant derivative  $\bar{\nabla}_i$ , which should replace  $\partial_i$  in the constraint  $T_i$ , reads as

$$T_i = \bar{\nabla}_i T \equiv \sqrt{\sigma} \partial_i \frac{w N H_\perp}{\sqrt{\sigma}}. \quad (3.19)$$

#### IV. ALGEBRA OF CONSTRAINTS FOR TWO BRANCHES OF GUMG THEORY

As we saw, the sets of constraints are different for two bifurcating branches. Here we consider their Poisson bracket algebras and begin with the  $T = 0$  case. In this case we have primary  $P_i$ , secondary  $H_i$ , and tertiary  $T$  constraints which all belong in the Dirac terminology to the first class, because their commutators with each other are vanishing on their full constraint surface  $(P_i, H_i, T) = 0$ ,

$$\{P_i(x), P_j(y)\} = \{P_i(x), H_j(y)\} = \{P_i(x), T(y)\} = 0, \quad (4.1)$$

$$\{H_i(x), H_j(y)\} = H_j(x)\partial_i\delta(x, y) - (i, x \leftrightarrow j, y), \quad (4.2)$$

$$\{H_i(x), T(y)\} = -\delta(x, y)\partial_i T - (\Omega T)(y)\partial_i\delta(y, x), \quad (4.3)$$

$$\{T(x), T(y)\} = (wN)(y)(\gamma^{ij}wNH_i)(x)\partial_j\delta(x, y) - (x \leftrightarrow y), \quad (4.4)$$

and they also commute with the Hamiltonian

$$\{\mathcal{H}, P_i\} = H_i, \quad (4.5)$$

$$\{\mathcal{H}, H_i\} = -\partial_k(N^k H_i) - H_k\partial_i N^k - \partial_i T, \quad (4.6)$$

$$\{\mathcal{H}, T\} = -w\partial_k(N^2\gamma^{kl}H_l) - N^k\partial_k T - TS. \quad (4.7)$$

For this branch of the theory the stress tensor of the effective perfect fluid is obviously vanishing because  $\varepsilon = -T/w\sqrt{\gamma}N = 0$ , and the equations of motion coincide with the vacuum Einstein equations. Obvious interpretation of this GUMG branch is that it is a partial gauge fixing of the Einstein theory in the gauge (1.1).

Interestingly, the algebra of constraints in the  $T = 0$  branch of the theory is very similar to the case of the unimodular gravity, even though in UMG with  $w = -1$  and  $N(\gamma) = 1/\sqrt{\gamma}$ , the function  $T = -H_\perp/\sqrt{\gamma}$  is no longer a constraint, but rather a constant of motion fixed by initial conditions. Indeed, in this case  $\Omega \equiv 0$  and the  $S$ -constraint

(3.11) is absent, so that  $P_i$ ,  $H_i$  and  $T_i$  form the full set of constraints. Moreover, their Poisson bracket algebra is closed—the first two sets of commutators coincide with (4.1) and (4.2), while the rest of them read

$$\{H_i(x), T_j(y)\} = -\partial_j(T_i\delta(y, x)), \quad (4.8)$$

$$\{T_i(x), T_j(y)\} = \partial_i^x\partial_j^y(N(y)(\gamma^{kl}NH_k)(x)\partial_l\delta(x, y)) - (i, x \leftrightarrow j, y). \quad (4.9)$$

These constraints also commute with the Hamiltonian as (4.5) and

$$\{\mathcal{H}, H_i\} = -\partial_k(N^k H_i) - H_k\partial_i N^k - T_i, \quad (4.10)$$

$$\{\mathcal{H}, T_i\} = \partial_i(\partial_k(N^2\gamma^{kl}H_l) - N^k T_k). \quad (4.11)$$

Thus, in unimodular gravity, just like in general relativity, all constraints belong to the first class [2,16,17].

The second branch of  $T \neq 0$  and  $S_i = 0$  is nontrivial and physically much more interesting because it incorporates the effective perfect fluid simulating the role of dark energy. It has four generations of constraints from primary to quaternary ones for which we will use the collective notation

$$\phi_I = (P_i, H_i, T_i, S_i). \quad (4.12)$$

Their Poisson bracket commutators form the matrix

$$\{\phi_I(x), \phi_J(y)\} = \begin{bmatrix} 0 & 0 & 0 & \{P_i(x), S_j(y)\} \\ 0 & 0 & \{H_i(x), T_j(y)\} & \{H_i(x), S_j(y)\} \\ 0 & \{T_i(x), H_j(y)\} & 0 & \{T_i(x), S_j(y)\} \\ \{S_i(x), P_j(y)\} & \{S_i(x), H_j(y)\} & \{S_i(x), T_j(y)\} & 0 \end{bmatrix}. \quad (4.13)$$

Its nonvanishing elements imply that the first class constraints, which commute with each other and with the Hamiltonian, should be disentangled from (4.12) as linear combinations of  $\phi_J(x)$  with some coefficients  $U_A^J$

$$\begin{aligned} \phi_A &= \int d^3x U_A^J(x)\phi_J(x) \\ &= \int d^3x [a_A^i(x)P_i(x) + b_A^j(x)H_j(x) \\ &\quad + c_A^j(x)T_j(x) + d_A^j(x)S_j(x)]. \end{aligned} \quad (4.14)$$

Here we use the first part of the alphabet with capital roman letters  $A, B, C, \dots$  to label the first class constraints

in contrast to indices from the second part of the alphabet  $I, J, \dots$  labeling the original set of spatially local constraints. The nature of these indices can be very different depending on the choice of basis in the space of  $\phi_A$ . To begin with, the constraints  $\phi_A$  in contrast to  $\phi_I$  almost always are spatially nonlocal like, say, irreducible components of transverse vectors and tensors. Therefore, these indices should include both discrete labels with a finite range and continuous or countable labels with an infinite range, like for example a momentum of the Fourier transform  $\mathbf{k}$ ,  $A \mapsto A = 1, 2, \dots; \mathbf{k}$ . We will not specify them here, because they might be very different in different cases of closed compact or infinite open space  $\Sigma$ . We will only assume the DeWitt rule for contraction of these repeated

indices, which implies not only the summation over the discrete labels in  $A$ , but also the integration over its continuous part.

With the choice of first class constraints in the form (4.14) the coefficients  $U_A^J = a_A^j, b_A^j, c_A^j, d_A^j$  should form on the subspace of vanishing constraints in phase space a zero eigenvalue eigenvector of the matrix of constraint commutators

$$\{\phi_A, \phi_I(y)\}|_{\phi_I=0} \equiv \int d^3x U_A^J(x) \{\phi_J(x), \phi_I(y)\}|_{\phi_I=0} = 0. \quad (4.15)$$

Using the following nonvanishing elements of this matrix

$$\begin{aligned} \{P_i(x), S_j(y)\} &= -\partial_j^y(\Omega(y)\partial_i\delta(y,x)), \\ \{H_i(x), H_j(y)\} &= H_j(x)\partial_i\delta(x,y) - (i,x \leftrightarrow j,y), \\ \{H_i(x), T_j(y)\} &= -T(y)\partial_j^y(\Omega(y)\partial_i\delta(y,x)) \\ &\quad - \Omega(y)T_j(y)\partial_i\delta(y,x) - \partial_j^y(T_i(y)\delta(y,x)), \\ \{H_i(x), S_j(y)\} &= -\partial_j^y \left[ S_i\delta(y,x) - \Omega\partial_i\partial_k N^k\delta(y,x) \right. \\ &\quad \left. + \frac{d\Omega(y)}{d\ln\gamma} \left( 2\partial_k N^k - \frac{\pi N}{\sqrt{\gamma}} \right) (y)\partial_i\delta(y,x) \right], \\ \{T_i(x), T_j(y)\} &= \partial_i\partial_j^y((wN)(y)(wN\gamma^{kl}H_k)(x)\partial_l\delta(x,y)) \\ &\quad - (x \leftrightarrow y), \\ \{T_i(x), S_j(y)\} &= \partial_i\partial_j^y(\dots), \end{aligned}$$

(where we do not explicitly specify the last rather complicated commutator because it will not be needed in what follows) one has the set of explicit equations on  $U_A^J(x)$

$$\begin{aligned} \partial_i(\Omega\partial_j d_A^j) &= 0, \\ \partial_i(\Omega T\partial_j c_A^j) + \int d^3y \{H_i(x), S(y)\}\partial_j d_A^j(y) &= 0, \\ T\partial_i(\Omega\partial_j b_A^j) - \int d^3y \{T_i(x), S(y)\}\partial_j d_A^j(y) &= 0, \\ \partial_i(\Omega\partial_j a_A^j) + \int d^3y \{S_i(x), H_j(y)\}b_A^j(y) \\ - \int d^3y \{S_i(x), T(y)\}\partial_j c_A^j(y) &= 0. \end{aligned}$$

The solution of the first three equations gives the condition of transversality of  $d_A^j, c_A^j$ , and  $b_A^j$

$$\partial_j d_A^j = 0, \quad \partial_j c_A^j = 0, \quad \partial_j b_A^j = 0, \quad (4.16)$$

whereas in view of the relation

$$\int d^3y \{S_i(x), H_j(y)\}b_A^j(y) = -\partial_i[\Omega\partial_j(b_A^j\partial_k N^k) - b_A^j S_j], \quad (4.17)$$

the fourth equation takes the form  $\partial_i[\Omega\partial_j(a_A^j - b_A^j\partial_k N^k)] = 0$  and has two solutions. The first is  $a_A^j = b_A^j\partial_k N^k$  when  $b_A^j \neq 0$  and the second is  $\partial_j a_A^j = 0$ , otherwise. Therefore, the full set of first-class constraints is

$$P_A = \int d^3x a_A^i(x) P_i(x), \quad (4.18)$$

$$H_A = \int d^3x b_A^i(x) (H_i(x) + \partial_k N^k(x) P_i(x)). \quad (4.19)$$

Here  $a_A^i(x)$  and  $b_A^i(x)$  both form a complete set of transversal vectors enumerated by condensed indices which were discussed above. Without losing generality these two sets can be identified, which we will do in what follows,

$$a_A^i(x) = b_A^i(x) = e_A^i(x), \quad \partial_i e_A^i(x) = 0, \quad (4.20)$$

$e_A^i(x)$  representing some complete basis in the space of all transverse vector fields. In addition to this basis it is useful to introduce the dual basis  $e_i^B(x)$  which has the following properties

$$\begin{aligned} \int d^3x e_A^i(x) e_i^B(x) &= \delta_A^B, \\ e_A^i(x) e_j^A(y) &= \Pi_j^i(x, y), \quad \partial_i \Pi_j^i(x, y) = 0. \end{aligned} \quad (4.21)$$

Here  $\Pi_j^i(x, y)$  is a projector on the space of transverse vectors. This dual basis allows one to show that the first class constraints form the subalgebra—their commutators express as their own linear combinations without contributions of second class constraints. Calculating the commutators of (4.18)–(4.19) one can see that in the resulting expressions the original constraints  $P_i(x)$  and  $H_i(x)$  get functionally contracted (that is with summation over the discrete index  $i$  and integration over space) with some transverse vectors  $V^i(x)$ . Since  $\Pi_j^i V^j \equiv \int d^3y \Pi_j^i(x, y) V^j(y) = V^i(x)$ , one can insert into these contractions the projector  $\Pi_j^i$  and then use its bilinear decomposition in the basis of  $e_A^i$  and its dual. This converts the resulting answer into the linear combinations of  $P_A$  and  $H_A$  which read

$$\{P_A, P_B\} = \{P_A, H_B\} = 0, \quad (4.22)$$

$$\begin{aligned} \{H_A, H_B\} &= H_C \int d^3x e_i^C (e_A^j \partial_j e_B^i - e_B^j \partial_j e_A^i) \\ &\quad + P_C \int d^3x e_i^C \partial_j [\partial_k N^k (e_A^i e_B^j - e_B^i e_A^j)]. \end{aligned} \quad (4.23)$$



Similarly we have Poisson brackets of the first-class and second-class constraints

$$\{P_A, P_i\} = \{P_A, H_i\} = \{P_A, T_i\} = \{P_A, S_i\} = 0, \quad (4.24)$$

$$\{H_A, P_i\} = -\partial_i(e_A^k P_k), \quad (4.25)$$

$$\{H_A, H_i\} = -e_A^k \partial_k H_i - H_k \partial_i e_A^k, \quad (4.26)$$

$$\{H_A, T_i\} = -\partial_i(e_A^k T_k), \quad (4.27)$$

$$\{H_A, S_i\} = -\partial_i(e_A^k S_k), \quad (4.28)$$

and the Poisson brackets of first-class constraints with the Hamiltonian

$$\{\mathcal{H}, P_A\} = \int d^3x e_A^i H_i, \quad (4.29)$$

$$\{\mathcal{H}, H_A\} = \int d^3x H_i \partial_j [N^j e_A^i - N^i e_A^j] \quad (4.30)$$

## V. CANONICAL REALIZATION OF LAGRANGIAN SYMMETRIES

According to the conventional canonical formalism of gauge constrained systems [14], the number of Lagrangian gauge symmetries of the theory equals the number of the primary first class constraints, which is the essence of the so-called Dirac conjecture. Under a number of assumptions on the structure of constraint algebra and the distribution of constraints in several generations (primary, secondary, etc.) this conjecture was proven in [14]. In our case of GUMG theory in its physically interesting  $S_i = 0$ ,  $T \neq 0$  branch the situation seems very involved and even contradictory—the number of anticipated local gauge symmetries of the Lagrangian action is three, while there are four first class constraints. Moreover, these constraints do not belong to a concrete generation, but represent linear combinations of the original primary and secondary constraints. In addition we have second class constraints in all four generations, not to say that the tertiary,  $T_i = 0$ , and quaternary,  $S_i = 0$ , constraints are reducible in view of their longitudinal nature. In this section we will apply a modified version of the Henneaux-Teitelboim procedure [14] to recover the Lagrangian symmetries of the theory and show that it actually satisfies the Dirac conjecture, even though it does not satisfy the set of assumptions made in [14]. Quite interestingly, this canonical procedure recovers only two gauge symmetries of the GUMG action—spatial diffeomorphisms with the transverse vector field (2.15) and prohibits the  $\xi^0$ -diffeomorphism (2.16) even despite the fact that at the Lagrangian level the latter falls out of the category of local gauge symmetries entirely due to subtle nonlocal behavior at spacetime boundaries.

### A. Symmetries of the extended and total actions

The recovery of Lagrangian symmetries of generic constrained systems begins in [14] with the construction of the *extended* canonical action

$$S_E = \int dt (p_k \dot{q}^k - \mathcal{H} - u^m \phi_m), \quad (5.1)$$

which includes together with the primary constraints  $\phi_{m_1}$  a full set of all the constraints  $\phi_m$  and their conjugated Lagrange multipliers  $u^m$ , consisting of  $M$  (primary, secondary, etc.) generations,

$$\phi_m = (\phi_{m_1}, \phi_{m_2}, \dots, \phi_{m_M}), \quad m = m_1, \dots, m_M. \quad (5.2)$$

Here  $m_i$ ,  $i = 1, \dots, M$ , enumerates the members of the  $i$ th generation. In general, the first class constraints  $\phi_A$  do not belong to a concrete generation but rather represent a linear combination of  $\phi_m$  mixing the constraints from different generations with some coefficients  $U_A^m$ —zero-vectors of the matrix  $\{\phi_n, \phi_m\}|_{\phi=0}$  [cf., Eq. (4.15)],

$$\phi_A = U_A^m \phi_m. \quad (5.3)$$

Off the constraint surface in phase space they satisfy the following commutation relations with some structure functions,

$$\{\phi_A, \phi_m\} = C_{Am}^n \phi_n, \quad \{\mathcal{H}, \phi_A\} = V_A^m \phi_m, \quad (5.4)$$

and allow one to have a canonical transformation of phase-space variables which preserves the constraints and the Hamiltonian

$$\delta \left( \begin{matrix} q^n \\ p_n \end{matrix} \right) = \left\{ \left( \begin{matrix} q^n \\ p_n \end{matrix} \right), \phi_A \mu^A \right\}. \quad (5.5)$$

Here  $\mu^A$  are some gauge parameters generally depending explicitly on time, phase-space variables and Lagrange multipliers,  $\mu^A = \mu^A(t, q, p, u)$ .

This canonical transformation can be accompanied by the transformation of the Lagrange multipliers  $u^m$  which together with (5.5) leaves the extended action invariant when the parameters  $\mu^A$  have in time a compact support. In view of the transformations of the symplectic term

$$\delta(p_k \dot{q}^k) = \frac{d}{dt} \left[ \left( p_k \frac{\partial}{\partial p_k} - 1 \right) \phi_A \mu^A \right] + \phi_A \frac{D\mu^A}{Dt}, \quad (5.6)$$

where  $D\mu^A/Dt$  denotes the partial time derivative acting only on explicit time dependence on time and Lagrange multipliers

$$\begin{aligned} \frac{D\mu^A}{Dt} &= \left( \partial_t + \dot{u}^m \frac{\partial}{\partial u^m} \right) \mu^A(t, q, p, u) \\ &= \left( \frac{d}{dt} - \dot{q}^k \frac{\partial}{\partial q^k} - \dot{p}_k \frac{\partial}{\partial p_k} \right) \mu^A(t, q, p, u), \end{aligned} \quad (5.7)$$

the variation of the extended action (after omitting the total derivative term) takes the form

$$\begin{aligned} \delta S_E &= \int dt \phi_m \left[ U_A^m \left( \frac{D\mu^A}{Dt} + \{ \mu^A, \mathcal{H} + u^n \phi_n \} \right) \right. \\ &\quad \left. + C_{An}^m \mu^A u^n - V_A^m \mu^A - \delta u^m \right]. \end{aligned} \quad (5.8)$$

This vanishes under an obvious transformation law for the Lagrange multipliers

$$\begin{aligned} \delta u^m &= U_A^m \left( \frac{D\mu^A}{Dt} + \{ \mu^A, \mathcal{H} + u^n \phi_n \} \right) \\ &\quad + C_{An}^m \mu^A u^n - V_A^m \mu^A. \end{aligned} \quad (5.9)$$

Note that this equation includes as a particular case the situation when the original set of constraints was already split into the first and second class constraints. In this case  $U_A^m = \delta_A^m$ , and the transformations of Lagrange multipliers also split into those of the first class constraints with the explicit time derivative of  $\mu^A$  and those of the second class ones without this time derivative.

Transformations (5.5) and (5.9) leave the extended action invariant *off shell* for arbitrary values of all (primary and higher) Lagrange multipliers  $u^m$ , and their number equals the number of *all first class constraints*. The number of Lagrangian gauge transformations is smaller and they act on the Lagrangian action functional having a smaller number of arguments—only Lagrangian coordinates  $q^k$ . These transformations are in one-to-one correspondence with the transformations of the so-called *total* canonical action [14] including, in contrast to the extended action (5.1), only the primary constraints,

$$S_T = \int dt (p_k \dot{q}^k - \mathcal{H} - u^{m_1} \phi_{m_1}). \quad (5.10)$$

This action can be regarded as a partial gauge fixing of the gauge symmetries of the extended action in a special gauge of vanishing secondary and higher Lagrange multipliers  $u^{m_i} = 0$ ,  $i = 2, \dots, M$ . Therefore, its invariance transformations are residual gauge transformations (5.5) and (5.9) with the special values of parameters  $\mu^A = \mu^A(t, q, p, u)$  preserving this gauge,  $\delta u^{m_i} = 0$  for  $i = 2, \dots, M$ . In view of the above equation for  $\delta u^m$  this leads to the relation

$$\begin{aligned} U_A^{m_i} \left( \frac{D\mu^A}{Dt} + \{ \mu^A, \mathcal{H} + u^{m_1} \phi_{m_1} \} \right) \\ + C_{Am_1}^{m_i} \mu^A u^{m_1} - V_A^{m_i} \mu^A = 0, \quad i = 2, \dots, M, \end{aligned} \quad (5.11)$$

which should be treated as a set of equations on  $\mu^A$ . When this set can be solved for a subset of parameters  $\mu^A$  *ultralocally in time* in terms of a complementary subset of independent parameters and their time derivatives, then the resulting transformations (5.5) and (5.9) with  $m = m_1$  will form the gauge symmetries of the total action which are equivalent to the Lagrangian symmetries of the theory. Obviously, the number of these symmetries coincides with the range of this subset of independent parameters  $\mu^A$  complementary to the dependent ones.

When all the constraints are of the first class, Eq. (5.11) can be further simplified. Indeed, in this case  $U_A^m$  becomes an identity matrix, so that the indices  $m$  and  $A$  become identified. Moreover, the procedure of constraints derivation implies the following simplifications of structure functions:  $V_{m_j}^{m_i} = 0$  when  $i > j + 1$  and  $C_{m_j, m_1}^{m_i} = 0$  when  $i > j$  (following from the fact that the  $(i + 1)$ th generation of constraint is generated only by the first  $i$  generations, and because nonvanishing values of  $C_{m_j, m_1}^{m_i}$  with  $i > j + 1$  would imply the equation on the Lagrange multiplier  $u^{m_1}$  rather than the new constraint [14]). This allows one to rewrite Eq. (5.11) as the following chain of equations

$$\begin{aligned} V_{m_{i-1}}^{m_i} \mu^{m_{i-1}} &= \frac{D\mu^{m_i}}{Dt} + \{ \mu^{m_i}, \mathcal{H} + u^{m_1} \phi_{m_1} \} \\ &\quad + \sum_{j=i}^M (C_{m_j, m_1}^{m_i} u^{m_1} - V_{m_j}^{m_i}) \mu^{m_j} \end{aligned} \quad (5.12)$$

with  $i = 2, \dots, M$ , where the summation over generations runs for each  $i$  from  $i$  to  $M$ . This system of equations can be solved for  $\mu^{m_{i-1}}$  recursively starting from  $i = M$  and taking  $\mu^{m_M}$  as independent parameters. This procedure ends at the step  $i = 2$ , in which we evaluate  $\mu^{m_1}$  in terms of these independent parameters.

## B. Symmetries of the GUMG action

Here we apply the above procedure to two different bifurcating branches of the GUMG theory. The branch of  $T = 0$  physically is less interesting because it corresponds to the vacuum general relativistic model without the perfect fluid. However, to demonstrate the recovery of spacetime diffeomorphisms restricted by the condition (1.1), we start with this particular branch and the related case of the unimodular gravity—both having only first class constraints.

### 1. Symmetries of $T = 0$ branch

All the constraints  $P_i$ ,  $H_i$ , and  $T$  in the extended action of the  $T = 0$  branch

$$S_E = \int dt d^3x (\pi^{ij} \dot{\gamma}_{ij} + P_i \dot{N}^i - NH_{\perp} - N^i H_i - u_1^i P_i - u_2^i H_i - u_3 T), \quad (5.13)$$

belong to the first class and, according to the previous section, give rise to extended gauge transformations, which we derive below. For this purpose, we denote  $P_i(x)$ ,  $H_i(x)$ , and  $T(x)$  respectively as  $\phi_{m_1x}$ ,  $\phi_{m_2x}$ , and  $\phi_{3x} \equiv \phi_x$  and rewrite the algebra of these constraints (4.1) and (4.5)–(4.7) in condensed notations with  $m_i \mapsto m_{1x}, m_{2x}, m_{3x} \equiv x$ , the index  $i = 1, 2, 3$  marking in accordance with (5.2) the constraints generation. Then in view of (4.1)  $\{\phi_{m_1x}, \phi_{m_1y}\} = 0$ , and the corresponding structure functions turn out to be vanishing,  $C_{m_1x, m_1y}^{m_jz} = 0$ . In these notations the relations (4.5)–(4.7) read as

$$\{\mathcal{H}, \phi_{m_1x}\} = V_{m_1x}^{m_2y} \phi_{m_2y}, \quad (5.14)$$

$$\{\mathcal{H}, \phi_{m_2x}\} = V_{m_2x}^{n_2y} \phi_{n_2y} + V_{m_2x}^y \phi_y, \quad (5.15)$$

$$\{\mathcal{H}, \phi_x\} = V_x^{m_2y} \phi_{m_2y} + V_x^y \phi_y, \quad (5.16)$$

where the integration over  $y$  is implied and the remaining structure functions equal

$$\begin{aligned} V_{m_1x}^{m_2y} &= \delta(x, y) \delta_{m_1}^{m_2}, \\ V_{m_2x}^{n_2y} &= -\partial_k (N^k \delta(x, y)) \delta_{m_2}^{n_2} - \delta(x, y) \partial_{m_2} N^{n_2}, \\ V_{m_2x}^y &= -\partial_{m_2} \delta(x, y), \\ V_x^{m_2y} &= -w(x) \partial_k (N^2 \gamma^{km_2} \delta(x, y)), \\ V_x^y &= -N^k(x) \partial_k \delta(x, y) - S \delta(x, y). \end{aligned}$$

In the case of solely first class constraints the matrix  $U_A^m$  in (5.3) is a unit one,  $U_A^m = \delta_A^m$ , and Eq. (5.12) for the parameters of residual gauge transformations,  $\mu^A \mapsto (\mu_1^{m_1y}, \mu_2^{m_2y}, \mu_3^y)$ , takes the form of the following two equations

$$\mu_2^{m_2y} V_{m_2y}^x = \frac{D\mu_3^x}{Dt} + \{\mu_3^x, \mathcal{H}\} + u^{m_1y} \{\mu_3^x, \phi_{m_1y}\} - \mu_3^y V_y^x, \quad (5.17)$$

$$\begin{aligned} \mu_1^{m_1y} V_{m_1y}^{m_2x} &= \frac{D\mu_2^{m_2x}}{Dt} + \{\mu_2^{m_2x}, \mathcal{H}\} + u^{m_1y} \{\mu_2^{m_2x}, \phi_{m_1y}\} \\ &\quad - \mu_2^{m_2y} V_{m_2y}^{m_2x} - \mu_3^y V_y^{m_2x}. \end{aligned} \quad (5.18)$$

When the parameter  $\mu_3$  is rewritten in terms of a new independent parameter  $\xi^0$ ,  $\mu_3 = \xi^0/w$ , the first equation (5.17) explicitly reads as

$$\begin{aligned} \partial_k \mu_2^k &= \frac{1}{w} (\dot{\xi}^0 - N^k \partial_k \xi^0) \\ &\quad - [\partial_k N^k + \{\ln w, \mathcal{H}\} - N^k \partial_k \ln w - S] \xi^0, \end{aligned} \quad (5.19)$$

and in view of the expression for  $S$ ,  $S = (1+w)\partial_k N^k + \{\ln w, \mathcal{H}\} - N^k \partial_k \ln w$ —a simple corollary of Eqs. (3.11)–(3.12), it finally takes the form

$$\partial_k (\mu_2^k - \xi^0 N^k) = \frac{1}{w} \mathcal{D}_t \xi^0, \quad (5.20)$$

where  $\mathcal{D}_t$  is defined in Eq. (2.16),  $\mathcal{D}_t = \partial_t - (1+w)N^k \partial_k$ . One solution of this equation is parametrized by an arbitrary function  $\xi^0$ ,

$$\mu_2^i = \xi^0 N^i + \partial^i \frac{1}{w\Delta} \mathcal{D}_t \xi^0, \quad (5.21)$$

while another one is an arbitrary transverse vector with  $\xi^0 \equiv 0$ ,

$$\mu_2^i = \xi_{\perp}^i, \quad \partial_k \xi_{\perp}^k = 0. \quad (5.22)$$

The explicit form of the second equation (5.18),

$$\begin{aligned} \mu_1^i &= \frac{D\mu_2^i}{Dt} + \left\{ \mu_2^i, \mathcal{H} + \int d^3y u_1^k P_k \right\} + \mu_2^k \partial_k N^i \\ &\quad - N^k \partial_k \mu_2^i - N^2 \gamma^{ik} \partial_k (w\mu_3), \end{aligned} \quad (5.23)$$

can now be used for the determination of the parameter  $\mu_1^i$ . For the solution (5.21) it can be rewritten as

$$\begin{aligned} \mu_1^i &= \dot{\xi}^0 N^i + \xi^0 u_1^i - (N^2 \gamma^{ij} + N^i N^j) \partial_j \xi^0 \\ &\quad + \left\{ \partial^i \frac{1}{w\Delta} \mathcal{D}_t \xi^0, \mathcal{H} + \int d^3x u_1^k P_k \right\} + \partial^i \frac{1}{w\Delta} \mathcal{D}_t \xi^0 \\ &\quad + (\partial_j N^i - \delta_j^i N^l \partial_l) \partial^j \frac{1}{w\Delta} \mathcal{D}_t \xi^0, \end{aligned} \quad (5.24)$$

and

$$\mu_1^i = (\partial_t - N^k \partial_k) \xi_{\perp}^i + \xi_{\perp}^k \partial_k N^i \quad (5.25)$$

for the solution (5.22).

Thus, we have the following two canonical generators of gauge transformations of the total GUMG action (3.6) in the branch  $T = 0$ . The first one is

$$G_0 = \int d^3x \left( P_i \mu_1^i + H_i \mu_2^i + T \frac{1}{w} \xi^0 \right), \quad (5.26)$$

where  $\xi^0$  is an independent parameter and the parameters  $\mu_1$  and  $\mu_2$  are defined in terms of  $\xi^0$  respectively by (5.24) and (5.21). The second one is

$$G_{\perp} = \int d^3x (P_i \mu_1^i + H_i \xi_{\perp}^i), \quad (5.27)$$

with an arbitrary transverse vector  $\xi_{\perp}^i$  and the vector  $\mu_1^i$  defined in terms of  $\xi_{\perp}^i$  by Eq. (5.25).

To compare the action of these generators on the variables  $N^i$  and  $\gamma_{ij}$  with their spacetime diffeomorphisms in the Lagrangian formalism, consider first the transformation of  $N^i$  by the generator  $G_0$ ,  $\delta_0 N^i = \{N^i, G_0\} = \mu_1^i$ . Note that for the Lagrangian values of canonical momenta (that is, when the equation of motion for the momenta is satisfied)  $u_1^i = \dot{N}^i$  and the two (fourth and fifth) terms in (5.24) can be rewritten as

$$\begin{aligned} & \left\{ \partial^i \frac{1}{w\Delta} \mathcal{D}_i \xi^0, \mathcal{H} + \int d^3x u_1^i P_i \right\} + \partial^i \frac{1}{w\Delta} \mathcal{D}_i \xi^0 \\ &= \partial_t \partial^i \frac{1}{w\Delta} \mathcal{D}_i \xi^0, \end{aligned} \quad (5.28)$$

whence

$$\begin{aligned} \delta_0 N^i &= N^i \dot{\xi}^0 + \dot{N}^i \xi^0 - (N^2 \gamma^{ij} + N^i N^j) \partial_j \xi^0 \\ &+ [\delta_j^i (\partial_t - N^l \partial_l) + \partial_j N^i] \partial^j \frac{1}{w\Delta} \mathcal{D}_i \xi^0, \end{aligned} \quad (5.29)$$

which is just the diffeomorphism transformation law of the shift function (2.12) with the spacetime vector parameter (2.16). A similar calculation for the canonical transformation  $\delta_{\perp} N^i = \{N^i, G_{\perp}\} = \mu_1^i$  with  $\mu_1^i$  given by (5.25) shows that it is just a diffeomorphism with the parameter (2.16).

Next, considering the gauge transformations of  $\gamma_{ij}$  take into account that with the Lagrangian value of the momentum  $P_i = 0$  the first term in (5.26) and (5.27) vanishes, so that the commutator of  $\gamma_{ij}$  with  $\mu_1^i$  does not contribute at all. Moreover, note that the generator of the form  $G = \int d^3x (H_{\perp} F^{\perp} + H_i F^i)$  gives rise to the diffeomorphism of  $\gamma_{ij}$  with the vector parameter  $\xi^{\mu}$  defined by the relations  $\xi^0 = F^{\perp}/N$  and  $\xi^i = F^i - N^i F^{\perp}/N$  ( $F^{\perp}$  and  $F^i$  are respectively the normal and tangential projections of  $\xi^{\mu}$  on the spatial slice of constant  $x^0 = t$ ). Therefore, comparison with the generators (5.26) and (5.27) (remember that  $T = wNH_{\perp}$ ) shows that  $G_0$  defines the diffeomorphism of  $\gamma_{ij}$  with the parameter (2.16), while  $G_{\perp}$  defines a spatial diffeomorphism with the parameter (2.15), which is in a complete agreement with the gauge symmetries of the Lagrangian formalism.

## 2. Symmetries of the unimodular gravity theory

The UMG theory, which is directly related to the  $T = 0$  branch of GUMG, also demonstrates the canonical realization of all three diffeomorphism symmetries. In this case the extended action differs from (5.13) only by the last term which should read as  $-u_3^i T_i$ . Similarly, the full set of first

class constraints  $P_m(x)$ ,  $H_m(x)$ , and  $T_m(x)$  denoted respectively as  $\phi_{m_1x}$ ,  $\phi_{m_2x}$ , and  $\phi_{m_3x}$  differs from that of the  $T = 0$  branch by the replacement  $\phi_x \rightarrow \phi_{m_3x}$ . Their algebra again has structure functions with vanishing components  $C_{m_1x, m_iy}^{m_jz} = 0$ , and their commutators with the Hamiltonian repeat the relations (5.14)–(5.16) with obvious replacements  $\phi_x \rightarrow \phi_{m_3x}$ ,  $V_{m_2x}^y \rightarrow V_{m_2x}^{m_3y}$ ,  $V_x^{m_2y} \rightarrow V_{m_3x}^{m_2y}$ , and  $V_x^y \rightarrow V_{m_3x}^{n_3y}$ , where

$$V_{m_2x}^{m_3y} = -\delta_{m_2}^{m_3} \delta(x, y), \quad (5.30)$$

$$V_{m_3x}^{m_2y} = \partial_{m_3} \partial_k (N^2 \gamma^{km_2} \delta(x, y)), \quad (5.31)$$

$$V_{m_3x}^{n_3y} = -\partial_{m_3} (N^{n_3} \delta(x, y)). \quad (5.32)$$

Correspondingly, the Eqs. (5.12) take the form

$$\begin{aligned} (V_{m_2y}^{m_3x} \mu_2^{m_2y})_{\parallel} &= \left( \frac{D\mu_3^{m_3x}}{Dt} + \{\mu_3^{m_3x}, \mathcal{H}\} \right. \\ &+ \left. \{\mu_3^{m_3x}, \phi_{m_1y}\} u^{m_1y} - V_{n_3y}^{m_3x} \mu_3^{n_3y} \right)_{\parallel}, \end{aligned} \quad (5.33)$$

$$\begin{aligned} \mu_1^{m_2x} &= \frac{D\mu_2^{m_2x}}{Dt} + \{\mu_2^{m_2x}, \mathcal{H}\} + \{\mu_2^{m_2x}, \phi_{n_1y}\} u^{n_1y} \\ &- V_{n_2y}^{m_2x} \mu_2^{n_2y} - V_{n_3y}^{m_2x} \mu_3^{n_3y}, \end{aligned} \quad (5.34)$$

where only the longitudinal part of the first vector equation is enforced, since it is contracted in the transformation of the action with the longitudinal vector  $T_{m_3} \equiv \partial_{m_3} T$ .

Choosing independent longitudinal vector  $\mu_3^k$  as  $\mu_3^k = \partial^k (1/\Delta) \xi^0$  with an arbitrary scalar parameter  $\xi^0$ , one finds a general solution of Eq. (5.33)

$$\mu_2^k = -\partial^k \frac{1}{\Delta} \xi^0 + N^k \xi^0 + \xi_{\perp}^k, \quad (5.35)$$

where  $\xi_{\perp}^k$  is an arbitrary transverse vector. Therefore, it follows from (5.34) that for the  $\xi^0$ -transformation with  $\xi_{\perp}^k = 0$

$$\begin{aligned} \mu_1^i &= \dot{\xi}^0 N^i + \xi^0 u_1^i - (N^2 \gamma^{ij} + N^i N^j) \partial_j \xi^0 \\ &- [\delta_j^i (\partial_t - N^l \partial_l) + \partial_j N^i] \partial^j \frac{1}{\Delta} \xi^0, \end{aligned} \quad (5.36)$$

where we took into account that in UMG theory  $w = -1$ , the operator  $\mathcal{D}_i \equiv \partial_i$  is field independent and, therefore, does not contribute to the Poisson bracket commutator term in Eq. (5.28). Similarly  $\mu_1^i = (\partial_t - N^k \partial_k) \xi_{\perp}^i + \partial_k N^i \xi_{\perp}^k$  for transformations with  $\xi^0 = 0$ . Thus, as in the  $T = 0$  branch of GUMG theory, we again obtain two generating functions (5.26) and (5.27) with  $T/w = NH_{\perp}$ , which reproduce at the Lagrangian level the diffeomorphisms with volume preserving vector parameters (2.16) and (2.15),  $\partial_{\mu} \xi^{\mu} = 0$ .

### 3. Symmetries of $T \neq 0$ , $S_i = 0$ branch

Extended action of this GUMG branch is of the form

$$S_E = \int dt d^3x (\pi^{ij} \dot{\gamma}_{ij} + P_i \dot{N}^i - N(\gamma) H_\perp - N^i H_i - u_1^i P_i - u_2^i H_i - u_3^i T_i - u_4^i S_i), \quad (5.37)$$

where the first-class constraints  $\phi_A = (P_{A_1}, H_{A_2})$ ,  $A \mapsto A_1, A_2$ , are hidden in the full set of constraints  $\phi_{mx} = (P_{m_1x}, H_{m_2x}, T_{m_3x}, S_{m_4x})$  as the linear combinations (4.18) and (4.19),  $\phi_A = U_A^{mx} \phi_{mx}$  with the following nonvanishing components of  $U_A^{mx}$ ,

$$\begin{aligned} U_{A_1}^{m_1x} &= e_{A_1}^{m_1}(x), & U_{A_2}^{m_1x} &= e_{A_2}^{m_1}(x) \partial_k N^k(x), \\ U_{A_2}^{m_2x} &= e_{A_2}^{m_2}(x). \end{aligned} \quad (5.38)$$

In view of (4.24)–(4.25) and (4.29)–(4.30) their algebra reads as

$$\begin{aligned} \{P_{A_1}, \phi_{m_1x}\} &= \{P_{A_1}, \phi_{m_2x}\} = \{P_{A_1}, \phi_{m_3x}\} \\ &= \{P_{A_1}, \phi_{m_4x}\} = 0, \end{aligned} \quad (5.39)$$

$$\{H_{A_2}, \phi_{m_1x}\} = C_{A_2, m_1x}^{n_1y} \phi_{n_1y}, \quad (5.40)$$

$$\{\mathcal{H}, P_{A_1}\} = V_{A_1}^{m_2x} \phi_{m_2x}, \quad (5.41)$$

$$\{\mathcal{H}, H_{A_2}\} = V_{A_2}^{m_1x} \phi_{m_1x} + V_{A_2}^{m_2x} \phi_{m_2x}, \quad (5.42)$$

where

$$V_{A_1}^{m_2x} = e_{A_1}^{m_2}(x), \quad (5.43)$$

$$V_{A_2}^{m_2x} = \partial_k [N^k(x) e_{A_2}^{m_2}(x) - N^{m_2}(x) e_{A_2}^k(x)]. \quad (5.44)$$

To obtain the generator of canonical transformation which leaves the total action invariant, one should solve in accordance with the Henneaux-Teitelboim procedure the Eqs. (5.11) for  $\mu^A$  in terms of some independent parameters. For  $i = 3, 4$  these equations are satisfied identically since

$$U_A^{m_i x} = C_{A, m_i y}^{m_i x} = V_A^{m_i x} = 0, \quad i = 3, 4. \quad (5.45)$$

In the case of  $i = 2$  the equation takes the form

$$\begin{aligned} U_A^{m_2x} \left( \frac{D\mu^A}{Dt} + \{\mu^A, \mathcal{H} + u^{m_1y} \phi_{m_1y}\} \right) \\ + C_{A, m_1y}^{m_2x} \mu^A u^{m_1y} - V_A^{m_2x} \mu^A = 0. \end{aligned} \quad (5.46)$$

Using the explicit expressions (5.43), (5.44) and the fact that  $C_{A, m_1y}^{m_2x} = 0$  due to Eqs. (5.39)–(5.40), one obtains the relation

$$\mu_1^m = \dot{\xi}_\perp^m - \partial_k (N^k \xi_\perp^m - N^m \xi_\perp^k), \quad (5.47)$$

written down in terms of transverse vectors  $\mu_1^m(x)$  and  $\xi_\perp^m(x)$  which are in one-to-one correspondence with  $\mu_1^A$  and  $\mu_2^A$

$$\begin{aligned} \mu_1^m(x) &= e_A^m(x) \mu_1^A, & \xi_\perp^m(x) &= e_A^m(x) \mu_2^A, \\ \mu_1^A &= \int d^3x e_m^A(x) \mu_{\perp 1}^m(x), & \mu_2^A &= \int d^3x e_m^A(x) \xi_\perp^m(x). \end{aligned}$$

Therefore, the canonical generator  $G_\perp = P_A \mu_1^A + H_A \mu_2^A$ , with  $\mu_1^A$  and  $\mu_2^A$  expressed via (5.47) in terms of an arbitrary transverse vector  $\xi_\perp^i$  can be rewritten in the reducible basis of transverse vectors as

$$G_\perp = \int d^3x [(\dot{\xi}_\perp^i - N^k \partial_k \xi_\perp^i + \xi_\perp^k \partial_k N^i) P_i + \xi_\perp^i H_i]. \quad (5.48)$$

Obviously this is just the generator (5.27) with the vector  $\mu_1^i$  given by (5.25), derived above for the case of  $T = 0$  branch, which is responsible for transverse spatial diffeomorphisms of both  $\gamma_{ij}$  and  $N^i$ .

But in contrast to the  $T = 0$  branch, this is the only set of two (per space point) local gauge transformations that exist in the  $S_i = 0$ ,  $T \neq 0$  branch of the theory. The number of local gauge transformations coincides with the number of *primary* first class constraints, which is less than four—the full number of first class constraints. The Dirac conjecture turns out to be true again, even though the set of assumptions listed in [14] as sufficient conditions for the validity of this conjecture are not satisfied. As one can see, this conclusion based on the canonical formalism of the model fully matches with the Lagrangian formalism accounting for the subtleties of gauge transformations at timelike and spacelike boundaries of spacetime, considered above in Sec. II. Below we reveal the detailed mechanism of this gauge invariance (and its violation) in the approximation of linearized theory on spatially closed and asymptotically flat Friedmann background.

## VI. LINEARIZED THEORY

We consider now the GUMG model in the linearized approximation on the background of the homogeneous Friedmann metric of positive or zero spatial curvature,  $k = +1$  or  $k = 0$ , respectively. In these cases we have the metric and curvature in terms of the scale factor  $a(t)$  and correspondingly the metric  $\sigma_{ij}$  of the 3-dimensional sphere of unit radius or the flat metric,

$$\gamma_{ij} = a^2(t) \sigma_{ij}(x), \quad N = N(a), \quad N^i = 0, \quad (6.1)$$

$${}^3R = \frac{6}{a^2} k, \quad K_{ij} = -\frac{a\dot{a}}{N} \sigma_{ij}, \quad H = \frac{\dot{a}}{a}. \quad (6.2)$$

$H$  denotes the Hubble factor of the Friedmann background—we hope that it will not be confused with the notation for the GR Hamiltonian and momentum constraints. Equations of motion for this background read

$$\frac{\delta S}{\delta \gamma_{ij}} = \sqrt{\sigma} \frac{a^3}{N} \left[ 2\dot{H} + 3(1-w)H^2 + (1+3w) \frac{N^2}{a^2} k \right] \sigma^{ij} = 0, \quad (6.3)$$

$$\frac{\delta S}{\delta N^i} \equiv 0, \quad (6.4)$$

where in view of homogeneity the shift component is identically satisfied. One can check that the first equation has the integral of motion with a constant  $C$

$$\frac{H^2}{N^2} + \frac{k}{a^2} = \frac{C}{3Na^3}, \quad (6.5)$$

which can be interpreted as the Friedmann equation with the physical Hubble factor in cosmic time,  $H/N = \dot{a}/Na$ , and the dark energy density  $\varepsilon = 2C/Na^3$  (the right-hand side being  $8\pi G\varepsilon/3$  in our units with  $G = 1/16\pi$ ). The nonzero constant  $C$  is what distinguishes the  $T \neq 0$  branch of the model from its general relativistic branch and simulates this dark energy density which identically satisfies the stress tensor conservation law

$$\frac{d\varepsilon}{da} = -3(1+w) \frac{\varepsilon}{a}. \quad (6.6)$$

where in addition to the rule of raising and lowering indices by the metric  $\sigma_{ij}$  all covariant derivatives are also defined with respect to this metric. One can check that all the terms except the last one coincide with the quadratic part of the Einstein action on the Friedmann background [18]. The last term is a modification due to GUMG generalization, which obviously vanishes for  $C = 0$  (GR branch of the model) and for  $w = -1$  when  $\Omega = 0$ —see Eq. (3.12). In this case this is just a unimodular gravity model.

Below we will discard the total derivative term in (6.10) and build the canonical formalism and the physical sector for the quadratic action  $S_{(2)}$ . Hamiltonian formalisms for (6.10) and (6.11) differ, of course, by a canonical transformation, but for  $S_{(2)}$  the formalism is simpler. Another remark concerns spatial surface terms in (6.11). The spatial

### A. Action of linearized theory and its gauge invariance properties

Now we expand the GUMG action to second order in perturbations of  $\gamma_{ij}$  and  $N^i$ , which we denote as follows

$$\delta \gamma_{ij} = a^2 s_{ij}, \quad \delta N^i = S^i, \quad (6.7)$$

along with the first and second order variations of  $N(\gamma)$ —the function of the perturbed 3-metric,

$$\delta N = NA, \quad \delta^2 N = \frac{1}{4} w N [(\Omega - 1)s^2 - 2s_{ij}^2], \quad (6.8)$$

Here in view of  $\delta N = wNs/2$

$$A = \frac{1}{2} ws, \quad s = \sigma^{ij} s_{ij}, \quad s_{ij}^2 \equiv \sigma^{im} \sigma^{jn} s_{ij} s_{mn}, \quad (6.9)$$

$\Omega$  is defined by Eq. (3.12), and everywhere here and in what follows all spatial indices are raised and lowered by the metric  $\sigma_{ij}$ . As a result the second order variation of the action reads

$$\frac{1}{2} \delta^2 S_{\text{GUMG}} = S_{(2)} + \int dt d^3x \frac{d}{dt} \left\{ \sqrt{\sigma} \frac{a^3 H}{N} \left( s_{ij} - \frac{1}{2} s^2 \right) \right\}, \quad (6.10)$$

with the first term given by

$$\begin{aligned} S_{(2)} = & \int dt d^3x \sqrt{\sigma} a^3 \left\{ \frac{1}{N} \left( \nabla_{(i} S_{j)} - \frac{\dot{s}_{ij}}{2} \right)^2 - \frac{1}{N} \left( \nabla_k S^k - \frac{\dot{s}}{2} \right)^2 - \frac{4H}{N} \left( \nabla_k S^k - \frac{\dot{s}}{2} \right) A \right. \\ & - \frac{6}{N} H^2 A^2 + \frac{1}{4} \frac{N}{a^2} (2\nabla^k s^{ij} \nabla_i s_{jk} - \nabla^k s^{ij} \nabla_k s_{ij} - 2\nabla_i s^{ij} \nabla_j s + \nabla^i s \nabla_i s) \\ & \left. + \frac{N}{a^2} k \left( s_{ij}^2 - \frac{1}{2} s^2 \right) - 2 \frac{N}{a^2} k s A - \frac{N}{a^2} (\nabla^j s_{ij} - \nabla^i s) \nabla_i A + \frac{1}{4a^3} C w \Omega s^2 \right\}, \quad (6.11) \end{aligned}$$

derivatives here are organized so that the variational principle is consistent under a fixed metric of the timelike boundary (only squares of first order derivatives and no second order derivatives are contained in  $S_{(2)}$ ). On the other hand, for spatially closed model ( $k = 1$ ) this problem is irrelevant and the same applies to asymptotically flat space, because the falloff condition for metric perturbations makes all quadratic surface terms arising from integrations by parts vanishing at infinity.

As usual in the linearized theory on the field background, the gauge transformations of perturbation fields are defined as the part of the transformations (2.11)–(2.13) of zeroth order in perturbations (their cross terms with  $\xi^\mu$  belong to the next order of perturbation theory). Therefore, these transformations read

$$\delta^\xi s_{ij} = 2H\sigma_{ij}\xi^0 + \nabla_i\xi_j + \nabla_j\xi_i, \quad \xi_i = \sigma_{ij}\xi^j, \quad (6.12)$$

$$\delta^\xi S^k = -\frac{N^2}{a^2}\nabla^k\xi^0 + \dot{\xi}^k, \quad \nabla^i = \sigma^{ij}\nabla_j, \quad (6.13)$$

and they of course reproduce for the vectors  $\xi^\mu$  satisfying the ( $\sigma$ -metric covariantized) Eq. (2.14) the general relativistic transformation law for  $A = ws/2$ ,  $\delta^\xi A = \partial_t(N\xi^0)/N$ . Under these transformations the quadratic action transforms as

$$\begin{aligned} \delta^\xi S_{(2)} = & \int dt \int_{\partial\Sigma} d^2\Sigma_i \sqrt{\sigma} Na [2(S^i(\Delta + 2k) - S_j \nabla^i \nabla^j) \xi^0 + (\dot{s}^{ij} \nabla_j - \dot{s} \nabla^i) \xi^0 \\ & + 4HA \nabla^i \xi^0 + s^{jk} \nabla^i \nabla_j \xi_k - s^{ij} \Delta \xi_j - 4kA \xi^i - \frac{2C}{Na} (1+w) S^i \xi^0] \\ & + \int d^3x \sqrt{\sigma} Na \left[ (-\nabla_j s^{ij} \nabla_i + \nabla^i s \nabla_i - 2ks) \xi^0 + \frac{C}{Na} (1+w) s \xi^0 \right] \Big|_{t_-}^{t_+} \\ & + 2C \int dt d^3x \sqrt{\sigma} \left[ \frac{d \ln w}{d \ln \gamma} s \xi^0 + (1+w) S^i \nabla_i \xi^0 \right], \end{aligned} \quad (6.14)$$

where we carefully took into account all surface terms originating from integration by parts.

As is expected this expression for general relativistic case of  $C = 0$  reduces to the surface integral and vanishes for  $\xi^\mu$  with a compact support inside the spacetime domain, which is the case of the theory invariant under local gauge transformations. For GUMG theory we have additional terms including spacetime volume integrals. This is strange because the gauge transformations of GUMG theory are just a subclass of gauge transformations of general relativity, so that the difference in the gauge transformation of the action can only be of the boundary terms type. It turns out indeed that this extra spacetime integral reduces to the surface integral at timelike and spacelike boundaries.

To see this perturb the exact equation (1.1) for  $\xi^\mu$  and obtain the relation

$$w \nabla_i \xi^i_{(1)} = - \left[ \frac{d \ln w}{d \ln \gamma} s \dot{\xi}^0_{(0)} + (1+w) S^i \nabla_i \xi^0_{(0)} \right] + \dot{\xi}^0_{(1)} \quad (6.15)$$

where  $\xi^i_{(1)} = \delta \xi^i$  and  $\xi^0_{(1)} = \delta \xi^0$  are the first order perturbations of  $\xi^\mu$  under the variations of the metric and shift functions on the Friedmann background (6.1)–(6.2), and we took into account that the background value of  $w$  is independent of space coordinates. Therefore, the last spacetime integral term in (6.14) also becomes a surface term

$$\begin{aligned} & 2C \int dt d^3x \sqrt{\sigma} \left[ \frac{d \ln w}{d \ln \gamma} s \dot{\xi}^0 + (1+w) S^i \nabla_i \xi^0 \right] \\ & = 2C \int_{\Sigma} d^3x \sqrt{\sigma} \xi^0_{(1)} \Big|_{t_-}^{t_+} - 2C \int dt w \int_{\partial\Sigma} d^2\Sigma_i \xi^i_{(1)}. \end{aligned} \quad (6.16)$$

As we show now, this relation implies that for both closed compact and open asymptotically Friedmann models the action is not invariant under  $\xi^0$ -transformation.

Indeed, for closed models timelike surface integrals are absent, but  $\xi^0$  cannot be a field independent variable with a compact support, because it should satisfy the Eq. (2.14) nontrivially depending on the fields. When expanded in powers of metric perturbations,  $\xi^\mu = \sum_n \xi^\mu_{(n)}$ , this equation implies that the spatially constant mode of the lowest order  $\xi^0_{(0)}$  is time independent,  $\int d^3x \sqrt{\sigma} \dot{\xi}^0_{(0)} = 0$ , other harmonics of  $\xi^0_{(0)}$  being arbitrary and admitting a compact support in time. On the contrary, the same mode of  $\xi^0_{(1)}$  is strongly restricted by the differential in time equation

$$\int d^3x \sqrt{\sigma} \left[ \dot{\xi}^0_{(1)} - \frac{d \ln w}{d \ln \gamma} s \dot{\xi}^0_{(0)} - (1+w) S^k \nabla_k \xi^0_{(0)} \right] = 0, \quad (6.17)$$

so that  $\xi^0_{(1)}$  cannot have a compact support. In particular,  $\xi^0_{(1)}(t_\pm) \neq 0$  and the gauge transformation (6.14) is nonzero in view of the first term of Eq. (6.16). Other terms are vanishing because they already contain one power of perturbation fields, so that  $\xi^0$  factor should include only the leading order part  $\xi^0_{(0)}$  which has a compact support in time. Thus, the closed model action is invariant only under the transverse spatial diffeomorphisms with the vectors  $\xi^\mu = (0, \xi^i_\perp)$  having a compact support in time. These are the only gauge transformations generated by the canonical first class constraints in the  $T \sim C \neq 0$  branch of the theory.

In the open models case with asymptotically flat space slices,  $\xi^0$  can be taken field independent, because the integral of Eq. (6.15) is no longer a restriction on the choice of  $\xi^0$ , but just the relation which determines the flux of the vector  $\xi^i_{(1)}$  through the remote spatial boundary,  $\int_\infty d^2\Sigma_i \xi^i_{(1)}$ . Therefore,  $\xi^0$  can have a compact support in time, and all spacelike integrals at  $t_\pm$  in (6.14) and (6.16)

vanish. However,  $\xi_{(1)}^i(x) \sim 1/|x|^2$ ,  $|x| \rightarrow \infty$ , and this “side” surface flux of the form (2.24), generated by the last term of (6.16), is nonvanishing. Note that this is also the only term that breaks gauge invariance of the action, because other “side” surface integrals are zero in view of the falloff conditions for  $s_{ij}(x)$ ,  $S^i(x)$ ,  $\xi^\mu(x)$ .<sup>6</sup>

Thus, in both closed and open cases the  $\xi^0$ -transformation of the action is given by the nonvanishing 3-dimensional integral over the full boundary of spacetime (6.16),

$$\delta^\xi S_{(2)} = 2C \int_{\partial^{(4M)}} d^3 \Sigma_\mu \Xi_{(1)}^\mu, \quad (6.18)$$

$${}^4M = [t_-, t_+] \times {}^3\Sigma, \quad \Xi_{(1)}^\mu = (\xi_{(1)}^0, -w\xi_{(1)}^i),$$

with a zero timelike or “side” part in a closed model case and a zero spacelike part,  $\Xi_{(1)}^0 = 0$ , for an asymptotically flat model. The GUMG action is gauge invariant only under two spatial diffeomorphisms with a transverse 3-vector having a compact support both in space and time. All this is in full accordance with the canonical formalism of the  $T \sim C \neq 0$  branch, considered above. Note that UMG theory with  $w = -1$  is an exception from this rule even for  $C \neq 0$ , because the Eq. (1.1) in this case is field independent, and  $\xi_{(1)}^\mu = 0$ .

## B. Physical sector

The number of degrees of freedom in the physical sector is determined by the number of first and second class constraints and, therefore, is different in different branches of the model. In the  $T \neq 0$  branch their counting gives  $18 - 2 \times 4 - 4 = 2 \times 3$ —three local degrees of freedom, where 18 is the number of phase space variables  $N^i$ ,  $P_i$ ,  $\gamma_{ij}$ ,  $\pi^{ij}$ , 4 is the number of first class constraints  $P_A$ ,  $H_A$  and 4 is the number of second class constraints—two longitudinal components of  $P_i$  and  $H_i$  and two independent constraints among  $T_i$  and  $S_i$  (since  $T_i = \partial_i T$  and  $S_i = \partial_i S$  only one longitudinal component counts for each of them). In the  $T = 0$  branch of GUMG and in UMG, just like in GR, the counting is different and obviously gives two degrees of freedom,  $18 - 2 \times 7 = 2 \times 2$ , where 7 is the full set of first class constraints  $P_i$ ,  $H_i$ , and  $T$  [16].

In the linearized theory disentangling these degrees of freedom runs via the decomposition of the metric perturbations into the linear combination of transverse-traceless tensor modes  $t_{ij}$ , transverse vector modes  $F_i$ ,  $V_i$ , and the scalar modes  $\psi$ ,  $E$ ,  $B$ ,

<sup>6</sup>Integrability condition for terms without spatial derivatives of metric perturbations in (6.11) imply that  $s_{ij}(x) \sim 1/|x|^2$ ,  $S^i(x) \sim 1/|x|$ ,  $\xi^i(x) \sim 1/|x|$  and  $\xi^0(x) \sim 1/|x|^2$  [the latter two restrictions guarantee that the gauge transformation (2.13) with  $\dot{\gamma}_{ij} \sim \delta_{ij} \dot{a}/a$  should not violate the falloff condition for  $s_{ij}(x)$ ], which guarantees that all surface terms except the one in (6.16) vanish at spatial infinity  $|x| \rightarrow \infty$ .

$$s_{ij} = t_{ij} + 2\nabla_{(i} F_{j)} - 2\psi\sigma_{ij} + 2\nabla_i \nabla_j E, \quad (6.19)$$

$$t^i{}_i = \nabla^i t_{ij} = \nabla_i F^i = 0, \quad (6.20)$$

$$S_i = V_i + \nabla_i B, \quad \nabla_i V^i = 0. \quad (6.21)$$

Using this decomposition in (6.11) one finds that the tensor transverse-traceless modes decouple from the vector and scalar sectors,  $S_{(2)} = S_t + S_v + S_s$ , in the form of the action of two field-theoretical oscillators on the nonstatic Friedmann background

$$S_t = \int dt d^3x \sqrt{\sigma} a^3 \left[ \frac{\dot{t}_{ij}^2}{4N} - \frac{N}{4a^2} (\nabla_k t_{ij})^2 - \frac{kN}{2a^2} t_{ij}^2 \right]. \quad (6.22)$$

Transverse vector modes  $F_i$  and  $V_i$  enter only the action  $S_v$ , while the action of the scalar sector  $S_s$  equals

$$S_s = \int dt d^3x \sqrt{\sigma} a^3 \left[ -\frac{6}{N} (\dot{\psi} + HA)^2 - \frac{4}{N} (\dot{\psi} + HA) \Delta(B - \dot{E}) + \frac{2k}{N} (B - \dot{E}) \Delta(B - \dot{E}) - \frac{2N}{a^2} \psi (\Delta + 3k) \psi + \frac{4N}{a^2} A (\Delta + 3k) \psi \right] + C \int dt d^3x \sqrt{\sigma} \frac{\Omega}{w} A^2, \quad (6.23)$$

where the perturbation of the lapse function  $A$ , cf., Eq. (6.7), reads in terms of the scalar modes of the metric perturbations as

$$A = w(\Delta E - 3\psi). \quad (6.24)$$

Obviously, it is this sector that contains in the case of the  $T \neq 0$  GUMG branch the third physical d.o.f. To disentangle it one has to build the canonical formalism of the linearized theory and solve the relevant constraints and canonical gauge conditions in a conventional Hamiltonian reduction procedure. Four gauge conditions, necessary to gauge out the symmetries generated by the four first class constraints  $P_A$  and  $H_A$  at the linearized level are used for the determination of the four transverse vector modes ( $F_i$ ,  $V_i$ ). But their quadratic action  $S_v$  vanishes on the linearized constraints, and their choice dynamically affect neither the transverse-traceless, nor the scalar sector of the theory. Therefore, we do not specify vector modes at all and consider only the scalar sector in what follows. The reduction of this scalar sector to physical modes follows only from solving its canonical first and second class constraints—the linearized version of the constraints of the full nonlinear theory considered above.

As one can see now, this reduction goes very differently in the spatially homogeneous and inhomogeneous subspaces of the model. To begin with, for spatially constant modes  $\psi_0$ ,  $E_0$ , and  $B_0$  the quadratic action (6.23) reads as



$$S_s^{(0)} = \int dt d^3x \sqrt{\sigma} a^3 \left\{ -\frac{6}{N} (\dot{\psi}_0 - 3wH\psi_0)^2 + \left[ \frac{9C}{a^3} w\Omega - \frac{6N}{a^2} k(1+6w) \right] (\psi_0)^2 \right\}, \quad (6.25)$$

and does not generate any constraints. This is the action of one global physical mode which is in fact the first order perturbation of the background solution of Eq. (6.3). This can be easily verified by noting that  $S_s^{(0)}$  coincides with the second order variation of the minisuperspace action

$$S_s^{(0)} = \frac{1}{2} \delta^2 \int dt d^3x \sqrt{\sigma} \left[ 6kN(a)a - \frac{6a^3}{N(a)} \dot{a}^2 \right] \quad (6.26)$$

under the variations of the scale factor  $\delta a = a\psi_0$  and the  $a$ -dependent lapse function  $\delta N(a) = 3wN(a)\psi_0$ . This action generates the second order equation of motion for  $\psi_0$  which has a runaway solution corresponding to either the cosmological acceleration or inflation, if it is applied in context of inflationary cosmology. This single mechanical (rather than field-theoretical) mode has a ghost nature because of the negative sign of its kinetic term and does not differ much from the scale factor mode in GR, except that  $\psi_0$  in GUMG is dynamically independent and its non-vanishing constant of motion  $C$  is freely specified by initial conditions.

For spatially inhomogeneous modes with  $\Delta \neq 0$  the situation is different—this is the case of the constrained system. The definition of the canonical momenta,

$$\Pi_\psi = -4\sqrt{\sigma} \frac{a^3}{N} [3(\dot{\psi} + HA) + \Delta(B - \dot{E})], \quad (6.27)$$

$$\Pi_E = 4\sqrt{\sigma} \frac{a^3}{N} [\Delta(\dot{\psi} + HA) - k\Delta(B - \dot{E})], \quad (6.28)$$

$$\Pi_B = 0, \quad (6.29)$$

implies one primary constraint. On the subspace of invertible operator  $\Delta + 3k$ , where the equations

$$(\Delta + 3k)(\dot{\psi} + HA) = \frac{N}{4a^3} \frac{1}{\sqrt{\sigma}} (\Pi_E - k\Pi_\psi), \quad (6.30)$$

$$(\Delta + 3k)\Delta(B - \dot{E}) = -\frac{N}{4a^3} \frac{1}{\sqrt{\sigma}} (\Delta\Pi_\psi + 3\Pi_E) \quad (6.31)$$

can be uniquely solved for  $\dot{\psi}$  and  $\dot{E}$ , the Legendre transform with respect to these velocities gives the total canonical action which includes the Hamiltonian  $\mathcal{H}$  and this primary constraint with the Lagrange multiplier  $u$ ,

$$S_s^{(>0)} = \int dt \left\{ \int d^3x (\Pi_\psi \dot{\psi} + \Pi_E \dot{E} + \Pi_B \dot{B}) - \mathcal{H} - \int d^3x u \Pi_B \right\}, \quad (6.32)$$

$$\begin{aligned} \mathcal{H} = & \frac{N}{8a^3} \int \frac{d^3x}{\sqrt{\sigma}} \left\{ -\Pi_\psi \frac{k}{\Delta + 3k} \Pi_\psi + 2\Pi_\psi \frac{1}{\Delta + 3k} \Pi_E \right. \\ & \left. + 3\Pi_E \frac{1}{\Delta + 3k} \frac{1}{\Delta} \Pi_E \right\} \\ & + \int d^3x \left\{ 2Na\sqrt{\sigma}\psi(\Delta + 3k) - A[H\Pi_\psi \right. \\ & \left. + 4\sqrt{\sigma}Na(\Delta + 3k)\psi] - C\sqrt{\sigma}\frac{\Omega}{w}A^2 + B\Pi_E \right\}. \quad (6.33) \end{aligned}$$

The conservation of the primary constraint (6.29) leads to the sequence of secondary, tertiary, and quaternary constraints

$$\{\Pi_B, \mathcal{H}\} = -\Pi_E = 0, \quad (6.34)$$

$$\{\Pi_E, \mathcal{H}\} = T = 0, \quad (6.35)$$

$$T \equiv \sqrt{\sigma} \Delta \left( 2C\Omega A + w \left[ H \frac{\Pi_\psi}{\sqrt{\sigma}} + 4Na(\Delta + 3k)\psi \right] \right), \quad (6.36)$$

$$\begin{aligned} \frac{\partial T}{\partial t} + \{T, \mathcal{H}\} = & Cw\sqrt{\sigma}\Delta \left( 12H \frac{d\Omega}{d \ln \gamma} \frac{A}{w} \right. \\ & \left. + \frac{N}{a^3} \frac{d \ln w}{d \ln \gamma} \frac{\Pi_\psi}{\sqrt{\sigma}} + 2\Omega\Delta B \right) = 0, \quad (6.37) \end{aligned}$$

where the explicit time derivative  $\partial/\partial t$  acts only on the background variables in  $T$ . The conservation of the last constraint leads to the equation on the Lagrange multiplier  $u$ , so that the sequence of these constraints terminates at the quaternary one.<sup>7</sup>

There is an additional subtlety for a spatially closed model,  $k = +1$ , when the operator  $\Delta + 3$  has discrete zero modes, and the nonlocal terms in the Hamiltonian (6.33) seem to be ill defined. This is the set of four eigenmodes of the Laplacian operator on the 3-sphere,  $\Delta Z_n(x) = -n(n+2)Z_n(x)$ , labeled by  $n = 1$ , which have the property  $\nabla_i \nabla_j Z_1(x) = -\sigma_{ij} Z_1(x)$ . This mode, however, does not contribute to the physical sector of the theory. To see this, note that in the subspace of this mode the left-hand

<sup>7</sup>These constraints are equivalent to but not directly coincide with the longitudinal parts of linearized constraints  $H_i, T_i,$  and  $S_i$  of the full nonlinear theory considered above. This is because the canonical formalisms of the full theory and its linearized version are related by the canonical transformation corresponding to the total derivative term in (6.10)—see discussion following Eq. (6.11).

sides of Eqs. (6.30)–(6.31) are vanishing, which means that we have an extra primary first class constraint

$$\Pi_E^1 - \Pi_\psi^1 = 0. \quad (6.38)$$

This constraint generates the gauge transformation  $E_1 \rightarrow E_1 + \zeta$ ,  $\psi_1 \rightarrow \psi_1 - \zeta$ , which leaves the linearized 3-metric (6.19) invariant because its scalar part equals  $-2\psi_1\sigma_{ij} + 2\nabla_i\nabla_j E_1 = -2\sigma_{ij}(\psi + E)_1$ . Indistinguishability of these two terms in the decomposition of  $s_{ij}$  into irreducible components requires additional gauge condition to fix  $\psi_1$  and  $E_1$  separately, while the physical sector depends only on their sum. The dynamical effect of this extra constraint is that the kinetic term of the  $n = 1$  mode completely vanishes in the Hamiltonian, so that ill-defined nonlocal terms disappear. Moreover, in view of the secondary constraint  $\Pi_E^1 = 0$  both momenta in the  $n = 1$  sector vanish, and the tertiary and quaternary constraints completely kill its physical mode, because both  $(\psi + E)_1$  and  $B_1$  turn out to be zero due to (6.36) and (6.37).<sup>8</sup>

Thus, for spatially inhomogeneous modes with  $\Delta \neq 0$  the tertiary and quaternary constraints (6.36)–(6.37) can be algebraically solved for  $A$  and  $B$  in terms of  $\psi$  and  $\Pi_\psi$  and used in (6.32) along with primary and secondary constraints. This gives the sector of one local scalar d.o.f. with the physical Hamiltonian  $\mathcal{H}_*[\psi, \Pi_\psi]$

$$S_s^{(>0)}[\psi, \Pi_\psi] = \int dt \left\{ \int d^3x \Pi_\psi \dot{\psi} - \mathcal{H}_*[\psi, \Pi_\psi] \right\}, \quad (6.39)$$

$$\begin{aligned} \mathcal{H}_*[\psi, \Pi_\psi] = \int d^3x \left[ \frac{1}{2} \Pi_\psi G(\Delta) \Pi_\psi + \Pi_\psi V(\Delta) \psi \right. \\ \left. + \frac{1}{2} \psi U(\Delta) \psi \right], \end{aligned} \quad (6.40)$$

where the coefficients  $G(\Delta)$ ,  $U(\Delta)$ , and  $V(\Delta)$  are the following (non)local operators—functions of the covariant Laplacian,<sup>9</sup>

$$G(\Delta) = \frac{1}{4\sqrt{\sigma}} \left[ \frac{2w}{C\Omega} H^2 - \frac{N}{a^3} \frac{k}{\Delta + 3k} \right], \quad (6.41)$$

<sup>8</sup>For the open model with  $k = 0$  the difficulty of the above type becomes the problem of the dipole mode which seems to invalidate the Hamiltonian (6.33) because of the infrared divergent integral  $\int d^3x \Pi_E (1/\Delta^2) \Pi_E = \int d^3x [(1/\Delta) \Pi_E]^2$  with  $(1/\Delta) \Pi_E(x) \sim 1/|x|$  at  $|x| \rightarrow \infty$ . But again in the physical sector, when the secondary constraint  $\Pi_E = 0$  is enforced, this problem is easily circumvented.

<sup>9</sup>Local degrees of freedom with a nonlocal Hamiltonian sounds as an euphemism, but this is a usual terminology accepted to distinguish field-theoretical modes associated with location in coordinate or dual momentum space from global mechanical ones.

$$V(\Delta) = \frac{2w}{C\Omega} NaH(\Delta + 3k), \quad (6.42)$$

$$U(\Delta) = 4Na\sqrt{\sigma}(\Delta + 3k) \left[ 1 + \frac{2wNa}{C\Omega} (\Delta + 3k) \right]. \quad (6.43)$$

The above equations for  $S_s^{(>0)}[\psi, \Pi_\psi]$  and  $\mathcal{H}_*[\psi, \Pi_\psi]$  apply only in the spatially inhomogeneous sector of the theory. In the closed cosmological model with the discrete set of modes on the 3-sphere,  $\psi(x) \rightarrow \psi_n$ , this means that the space integral should be replaced as  $\int d^3x \rightarrow \sum_{n=2}^{\infty}$ ,  $G(\Delta) \rightarrow G(-n(n+2))$ , etc. For spatially flat Friedmann background,  $k = 0$ , the spectrum of degrees of freedom is continuous and can be represented by Fourier modes,  $\psi(x) \rightarrow \hat{\psi}(p)$ ,  $\int d^3x \rightarrow \int d^3p$ ,  $G(\Delta) \rightarrow G(-p^2)$ , which are adjacent at  $p \rightarrow 0$  to the discrete homogeneous mode  $\psi_0$ .

### C. Perturbative (in)stability

A nontrivial effect of solving the constraints is a complicated form of the coefficient  $G(\Delta)$  in the kinetic term of (6.40)—in contrast to the ghost mode in the homogeneous sector, its sign strongly depends on the details of the model. The positivity of this coefficient guarantees absence of ghost modes and conversion of the action to the Lagrangian form with the canonical normalization of the physical mode,  $\psi \rightarrow \varphi$ . Exclusion of the canonical momentum via its equation of motion,  $\Pi_\psi = G^{-1}(\dot{\psi} - V\psi)$ , gives the Lagrangian action

$$\begin{aligned} S_s^{(>0)}[\psi] = \int dt d^3x \left[ \frac{1}{2} \dot{\psi} G^{-1} \dot{\psi} - \dot{\psi} G^{-1} V \psi \right. \\ \left. - \frac{1}{2} \psi (U - V G^{-1} V) \psi \right]. \end{aligned} \quad (6.44)$$

For positive  $G$  in terms of the canonically normalized field  $\varphi$ ,

$$\psi = \sigma^{1/4} \sqrt{G} \varphi, \quad (6.45)$$

it reads after several integrations by parts

$$\begin{aligned} S_s^{(>0)}[\psi] = \frac{1}{2} \int dt d^3x \sqrt{\sigma} \left\{ \dot{\varphi}^2 - \varphi \left[ GU - V^2 \right. \right. \\ \left. \left. + \frac{1}{2} \partial_t \left( \frac{\dot{G}}{G} \right) - \frac{1}{4} \left( \frac{\dot{G}}{G} \right)^2 + V \partial_t \ln \frac{G}{V} \right] \varphi \right\}. \end{aligned} \quad (6.46)$$

A remarkable property of the coefficients (6.41)–(6.43) is that the combination  $GU - V^2$  is linear in  $\Delta$ , so that nonpolynomial in spatial derivatives (nonlocal) nature of the action takes place only for a closed model and is significant only for long wavelengths modes with a small  $\Delta$ ,

$$G = \frac{wH^2}{2\sqrt{\sigma}C\Omega} + O\left(\frac{k}{\Delta}\right). \quad (6.47)$$

In this limit the potential part of the action represents the sum of gradient squared and the mass term with the effective mass parameter  $m_{\text{eff}}^2 = m_{\text{eff}}^2(a)$  which explicitly depends on  $a$ ,

$$\begin{aligned} GU - V^2 + \frac{1}{2}\partial_t\left(\frac{\dot{G}}{G}\right) - \frac{1}{4}\left(\frac{\dot{G}}{G}\right)^2 + V\partial_t\ln\frac{G}{V} \\ = -c_s^2\frac{\Delta}{a^2}N^2 + m_{\text{eff}}^2N^2 + O\left(\frac{k}{\Delta}\right), \end{aligned} \quad (6.48)$$

and  $c_s$ —the speed of sound parameter defined by

$$c_s^2 = \frac{w(1+w)}{\Omega}. \quad (6.49)$$

Thus, the short wavelengths part of the action—the limit of big  $\Delta$ —takes the form

$$S_s^{(>0)}[\psi] = \frac{1}{2}\int dt d^3x \sqrt{\sigma} \left\{ \dot{\varphi}^2 + N^2\varphi \left[ c_s^2\frac{\Delta}{a^2} - m_{\text{eff}}^2 \right] \varphi \right\}. \quad (6.50)$$

For spatially flat Friedmann model this representation is exact for the full range of spatial gradients  $\Delta = -p^2$ —the square of the comoving momentum.

Absence of ghost instability,  $G > 0$ , and gradient instability,  $c_s^2 > 0$ , implies in view of (6.47), (6.49), and the positivity of the constant  $C$  [cf., Eq. (6.5)] the following two inequalities

$$\frac{w}{\Omega} > 0, \quad 1 + w > 0. \quad (6.51)$$

Outside of this range generic GUMG theory is unstable either due to ghost or gradient instability of the infinite set of modes extending to UV limit. Note that the second inequality is consistent with conclusions that violation of null energy condition is associated with perturbative instabilities [19,20].

Of course, there is always a ghost spatially homogeneous mode which, as was mentioned above, is just the perturbation of the minisuperspace Friedmann background. This mode can hardly disprove the GUMG theory, because this type of instability, or presence of a runaway solution, is just the essence of the cosmological acceleration or inflation phenomenon. A wrong sign of the kinetic term of the homogeneous background mode does not make general relativity an unstable theory, which is usually explained by the fact that this mode is not dynamically independent. In GUMG theory this mode is the physical one, and its effect is measured by the magnitude of the constant  $C$  which is a part of freely specifiable initial conditions—just like the

cosmological constant as a constant of integration in UMG theory. Otherwise, it works dynamically exactly the same way as the scale factor mode in GR.

A possible objection to the stability of the GUMG model might be that simultaneous growth of negative energy background mode and the positive energy inhomogeneous modes is not prohibited by the conservation law for the total energy. But this objection is likely to be refuted by a simple argument that this is just a back reaction of cosmological perturbations on the Friedmann background, and the issue of this problem, or the magnitude of this effect, is far from being settled both at the classical and, even more so, at the quantum level [21,22]. Thus, GUMG theory can be added to the list of reasons [4,23] why in certain cases ghost modes are not harmful in physical models.

The linearized theory analysis above does not stand a smooth limit  $w = -1$  which corresponds to the UMG theory, because of the singularity caused by  $\Omega = 0$ . This is because the number of local degrees of freedom drops in this limit to two, and the theory still remains stable even though it falls out of the domain (6.51). Finally, below the phantom divide line  $w = -1$  the extra scalar mode again becomes dynamical, but it suffers gradient instability with a negative  $c_s^2$ .

We will not discuss here other sources of instability, like the tachyonic one associated with the sign of the effective mass term in (6.50). The criteria of stability of the theory on a nonstationary background are very sophisticated when the positivity of the Hamiltonian, which is not conserved, becomes parametrization dependent and no longer indicative of the consistency of the model [24]. In particular, competing contributions of time dependent comoving momentum and masslike terms in (6.50) are subject to cosmological perturbation theory which reveals the particle creation and formation of inflationary power spectra. These phenomena will be considered elsewhere [25], whereas below we will only briefly discuss the prospects of applying the GUMG model in context of dark energy and inflation theories.

## VII. DISCUSSION AND CONCLUSIONS

Generalized unimodular gravity theory turned out to have a very rich dynamical structure. Its canonical formalism incorporates four generations of constraints of both the first and the second class, the first class ones being nontrivial linear combinations of the constraints belonging to different (primary and secondary) generations. Moreover, the number of primary first class constraints seemingly contradicts the originally claimed local symmetries of the Lagrangian action of GUMG model [1], which would indicate the breakdown of the Dirac conjecture in the theory of constrained dynamics. It turned out, however, that this conjecture still applies in GUMG model, because one of the originally assumed Lagrangian symmetries is actually either nonlocal in time or violates

boundary conditions at spatial infinity. The constrained Hamiltonian formalism clearly reveals this peculiarity, which is deeply hidden in the Lagrangian framework, and allows one to recover true local symmetries as canonical transformations acting in the phase space of the theory.

Another peculiarity of this model is the so-called bifurcation of the system of constraints, which means that the theory has two branches with different numbers of constraints belonging to different classes. One branch is characterized by the set of first class constraints and can be interpreted as general relativity within a partial gauge fixation of spacetime diffeomorphisms, corresponding to the kinematical restriction on metric coefficients in GUMG model—the lapse function as a rather generic function of the 3-metric determinant. The second, physically most interesting branch, is the one in which this restriction gives rise to the effective perfect fluid originally suggested as a candidate for dark energy. This dark fluid has a barotropic equation of state  $p = w\varepsilon$  with a variable parameter  $w = w(a)$  depending on the cosmological scale factor [1]. This branch has only two local diffeomorphism symmetries realized as canonical transformations on phase space—this enlarges the physical sector of the theory from two general relativistic degrees of freedom to three degrees of freedom. Perturbative analyses on the closed and spatially flat Friedmann background shows that this extra degree of freedom—the scalar graviton—can be free of ghost and gradient instabilities in a wide class of GUMG models satisfying the restrictions (6.51). The second branch also includes a well-known unimodular gravity model corresponding to a constant value  $w = -1$  which belongs to the boundary of this stability domain, but the theory is still stable, because the scalar mode is not dynamical one—all the constraints of UMG model are the first class ones and they rule this mode out of the physical sector.

Even though GUMG theory was originally suggested as a model for cosmological acceleration, by and large it fails to accommodate the dark energy phenomenology. Cosmological data suggests [5,6] that the effective parameter  $w$  could be below  $-1$  at small  $z$ , which contradicts (6.51). One of the possibilities to relax this stability criterion could be an attempt to construct a GUMG model in which the unstable scalar mode  $\psi$  is not physical—the method analogous to the elimination of ghost modes by canonical constraints in [26]. Quite interestingly, it is possible.

If one chooses the function  $w(a)$  in such a way that  $\Omega = 0$ , the sequence of canonical constraints (6.34)–(6.37) does not terminate at the fourth generation, because the conservation of (6.37) is no longer an equation on the Lagrange multiplier. In addition, two more constraints are generated,  $\Delta(\Delta + 3k)A = 0$  and  $\Delta B = 0$ . As a result all six scalar variables ( $\psi, \Pi_\psi, E, \Pi_E, B, \Pi_B$ ) are ruled out by the six constraints, and only harmless transverse-traceless

tensor modes survive in the inhomogeneous sector of the model. In view of the expression (3.12) for  $\Omega$  its zero value implies the following dependence of  $w$  and  $N$  on the scale factor

$$w = \frac{v}{a^3 - v}, \quad N = n \frac{v - a^3}{va^3}, \quad (7.1)$$

where  $v$  and  $n$  are some integration constants. The interesting case of  $w < -1$  implies that  $v > a^3$ . In the course of cosmological expansion when  $a$  grows to  $a_0 = v^{1/3}$  the barotropic parameter  $w \rightarrow -\infty$  and  $N \rightarrow 0$ , so that  $\dot{a} \rightarrow 0$  as it follows from the modified Friedmann equation (6.5). The scale factor reaches maximum with the infinite *physical* Hubble parameter  $\dot{a}/Na \sim (v - a^3)^{-1/2}$  and infinite acceleration in the cosmic (proper) time,  $d^2a/d\tau^2 = d^2a/(Ndt)^2 \rightarrow \infty$ . This is the analogue of the big rip singularity [27] beyond which the physical evolution cannot be analytically continued. For a negative  $v$  the parameter  $w$  stays above its value  $-1$  and smoothly grows to zero in the course of expansion, which is again different from the dark energy scenario, because it does not cross the phantom divide line  $w = -1$  and resembles more the exit from inflation picture. Altogether, including this case of nondynamical scalar graviton, it seems that the GUMG model is more interesting as a new candidate for inflation scenario, rather than the source of dark energy.

The GUMG model of inflation could be rather interesting, because the role of the inflaton is played by the scalar sector of the metric field. The situation is similar to the  $R^2$ -inflation in the Starobinsky model [11], when the role of inflaton is played by the dynamical conformal mode of the metric. Here the same is realized without higher order derivatives of the metric—entirely due to kinematical restriction on the lapse function of the theory. In view of the expression (3.12) for  $\Omega$ , stability conditions (6.51) for the scalar graviton at negative  $w$  imply that

$$\frac{dw}{d \ln a} > -3w(1 + w) > 0, \quad (7.2)$$

which is consistent with the inflation scenario, because  $w$  grows to zero at the exit from inflation, just like in the zero  $\Omega$  case above.

Moreover, the scalar graviton  $\psi$  with the action (6.50) has a dispersion equation with a nontrivial speed of sound (6.49) determined by the time-dependent equation of state of the effective perfect fluid  $p = w(a)\varepsilon$ . It is interesting that in the hydrodynamical formalism of inflation theories, incorporating also the class of k-essence models [28], the speed of sound (critically effecting the primordial power spectrum of perturbations) is determined by the effective equation of state of the inflaton field,  $c_s^2 = dp/d\varepsilon$ . In the GUMG model this expression equals

$$\frac{dp}{d\varepsilon} = w + \varepsilon \frac{dw/da}{d\varepsilon/da} = w - \frac{1}{3} \frac{1}{1+w} \frac{dw}{d \ln a}, \quad (7.3)$$

where we took into account the stress tensor conservation law (6.6) for  $\varepsilon$ . On the other hand, in view of the expression (3.12) for  $\Omega$ , the sound of speed of the GUMG scalar graviton (6.49) is the same up to quadratic order in  $dw/da$

$$c_s^2 = \frac{w}{1 + \frac{1}{3w(1+w)} \frac{dw}{d \ln a}} = \frac{dp}{d\varepsilon} + O((dw/da)^2). \quad (7.4)$$

For slowly varying functions  $w(a)$  these two parameters nearly coincide. For faster varying equations of state their discrepancy is not surprising, though, because the dynamics of the usual inflaton field additional to gravity is qualitatively different from the case when the inflaton belongs to the metric sector and acquires a dynamical nature due to nontrivial GUMG kinematics.

Thus, despite the fact that generalized unimodular gravity most likely represents a failed attempt to build a phenomenologically consistent model of dark energy, this theory can be regarded as a rather prospective source of inflation scenario, which is a subject of further studies [25].

## ACKNOWLEDGMENTS

We want to express special thanks to Alexander Kamenshchik with whom the idea of the generalized unimodular gravity model was originally put forward. Also we are indebted to Joseph Buchbinder, Mario Herrero-Valea, Nobu Ohta, Sergey Sibiryakov, Dima Sorokin, Philip Stamp, Gian Paolo Vacca, and Alex Vikman for stimulating discussions. One of the authors (A. B.) is grateful for hospitality of the Yukawa Institute for Theoretical Physics, Pacific Institute for Theoretical Physics and Green College of the University of British Columbia, where part of this work has been completed. This work was supported by the RFBR Grant No. 17-02-00651.

- 
- [1] A. O. Barvinsky and A. Yu. Kamenshchik, Darkness without dark matter and energy—generalized unimodular gravity, *Phys. Lett. B* **774**, 59 (2017).
- [2] J. J. van der Bij, H. van Dam, and Y. J. Ng, The exchange of massless spin two particles, *Physica (Amsterdam)* **116A**, 307 (1982); W. G. Unruh, A unimodular theory of canonical quantum gravity, *Phys. Rev. D* **40**, 1048 (1989); M. Henneaux and C. Teitelboim, The cosmological constant and general covariance, *Phys. Lett. B* **222**, 195 (1989); P. Jirouek and A. Vikman, New Weyl-invariant vector-tensor theory for the cosmological constant, *J. Cosmol. Astropart. Phys.* **04** (2019) 004.
- [3] R. R. Caldwell, A phantom menace? Cosmological consequences of a dark energy component with super-negative equation of state, *Phys. Lett. B* **545**, 23 (2002).
- [4] A. Vikman, Can dark energy evolve to the phantom?, *Phys. Rev. D* **71**, 023515 (2005).
- [5] P. Ade *et al.*, Planck 2015 results. XIII. Cosmological parameters, *Astron. Astrophys.* **594**, A13 (2016); Planck 2015 results. XIV. Dark energy and modified gravity, *Astron. Astrophys.* **594**, A14 (2016).
- [6] D. Wang, W. Zhang, and X.-H. Meng, Searching for the evidence of dynamical dark energy, *Eur. Phys. J. C* **79**, 211 (2019).
- [7] B. Ratra and P. G. E. Peebles, Cosmological consequences of a rolling homogeneous scalar field, *Phys. Rev. D* **37**, 3406 (1988).
- [8] A. Yu. Kamenshchik, U. Moschella, and V. Pasquier, An alternative to quintessence, *Phys. Lett. B* **511**, 265 (2001).
- [9] M. Shaposhnikov and D. Zehnhäusern, Scale invariance, unimodular gravity and dark energy, *Phys. Lett. B* **671**, 187 (2009).
- [10] D. Blas and S. Sibiryakov, Technically natural dark energy from Lorentz breaking, *J. Cosmol. Astropart. Phys.* **07** (2011) 026.
- [11] A. A. Starobinsky, A new type of isotropic cosmological models without singularity, *Phys. Lett.* **91B**, 99 (1980); Disappearing cosmological constant in f(R) gravity, *JETP Lett.* **86**, 157 (2007).
- [12] S. Deser and R. P. Woodard, Nonlocal Cosmology, *Phys. Rev. Lett.* **99**, 111301 (2007); Nonlocal cosmology II—Cosmic acceleration without fine tuning or dark energy, *J. Cosmol. Astropart. Phys.* **06** (2019) 034; A. O. Barvinsky, Serendipitous discoveries in nonlocal gravity theory, *Phys. Rev. D* **85**, 104018 (2012).
- [13] R. Bufalo and M. Oksanen, Canonical structure and extra mode of generalized unimodular gravity, *Phys. Rev. D* **97**, 044014 (2018).
- [14] M. Henneaux and C. Teitelboim, *Quantization of Gauge Systems* (Princeton University Press, Princeton, NJ, 1992).
- [15] R. L. Arnowitt, S. Deser, and C. W. Misner, The dynamics of general relativity, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), p. 227.
- [16] M. Henneaux, C. Teitelboim, and J. Zanelli, Gauge invariance and degree of freedom count, *Nucl. Phys.* **B332**, 169 (1990).
- [17] R. de León Ardón, N. Ohta, and R. Percacci, Path integral of unimodular gravity, *Phys. Rev. D* **97**, 026007 (2018).
- [18] J. Garriga, X. Montes, M. Sasaki, and T. Tanaka, Canonical quantization of cosmological perturbations in the one-bubble open universe, *Nucl. Phys.* **B513**, 343 (1998); Erratum, *Nucl. Phys.* **B551**, 511(E) (1999).
- [19] S. Dubovsky, T. Gregoire, A. Nicolis, and R. Rattazzi, Null energy condition and superluminal propagation, *J. High Energy Phys.* **03** (2006) 025.

- [20] I. Sawicki and A. Vikman, Hidden negative energies in strongly accelerated Universes, *Phys. Rev. D* **87**, 067301 (2013).
- [21] N. C. Tsamis and R. P. Woodard, Relaxing the cosmological constant, *Phys. Lett. B* **301**, 351 (1993); Quantum gravity slows inflation, *Nucl. Phys.* **B474**, 235 (1996).
- [22] J. Garriga and T. Tanaka, Can infrared gravitons screen Lambda?, *Phys. Rev. D* **77**, 024021 (2008).
- [23] A. Smilga, Classical and quantum dynamics of higher-derivative systems, *Int. J. Mod. Phys. A* **32**, 1730025 (2017).
- [24] A. Ganz, P. Karmakar, S. Matarrese, and D. Sorokin, Hamiltonian analysis of mimetic scalar gravity revisited, *Phys. Rev. D* **99**, 064009 (2019).
- [25] A. O. Barvinsky, A. Yu. Kamenshchik, and N. Kolganov (to be published).
- [26] D. Langlois and K. Noui, Hamiltonian analysis of higher derivative scalar-tensor theories, *J. Cosmol. Astropart. Phys.* **07** (2016) 016; K. Takahashi and T. Kobayashi, Extended mimetic gravity: Hamiltonian analysis and gradient instabilities, *J. Cosmol. Astropart. Phys.* **11** (2017) 038.
- [27] R. Caldwell, M. Kamionkowski, and N.N. Weinberg, Phantom Energy and Cosmic Doomsday, *Phys. Rev. Lett.* **91**, 071301 (2003).
- [28] V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, Cambridge, England, 2005).