Cosmological quantum fluctuations: Gauge-invariance and Gaussian states

Carsten Gundlach

Department of Applied Mathematics and Theoretical Physics,

University of Cambridge, Silver Street, Cambridge CB3 9EW, England

June 1992, revised 6 January 1993

Abstract

A quantum theory of cosmological perturbations is based on the action of Einstein gravity with matter expanded to quadratic order around a background solution of the field equations. If, in the Schroedinger picture, the wave function is assumed to factorize such that each factor is a function of only a finite number of degrees of freedom (e g Fourier components of the fields) all gauge constraints can be solved explicitly. This generalizes previous gauge-invariant quantizations while clarifying their limited scope. The result is cast into the more convenient Heisenberg picture. It can be found directly from the Lagrangian, the Hamiltonian approach serving only as a justification. The concept of a Fock vacuum is generalized to several linear but coupled degrees of freedom.

1. Introduction

The study of vacuum fluctuations in cosmology is motivated by the idea that the existence of structure such as galaxies may be a direct consequence of the uncertainty principle in conjunction with a period of inflation in the early universe.

In this paper I consider quantum perturbations of the matter and metric on a classical background solution, without a semiclassical backreaction. The background is assumed to be of minisuperspace type, which I define as a spacetime with matter fields possessing so much symmetry that there is a preferred coordinate system. In consequence it typically has a finite number of degrees of freedom, which I denote by $Q^A(t)$. The standard example is Friedmann-Robertson-Walker (FRW) space with a homogeneous scalar field, but by leaving the minisuperspace generic I avoid unnecessary technicalities. The quantum theory is based on canonical quantization of the action of gravity plus matter, expanded to quadratic order in perturbations $q^a(\vec{x}, t)$ without any symmetries around the background . The linear order vanishes as the background is assumed to obey the field equations. This purely quadratic action gives rise to a linear field theory.

Linear equations of motion can always be decoupled, and therefore the quadratic action can be written as

$$S = \int L \, d^3\vec{x} dt = \sum_n \int L_n(q_n, \dot{q}_n, t) \, dt = \sum_n S_n[t, q_n(t)], \tag{1}$$

where n is a discrete or continuous infinite label. $S_n = S_n(t)$ denotes the explicit time-dependence of the action arising from the classical background solution. The q^a have been decomposed as $q^a(\vec{x}, t) = \sum_n q^a_n(t) G_n(\vec{x}, t)$. On any homogeneous spacetime background the $G_n(\vec{x}, t)$ are the harmonics $G_n(\vec{x})$ with respect to the induced 3-metric on

t-hypersurfaces, because these are automatically orthogonal under square integration. On FRW spaces the $G_n(\vec{x})$ are known (Halliwell and Hawking 1985, D'Eath and Halliwell 1987). On flat FRW space they are just $\exp i\vec{n}\vec{x}$. The suffix n also includes a split into scalar, vector and tensor perturbations with respect to the background 3-metric, which arises naturally from the "Fourier" decomposition.

As in full, non-linearized gravity, some field components λ^i, those which do not lie within the t-hypersurfaces, have vanishing conjugate momenta and play the role of Lagrange multipliers in the action. The Hamiltonian is therefore of the form

$$H = \sum_n H_n(p_n, q_n, t) + \sum_{n,i} \lambda^i_n H^n_i(p_n, q_n, t). \tag{2}$$

H is quadratic in the p and q or λ. The H_n are therefore quadratic in the p_n and q_n. The H^n_i are linear in the p_n and q_n, and constrained to vanish by the λ^i_n. Note that, in contrast to full gravity, even the lapse constraint is linear, while the Hamiltonian has a nonvanishing (quadratic) part.

When we quantize in the Schroedinger picture, in the representation $p = -i\partial/\partial q$, the wave function $\chi(q,t)$ obeys the Schroedinger equation

$$\left(-i\frac{\partial}{\partial t} + \sum_n H_n\right)\chi(q,t) = 0 \tag{3}$$

and the linear constraints

$$H^n_i \chi = 0 \quad \text{for all} \quad i, n. \tag{4}$$

Equations (1-4) summarize our premises.

The principal remaining problem is now the gauge-invariance of (even linearized) gravity, i e the fact that it contains unphysical degrees of freedom which correspond only to (small) changes in the identification of points in the perturbed spacetime with coordinate points and hence points in the unperturbed background spacetime. In the Schroedinger picture these are in principle eliminated by the (linearized) lapse and shift constraints H_i^n. For a closed FRW universe with a single scalar field the constraints were given by Halliwell and Hawking (1985) and solved by Shirai and Wada (1988). That solution is complicated by the use of two different methods for solving the lapse and the shift constraints and is best reviewed in an appendix.

If we quantize the action (1) in the Heisenberg picture, all operators obey the classical evolution and constraint equations. The field operators can therefore be written as products of complex classical solutions, called the mode functions, and annihilation and creation operators:

$$q^a(\vec{x},t) = \sum_{n,r} G_n(\vec{x},t) q_n^{ar}(t) a_{nr} + G_n^*(\vec{x},t) q_n^{ar*}(t) a_{nr}^\dagger \tag{5}$$

where $[a_{nr}, a_{n'r'}^\dagger] = \delta_{rr'}\delta_{nn'}$. The problem is now that there cannot be as many different a_r as there are q^a for each n, or the constraints could not be satisfied. The question is then how the q^a are connected with the a_r. An approximate answer has been used in most of the literature on inflation up to the present day, following Starobinsky (1982), Hawking (1982), Guth and Pi (1982) and Bardeen et al (1983). It is to link the a_r one-to-one with the perturbations of the scalar fields (matter fields) present, and the metric perturbations implicitly to the a_r via their links to the matter fields. (See Salopek et al (1989) for a clear statement of this approach.) It is incorrect as the canonical momentum of a scalar field ϕ is its time derivative $\dot{\phi}$ **plus** metric perturbations.

The correct Heisenberg quantization was found by Mukhanov (1988) for the particular case of a scalar field on flat FRW space (See also Mukanov et al 1990). He found that when one substitutes the classical constraints $H_i^n = 0$ (in my notation) back into the action (1), it can be expressed in terms of a single degree of freedom, v. The action in v is just the action of a minimally coupled scalar field on Minkowski space with a time-dependent mass, making quantization obvious. It is not obvious that such a reduced action can be made the basis of a quantum theory. Furthermore the variable v was introduced ad hoc. In Mukhanov's example there is at least one other variable, u, in which the reduced action can also be expressed entirely (Deruelle, Gundlach and Polarski 1992), and its equation of motion is of the same form as that of v.

In my general treatment in section 2 I assume that the wave function in the Schroedinger picture factorizes in the q_n. Under this assumption I show that solving the lapse and shift constraints (and any other gauge constraints) does indeed correspond to reducing the phase space by means of a canonical transformation. The new Hamiltonian depends only on the true degrees of freedom. These can be uniquely characterized as linear combinations of the old degrees of freedom which are both linearly gauge-invariant in the classical sense of Bardeen (1980) and which contain only the spatial part of vector and tensor fields with respect to the time coordinate of the Schroedinger equation. In practice the true degrees of freedom can be defined and the reduced action be obtained directly from the action, after a choice of time variable has been made. The Hamiltonian framework is only needed at the formal level, as a justification.

Why a Hamiltonian/ Schroedinger picture approach at all? The solution of the gauge-problem lies in gauge-invariant variables, which arise naturally in section 2. But generalizations of Bardeen's (1980) perturbation variables make sense only around a background with a preferred coordinate system. This in turns makes Hamiltonian methods useful.

In section 3 I take the quantum theory of the true degrees of freedom over to the usual Heisenberg picture, and construct a class of states which are of interest in cosmology, and which can be characterized as the semiclassical, Gaussian, or Fock vacuum states. These are a generalization of the usual Fock vacua, which arise because there may be more than one true degree of freedom per n, for example if there are two or more scalar fields present. Section 4 contains my conclusions.

My notation is as follows: Q^A is a minisuperspace (background) coordinate, and P_A its conjugate momentum. q_n^a is a perturbation coordinate, where n denotes a spatial Fourier component or its generalization, and p_{an} is its conjugate momentum. λ_n^i is the perturbation of a Lagrange multiplier. An alternative set of coordinates to the $\{q_n^a\}$ is $\{y_n^a\} \equiv \{v_n^r, f_n^i\}$. Here the index a ranges over the same values as the indices r and i together, and i has the same range on λ_n^i and f_n^i. v and f denote the true and the pure gauge degrees of freedom, while y is the common name for both. The notation v^r is used in generalization of the v of Mukhanov (1988), and "vrai" and "faux" can serve as a mnemonic aid for v and f.

2. The true quantum degrees of freedom

We shall now assume that the wave function is of the form

$$\chi(q,t) = \prod_n \Psi_n(q_n,t). \tag{6}$$

This is assumed in most previous work, enforced by technical difficulty if not by other considerations. One should be aware that (6) is a genuine restriction on the quantum state $|\chi\rangle$. In particular,

$$\langle\chi|q_n^a q_m^b|\chi\rangle = \langle\Psi_n|q_n^a|\Psi_n\rangle\langle\Psi_m^b|q_m|\Psi_m\rangle = 0. \quad \text{for} \quad n \neq m. \tag{7}$$

(7) does not hold for the two-particle state $(a_n^{r\dagger} + a_m^{s\dagger})|0\rangle$, where $a_n^{r\dagger}$ and $a_m^{s\dagger}$ are the usual creation operators in the Heisenberg picture and $|0\rangle$ is the ground state (compare equation (5).) Therefore (6) excludes typical particle states. It does include the class of states which we construct in section 4. The assumption (6) simplifies the Schroedinger equation and constraints to

$$\left(-i\frac{\partial}{\partial t} + H_n(p_n,q_n,t)\right)\Psi_n(q_n,t) = 0, \tag{8}$$

$$H_i^n\Psi_n = 0 \quad \text{for all} \quad i \tag{9}$$

for each n. To summarize the task of this chapter, for each generalized Fourier component n of the perturbations (n includes the scalar-vector-tensor split) we are dealing with $a = 1...N+M$ quantized degrees of freedom. They are subject to a Schroedinger equation with a quadratic time-dependent Hamiltonian and $i = 1...M$ homogeneous linear time-dependent constraints.

Formulated like this, the task at hand is straightforward, brought to light from under an interpretation in terms of spacetime and gauge freedom. Part of the simplicity, namely the linearity in both q and p (or $-i\partial/\partial q$) derives of course from the fact that we have assumed a classical background.

Because of the finite number of degrees of freedom the quantization with constraints can be replaced by a classical solution of the constraints, leading to N true degrees of freedom, prior to quantization. The equivalence will be shown at the end of this section. Here and in the following section we are looking at one value of n at a time, and therefore we suppress the suffix n for clarity, except in H_n and H_i^n, which need to be distinguished from the total Hamiltonian H. The constraints are

$$H_i^n = K^i{}_a q^a + J^{ia} p_a. \tag{10}$$

We make a canonical transformation $H_n(t,p,q) \to \bar{H}_n(t,\pi,y)$ such that the π_i are identical with the constraints, the $y^i \equiv f^i$ are pure gauge, and the remaining $y^r \equiv v^r$ are completely unconstrained. The generating function is

$$\Phi(p,y,t) = -p_a (X^{-1}(t))^a{}_b y^b - \frac{1}{2}\Delta_{ij}(t) f^i f^j \tag{11}$$

with

$$X_a^i \equiv K_a^i, \qquad i = 1...M, \tag{12}$$

$$J^{ia} X_a^r \equiv 0, \qquad i = 1...M, \quad r = 1...N, \tag{13}$$

$$\Delta^{ij} \equiv J^{ia} K_a^j, \qquad \Delta_{ij} \equiv \left(\Delta^{ij}\right)^{-1}, \qquad i,j = 1...M, \tag{14}$$

$$\Delta^{ij} = \Delta^{ji}. \tag{15}$$

Staggered indices are contracted over the ranges $a, b = 1...N + M$ and $i, j = 1...M$. X is determined by equations (12,13) only up to a linear recombination of the v^r. $\Delta^{ij} - \Delta^{ji}$, which we have assumed to vanish, is just the commutator or Poisson bracket $\{H_i^n, H_j^n\}$ of two constraints. As they are linear in p and q, their commutator cannot contain p or q. The H_i^n are part of an expansion of the constraints of full general relativity in powers of the p and q. These constraints are closed under the Poisson bracket, which is also true in the expanded form. The zero order of that expansion are the background constraints, which hold on a background solution. Therefore we have

$$\{H_i^n, H_j^n\} = \text{linear in background constraints} = 0 \tag{16}$$

as assumed.

The first term in the generating function effects a point transformation $y = y(q, t)$, while the second term introduces an inhomogeneous term in the relation between the π and p. We read off

$$q^a = -\frac{\partial \Phi}{\partial p_a} = (X^{-1})^a_b y^b, \tag{17}$$

$$\pi_r = -\frac{\partial \Phi}{\partial v^r} = p_a (X^{-1})^a_r, \tag{18}$$

$$\pi_i = -\frac{\partial \Phi}{\partial f^i} = p_a (X^{-1})^a_i + \Delta_{ij} f^j, \tag{19}$$

$$\bar{H}_n = H_n + \frac{\partial \Phi}{\partial t} = H_n - p_a \left(\frac{\partial}{\partial t} (X^{-1})^a_b \right) y^b - \frac{1}{2} \left(\frac{\partial}{\partial t} \Delta_{ij} \right) f^i f^j. \tag{20}$$

In $\bar{H}$ the old variables q^a and p_a must be replaced using (17) and the inverse of (18,19),

$$p_a = \pi_r X^r_a + \left(\pi_j - \Delta_{ij} f^i \right) X^j_a. \tag{21}$$

We find further that

$$J^{ia}p_a + K^i_a q^a = \Delta^{ij}\pi_j, \tag{22}$$

so that the constraints (10) are replaced by $\pi_i = O$ for $r = 1...M$ in the new phase space coordinates, which was the purpose of this transformation. When the π_r are set equal to zero, that is, when the constraints are enforced, $\bar{H}_n$ of equation (20) turns into

$$H_n^{\rm GIP}(t, v^r, \pi_r) \equiv \bar{H}_n(\pi_i = 0) = \tag{23}$$

$$H_n\left(t,\ p_a \to \pi_b X^b_a - f^i\Delta_{ij}X^j_a,\ q^a \to (X^{-1})^a_b f^b\right)$$

$$+\,\pi_r(\frac{\partial}{\partial t}X)^r_a(X^{-1})^a_b y^b - f^i\Delta_{ij}(\frac{\partial}{\partial t}X)^j_a(X^{-1})^a_b y^b - \frac{1}{2}\left(\frac{\partial}{\partial t}\Delta_{ij}\right)f^i f^j.$$

This expression is the main result of this section. We still have to prove that $H_n^{\rm GIP}$ does not depend on the f^i. We have

$$\frac{d}{dt}H^n_i = \frac{\partial}{\partial t}H^n_i + \{H^n_i, H\} = \frac{\partial}{\partial t}H^n_i + \{H^n_i, H_n\} = \frac{\partial H_n}{\partial f^i}. \tag{24}$$

Here the first equality is Hamilton's principle and the second follows from (16). The third equality holds because we can choose phase space coordinates such that $H_i = \pi_i$, so that the the explicit time derivative vanishes and the Poisson bracket becomes simply a fundamental one. If $H^n_i = O$ is enforced to hold at all times, its (total) time derivative must also vanish. Therefore H_n can contain an f^i only when it is multiplied by a π_j. With $\pi_j = 0$ all f^i-dependence vanishes from $H_n^{\rm GIP}$.

We come back to the quantum theory and an important outstanding point, the equivalence of Dirac quantizing the linear constraints and solving them prior to quantization. In a choice of phase space coordinates in which the constraints are $\pi_i \chi = 0$, the Schroedinger wave function $\chi(q)$ cannot depend on the f^i. Dirac quantization of the constraints $H^i_n = 0$ leads us to the restriction of the wavefunction $\chi(q) = \chi(v, f)$ to the form $\chi(v)$ only. The gauge degrees of freedom f are simply not quantized. (This has been interpreted as the f^i being in their "ground state" (Halliwell and Hawking 1985), but see the appendix.) The Schroedinger equation and wave function are now just

$$\left(-i\frac{\partial}{\partial t} + H^{\mathrm{GIP}}_n\right)\Psi(v_n, t) = 0, \tag{25}$$

and there are no further constraints.

Calculating H^{GIP} from the action (1) by the route we have given here is tedious even for the simple case of a single scalar field on flat FRW space (see Deruelle, Gundlach and Polarski 1992, appendix). However, H^{GIP} can be derived directly from a reduced action which one obtains by varying (1) with respect to the λ^i_n and substituting the constraints thus obtained back into the action. One then has to guess what the true degrees of freedom are. In our framework they appear as the solutions to the system of linear equations (13). The calculation in this section not only justifies this Lagrangian shortcut, but also indicates what the true degrees of freedom are. They must be i) gauge-invariant in the classical sense reviewed above, i.e. they must be among (the generalizations of) Bardeen's gauge-invariant perturbations. This is precisely the meaning of equation (13) with $v^r \equiv X^r_a q^a$. They must also ii) not contain the lapse and shift perturbations, or any other λ^i_n which arise from matter gauge fields. These requirements fix the set of v^r uniquely.

We have provided a way of finding the true degrees of freedom v for a very general background and matter fields, and a justification of Mukhanov's reduced-Lagrangian shortcut, although strictly speaking only for quantum states of the form (6). A canonical transformation to solve the constraints first used by Shirai and Wada (1988) (see the appendix).

3. Heisenberg picture and Fock vacua

We now take the Schroedinger equation (25) as our new starting point. How can one then quantize in the Heisenberg picture, just using the Hamiltonian H_n^{GIP}? This is relevant in particular if we have (approximate) solutions of the equations of motion at hand, and just need to know appropriate initial values for the mode functions of equation (5). As we shall see, for a linear system quantization is synonymous with finding all (complex) classical solutions, not only in the Heisenberg but also in the Schroedinger picture. The unsual feature is that although the classical equations of motion are linear, they are typically coupled. Theoretically one could go to action-angle variables by a linear, time-dependent canonical transformation, thus reducing the quantum problem to N harmonic oscillators. However, this seems impractical even for $N = 2$, as one obtains nonlinear matrix differential equations. We therefore quantize in the coordinates y at hand. In cosmology, one typically considers a preferred quantum state. We construct a set of preferred solutions, which can be characterized equivalently as the Fock vacua, Gaussian states or WKB states. This set has $N(N+1)$ real parameters. The ground state at any given time is part of the set.

In order to simplify expressions we slightly change our notation from the preceding sections: Throughout we consider one "Fourier" component n at a time and suppress the suffix n everywhere. We forget about the existence of the f and go back to denoting v and π_v by the standard q and p. We also suppress the indices a and write the column vector q^a as q and the row vector p_a as p^t, so that the contraction $q^a p_a$ bcomes $p^t q$ or $q^t p$. Here and in the following all lower-case letters denote column vectors and all upper-case letters denote square matrices. M^t denotes the transpose, M^* the complex conjugate, and $M^+ \equiv M^{*t}$ the hermitian conjugate of a matrix M. We reserve the symbol † for the hermitian conjugate of an operator on Hilbert space.

Consider the phase space R^{2N} and denote a point in phase space by

$$q = \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix}, \quad p = \begin{pmatrix} p_1 \\ \vdots \\ p_N \end{pmatrix}, \quad \phi = \begin{pmatrix} p \\ q \end{pmatrix}. \tag{26}$$

Consider now the most general time-dependent Hamiltonian giving rise to linear dynamics,

$$H = \frac{1}{2}\phi^t M(t)\phi, \quad \text{where} \quad M = \begin{pmatrix} A & B \\ B^t & C \end{pmatrix}, \quad A = A^t, \quad C = C^t. \tag{27}$$

H_n^{GIP} in particular is of this form. The Hamilton equations are

$$\dot{\phi} = \epsilon M(t)\phi, \tag{28}$$

where ϵ is the $2N \times 2N$ matrix

$$\epsilon = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{29}$$

$\omega(\phi_1, \phi_2) = \phi_1^t \epsilon \phi_2$ is the conserved symplectic form on phase space.

Quantizing in the Heisenberg picture, we now consider p and q as time-dependent Hermitian operators with the usual commutation relations $[q_r, p_s] = \delta_{rs}$, or, in matrix notation,

$$\left[\phi, \phi^t\right] = i\epsilon. \tag{30}$$

The Hamilton equations arise again as the Heisenberg equations. We now separate the time evolution and noncommutativity aspects by defining

$$\phi(t) = \Phi(t)\phi_0. \tag{31}$$

p_0 and q_0 are time-independent operators, while Φ is a time-dependent matrix of numbers. Φ then obeys the matrix equation

$$\dot{\Phi} = \epsilon M \Phi. \tag{32}$$

Φ is real if ϕ is hermitian; and if we impose the commutation relations (30) also on ϕ_0, Φ must be symplectic,

$$\Phi\epsilon\Phi^t = \epsilon. \tag{33}$$

Defining ϕ_0 such $\phi(0) = \phi_0$, or $\Phi(0) = 1$, would be a possible choice, but we may consider any other symplectic initial value for $\Phi(0)$.

We now construct constant annihilation and creation operators,

$$\begin{pmatrix} a \\ a^\dagger \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} p_0 \\ q_0 \end{pmatrix}, \tag{34}$$

which obey the standard commutation relations $[a_r, a_s^\dagger] = \delta_{rs}$ etc. Of course we can express $p(t)$ and $q(t)$ in terms of a and $a^\dagger$ instead of p_0 and q_0:

$$\phi(t) = \Psi(t) \begin{pmatrix} a \\ a^\dagger \end{pmatrix} \quad \Rightarrow \quad \Phi = \Psi \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} \quad \Rightarrow \quad \Psi\epsilon\Psi^t = -i\epsilon. \tag{35}$$

Φ and Ψ can be parametrized as

$$\Psi = \begin{pmatrix} P & P^* \\ Q & Q^* \end{pmatrix}, \quad \Phi = \sqrt{2} \begin{pmatrix} Re\, P & Im\, P \\ Re\, Q & Im\, Q \end{pmatrix}, \tag{36}$$

where the matrices P and Q are subject to the conditions

$$PP^{+} - P^{*}P^{t} = 0, \quad QQ^{+} - Q^{*}Q^{t} = 0, \quad QP^{+} - Q^{*}P^{t} = i. \tag{37}$$

These follow from (33) and are therefore again equivalent to the canonical commutation relations. The matrix $Q(t)$ generalizes the positive frequency solutions, called mode functions, which appear in the Heisenberg quantization of a single degree of freedom. Its components are the $v_n^{sr}(t)$, which are among the $q_n^{ar}(t)$ of equation (5). s denotes the row and r the column of the matrix Q. The number of free real parameters (initial values) in P and Q is $N(2N+1)$, the same as in the group $\mathrm{Sp}(2N)$ of symplectic matrices.

Beginning with the vacuum state $|0\rangle$ defined by

$$a\,|0\rangle = 0, \quad \text{i.e.} \quad a_r\,|0\rangle = 0, \quad r = 1...N, \tag{38}$$

the Hilbert space of (time-independent) states can be constructed as the Fock space over $|0\rangle$ by the action of the creation operators $a^{\dagger}$ in the usual way. By construction $\langle p\rangle = \langle q\rangle = 0$ for all these states. Although the operator a in (38) is constant, $|0\rangle$ depends implicitly on the choice of P and Q: There is a family of states $|0\rangle$ which obeys an equation of the form (38), but which are physically different. For a time-**independent** Hamiltonian of the form (27), for example of the harmonic oscillator or of free fields in Minkowski spacetime, there is a unique ground state, that is, lowest energy eigenstate of the Hamiltonian. Therefore the terms vacuum, ground state and Fock vacuum (with the definite article) are often used interchangeably. Here we shall use the term Fock vacuum in the indefinite, meaning any Heisenberg picture state which is annihilated by a constant operator, and we investigate in the following the structure of the family of Fock vacua (38) for a given Hamiltonian of the form (27).

All physical information of a Fock vacuum (for free fields) is contained in the quadratic expectation values

$$\langle 0|\phi\phi^t|0\rangle \equiv \begin{pmatrix} \langle 0|pp^t|0\rangle & \langle 0|pq^t|0\rangle \\ \langle 0|qp^t|0\rangle & \langle 0|qq^t|0\rangle \end{pmatrix} = \Psi \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \Psi^t = \begin{pmatrix} PP^+ & PQ^+ \\ QP^+ & QQ^+ \end{pmatrix}, \tag{39}$$

Not all of the $N(2N+1)$ parameters in Ψ correspond to these physical expectation values. A common right $\mathrm{U}(N)$ factor in the matrices P and Q drops out of the physical quantities QQ^+, QP^+ and PP^+. N^2 parameters out of the $N(2N+1)$ can be taken to constitute this unphysical $\mathrm{U}(N)$ factor.

We now construct an explicit parametrization of P and Q in terms of a unitary matrix U without physical significance and a complex symmetric matrix Z which contains the physical information. For ease of notation we introduce the auxiliar quantities

$$A \equiv QQ^+ = A^* = A^t, \tag{40}$$

$$C \equiv PP^+ = C^* = C^t, \tag{41}$$

$$B \equiv QP^+ - \frac{i}{2} = B^*. \tag{42}$$

Now we "fix the $U(N)$ gauge" by applying a polar decomposition to Q, that is, we write

$$Q = RU, \quad P = TU, \quad \text{where} \quad R^+ = R, \quad UU^+ = 1. \tag{43}$$

R is uniquely defined and positive semidefinite. We shall see later that it is positive definite if the Fock vacuum is normalizable. Note that T has no special properties. Substituting into (40-42), we obtain subsequently

$$QQ^+ = R^2 = A \quad \Rightarrow \quad R = A^{\frac{1}{2}} \tag{44}$$

$$QP^+ = RT^+ = B + \frac{i}{2} \quad \Rightarrow \quad T^+ = A^{-\frac{1}{2}} \left(B + \frac{i}{2} \right), \tag{45}$$

$$PP^+ = TT^+ = B^t A^{-1} B + \frac{1}{4} A^{-1} - \frac{i}{2} \left(A^{-1} B - B^t A^{-1} \right) = C$$

$$\Rightarrow \quad A^{-1} B - B^t A^{-1} = 0. \tag{46}$$

$$\Rightarrow \quad C = A^{-1} \left(B^2 + \frac{1}{4} \right), \tag{47}$$

The square root of A is well-defined, as R^2 is positive definite. From our parametrization we can derive the Heisenberg inequality in the form

$$\mathrm{tr}\, \langle 0 | q q^t | 0 \rangle \langle 0 | p p^t | 0 \rangle = \mathrm{tr}\, AC = \mathrm{tr} \left(B^2 + \frac{1}{4} \right) \geq \frac{N}{4}. \tag{48}$$

(46) and (47) are constraints on the possible values (39) can take, while (44) and (45) give R and T in terms of these values.

The parameters A and B are not independent. We can reparametrize them together by a complex symmetric matrix Z, defined as

$$Z = Z^t = A^{-1} \left(B + \frac{i}{2} \right). \tag{49}$$

The symmetry of Z is precisely equivalent to (46). We are justified in labelling the Fock vacua $|0\rangle$ as $|Z\rangle$. (To be precise we should say that the free parameter is the initial value of Z at some fixed time.) We have chosen Z because both because it contains no unphysical parameters and because it allows an intuitive description of the class of Fock vacua either as the class of Gaussian states or the class of WKB states. These concepts are of course defined in the Schroedinger picture, to which we go over now.

One obtains the Schroedinger picture from the Heisenberg picture through a time-dependent unitary transformation on Hilbert space. As the states become time-dependent, so must the annihilation operators, as equation (40) remains invariant. The fundamental operators p and q become time-independent by the definition of the Schroedinger picture. We can take a shortcut to finding the unitary transformation to the Schroedinger picture by choosing its usual configuration space representation: we simply make the replacements

$$q_H(t) \to q_S, \quad p_H(t) \to p_S = -i\frac{\partial}{\partial q_S}, \tag{50}$$

where q_H and p_H denote the Heisenberg picture operators considered so far, while q_S acts as a c-number. The annihilation and creation operators, now time-dependent, become:

$$\begin{pmatrix} a_H \\ a_H^\dagger \end{pmatrix} = \Psi^{-1}(t)\begin{pmatrix} p_H(t) \\ q_H(t) \end{pmatrix} \to \begin{pmatrix} a_S(t) \\ a_S^\dagger(t) \end{pmatrix} = i\begin{pmatrix} Q^+ & -P^+ \\ -Q^t & P^t \end{pmatrix}\begin{pmatrix} -i\frac{\partial}{\partial q_S} \\ q_S \end{pmatrix}. \tag{51}$$

The Heisenberg state $|\psi\rangle$ is replaced by the wave function $\psi(q,t)$. In the following we remain in the Schroedinger picture and from now on q denotes q_S. The Fock vacuum $|Z\rangle$ in particular has the wave function

$$\Psi_Z(q,t) = \exp\left(-\frac{1}{2}f(t) + \frac{i}{2}q^t Z(t) q\right). \tag{52}$$

As

$$Z = \left(Q^+\right)^{-1} P^+, \tag{53}$$

we find that $a_S(t)\psi_Z(q,t) = 0$. Hence ψ_Z is a Fock vacuum, but now parametrized by the physical parameter Z rather than by P and Q, which contain the unphysical $\mathrm{U}(N)$ factor.

The Schroedinger equation

$$\left[-i\frac{\partial}{\partial t}+H\left(-i\frac{\partial}{\partial q},q,t\right)\right]\Psi_Z(q,t)=0 \tag{54}$$

is equivalent to a first-order equation for Z,

$$\dot{Z}+ZAZ+ZB+B^tZ+C=0, \tag{55}$$

and for the phase f,

$$\dot{f}=\mathrm{tr}AZ+\mathrm{tr}B \tag{56}$$

The initial values of f have no physical significance: The initial value of $Im\, f$ determines an arbitrary constant phase in Ψ_Z, and $Re\, f$ is given at all times by the normalization of Ψ_Z. The state given by Z is normalizable if and only if the imaginary part of Z is invertible. From (44) and (49) this is equivalent to R being positive definite. It can be checked that Z obeys (55), using (53) and the matrix identity $(M^{-1})^{\cdot}=-M^{-1}\dot{M}M^{-1}$.

It is instructive to calculate the quadratic expectation values directly from $\psi_Z(q,t)$. Using the identity

$$\frac{\int qq^t\, e^{-\frac{1}{2}q^tMq}\, d^Nq}{\int e^{-\frac{1}{2}q^tMq}\, d^Nq}=M^{-1} \tag{57}$$

we find first that

$$\langle Z|qq^t|Z\rangle=\frac{1}{2}(Im\, Z)^{-1}, \tag{58}$$

and hence

$$\langle Z|qp^t|Z\rangle=\langle Z|qq^t|Z\rangle Z=\frac{1}{2}(Im\, Z)^{-1}(Re\, Z)+\frac{i}{2}, \tag{59}$$

$$\langle Z|pp^t|Z\rangle=Z\langle Z|qq^t|Z\rangle Z-iZ=\frac{1}{2}(Re\,)Z(Im\, Z)^{-1}(Re\, Z)+\frac{1}{2}(Im\, Z)^{-1}. \tag{60}$$

These expressions follow also from (39-42,47,49).

An important motivation for considering Fock vacua for the states of linearized perturbations in cosmology is that they coincide with the semiclassical states, as we show now. It is clear that a classical action, or solution of the Hamilton-Jacobi equation, must be quadratic in the q, as the Hamiltonian is derived from a quadratic action functional. In fact,

$$S_{cl}(q,t) = \frac{1}{2} q^t Z(t) q \tag{61}$$

obeys the Hamilton-Jacobi equation

$$\frac{\partial S_{cl}}{\partial t} + H\left(\frac{\partial S_{cl}}{\partial q}, q, t\right) = 0 \tag{62}$$

precisely when Z obeys the equation of motion (55), the solutions of which are given by (53). S_{cl} describes a family of solutions of the Hamilton equations fibrating configuration space by means of the vector field

$$p_{cl}(q,t) = \left(\frac{\partial S_{cl}}{\partial q}\right)^t = P^*(t)Q^*(t)^{-1}q. \tag{63}$$

$S_{cl}(q_f, t_f)$ is therefore the classical action, up to $t = t_f$, of the solution q_{cl} which goes through $q = q_f$ at $t = t_f$,

$$q_{cl}(q_f, t_f, t) = Q^*(t)Q^*(t_f)^{-1}q_f. \tag{64}$$

q_{cl} is realized as a superposition of column vectors of $Q^*(t)$. Such a superposition exists uniquely as Q and P are invertible.

By a semiclassical state or WKB state we mean the usual thing, a solution of the Schroedinger equation to order $\hbar^2$ of the form

$$\psi(q,t) = A(q,t)\, e^{\frac{i}{\hbar} S_{cl}(q,t)} \tag{65}$$

where S_{cl} obeys the appropriate Hamilton-Jacobi equation and where A and S_{cl} do not depend on $\hbar$. We then find by comparison with (54) that for the case under consideration we have already found these states in the Fock vacua. It is peculiar to the linear dynamics considered here that $A(q,t) = A(t)$ and that the WKB ansatz gives an exact solution.

Finally we consider the relation between the Fock vacua and the ground state. For an explicitly time-dependent Hamiltonian the "ground state" (i e the lowest eigenstate of the Hamiltonian) at one time will not be the ground state at another time. The instantaneous ground state at $t = t_1$, say $|Z_1\rangle$, must be an instantaneous eigenstate of the Hamiltonian, i.e. $H(t_1)|Z_1\rangle = E_0(t_1)|Z_1\rangle$. We can find it in the configuration representation of the Schroedinger picture, where this condition becomes $i\partial/\partial t\, \psi_{Z_1}(t_1) = E_0 \psi_{Z_1}(t_1)$. But that just means that $\dot{Z}_1(t_1) = 0$, which gives $Z_1(t_1)$ from equation (55). Now it is clear that all the other eigenstates of the Hamiltonian at $t = t_1$ are those created by the action of the operators $a^\dagger$ on the particular Fock vacuum $|Z_1\rangle$. As they all have higher energies, this Fock vacuum is the instantaneous ground state. For perturbation modes of inflationary cosmology we often find that the Hamiltonian becomes time-indepent in the asymptotic past. It is then natural to consider the ground state in the asymptotic past, also called adiabatic vacuum, as a preferred (Fock vacuum) state. For the perturbations of a scalar field in a homogenous inflating universe this state has been extracted from the no boundary proposal (Laflamme 1987). The ground state energy can be calculated as

$E_0 = \frac{1}{2}\sqrt{tr(AC - B^t A^{-1} BA)}$, which is of course explicitly time-dependent. (As the sum over all field modes n of $E_0(t) - E_0(t + \Delta t)$ is infinite, the Bogoliubov transformation from the instantaneous ground state at one time to that at any other time will not be unitarily representable by an S-matrix for the entire field theory, although that is possible for a single n Fourier component.)

4. Conclusions

I have given a treatment of one-loop quantum gravity (or free quantum field theory in curved spacetime including linearized gravity) around a minisuperspace solution without backreaction terms. The constraints are solved by finding the true degrees of freedom before quantization. As we have assumed the factor ansatz (6) for the wave function, so that we are dealing only with a finite number of degrees of freedom at a time, the quantization given here is unique. It is an improvement on the approximate quantization schemes still prevalent in the literature on quantum fluctuations in inflation, and both a justification – although only for this form of the wave function! – and a generalization of the quantization scheme of Mukhanov (1988).

The new approach has already been used on a previously unattempted problem, namely to quantize the perturbations of extended chaotic inflation (Deruelle, Gundlach and Langlois 1992a, b), where there are two true degrees of freedom.

Technically the gauge-invariant approch relies on the existence of coordinates q_n for a given background spacetime which split the action into an infinite sum of noninteracting terms. These coordinates exist for FRW background spacetimes (D'Eath and Halliwell 1987). It should be possible to construct them also on other homogeneous background spacetimes, where they are just the spatial harmonics. This would include not only homogeneous cosmologies, but also, for example, bubble nucleation instantons. In principle they exist for any background.

The gauge-invariant approach trades gauge-invariance for covariance: The actions associated with the Hamiltonians for each component v_n cannot be reassembled into a covariant action for a field v. This happens at least for the true degree of freedom v arising from a single scalar field on $K = -1$ or $K = +1$ FRW space (Hawking et al 1992, see also Shirai and Wada 1988).

Noncovariance will make it impossible to renormalize a quantum energy momentum tensor based on the true degrees of freedom by the standard covariant methods (Birrell and Davies 1982). From a pragmatic point of view, this is no great loss: i)There are arguments that the approximation of a classical background is only valid to the extent to which the backreaction is small. ii) The interesting quantity to compare with cosmological observation is not the energy-momentum tensor, which arises from the quantum mechanical two-point function at the coincidence point. Rather it is Fourier components of the two-point functions on length scales interesting for the formation of cosmological structure. One may therefore disregard the ultraviolet divergence which must be dealt with in the stress tensor, as any renormalization prescription affects only the very short wavelength components of the two-point functions.

Acknowledgments

I am grateful to David Langlois, Klaas Landsmann and Chris Fewster for helpful discussions.

Appendix: Previous solutions of the constraints

The canonical transformation of section 2 has two predecessors. The first one is instructive to review for the following reason: To solve the perturbative lapse and shift constraints, we had to use the semiclassical limit for the background $\{Q^A\}$. This is not surprising, as the true degrees of freedom v left behind must be among the generalizations of the gauge-invariant perturbations that Bardeen (1980) has constructed for a FRW background. Although among those one can find enough linear combinations which depend only on the q_n^a and not on the lapse and shift perturbations λ_n^i, their coefficients contain time derivatives of the background Q^A. But time derivatives have a meaning only for a classical background. Vice versa, the solution of the shift constraints alone should not require the semiclassical limit, because it only corresponds to finding the true degrees of freedom within a single t-hypersurface.

These considerations are confirmed by the existence of the following approximate solution, given by Wada(1986), of the shift constraint, say of the scalar sector. The shift constraints of general relativity are linear in the momenta. This also holds for the midisuperspace ansatz, so that the shift constraints must be linear in either P or p. All perturbative constraints are also linear in either p or q. The perturbative shift constraint is therefore of the form

$$(F^a p_a + W_a^A P_A q^a)\Psi(Q,q) = 0. \tag{A.1}$$

where $P_A = -i\partial/\partial Q^A$ and $p = -i\partial/\partial q$. In the semiclassical limit $P_A = \partial S_0/\partial Q^A$ where $S_0(Q)$ is a solution of the Hamilton-Jacobi equation: (A.1) is of the form (10) again, and section 3 applies. But here we do not assume that limit. If we make the ansatz

$$\Psi(Q,q) = \Psi(\bar{Q},v), \tag{A.2}$$

$$v^r = X^r_a(Q)\, q^a, \tag{A.3}$$

$$\bar{Q}^A = Q^A + \frac{1}{2}\, \delta^A_{ab}(Q)\, q^a\, q^b, \tag{A.4}$$

substitute into the constraint and reorder in powers of q, we obtain

$$\left\{ \left(\left[\delta^A_{ab} F^b + W^A_a\right] q^a + W^B_a \delta^A_{bc,B} q^a q^b q^c \right) P_A + \left(\left[X^r_a F^a\right] + W^B_a X^r_{b,B} q^a q^b \right) \pi_r \right\} \Psi = 0. \tag{A.5}$$

Clearly it is necessary that all four (Q-dependent) coefficients vanish separately. The condition $X^r_a F^a = 0$ is the same as (14), as it must, specifying what are the true degrees of freedom. The δ^A_{ab} are underdetermined by the condition $\delta^A_{ab} F^b + W^A_a = 0$ alone, but overdetermined by the remaining set of derivative constraints. Nor can the X^r_a obeying $X^r_a F^a = 0$ also satisfy $W^B_a X^r_{b,B} = 0$. Therefore one will in general have to neglect the terms not in the square brackets. Now if one allows only states Ψ in which the expectation value of all the q and p is $O(\epsilon)$, the expectation values of those terms is of $O(\epsilon^3)$ and can be neglected with respect to the terms in square brackets, which are of $O(\epsilon)$. In this method it is unclear what the meaning of the $\bar{Q}$ is and therefore how to obtain the gauge-dependent quantitites back after quantization. The "$\bar{Q}$ method" of solving the shift constraints was used by Wada (1986) for the closed FRW background with a cosmological constant, and with a single scalar matter field by Shirai and Wada (1988). There it is accidentally possible to make all four terms in (A.5) vanish.

The other predecessor was first used by Halliwell and Hawking (1985), for the same background, but for the vector perturbations (called c_n there). As these are already pure gauge, the point transformation part of our canonical transformation is trivial. The shift in their momenta is obtained by making the wave function a Gaussian where the exponent is $-\frac{1}{2}\Delta_{ij}f^if^j$. A $p=-i\partial/\partial q$ acting on a Gaussian becomes a q. What Halliwell and Hawking call the "ground state" is really only the source of the last term in the canonical transformation (19). This "gauge Gaussian" method was applied to the scalar lapse constraint by Shirai and Wada (1988), after solving the shift constraints by the $\bar{Q}$ method. In that approach some of the terms in what is $H_n^{\rm GIP}$ in our notation arise as changes in the background Hamiltonian $H_0(Q,P)$ due to the pure point transformation from Q,q to $\bar{Q},q$, which changes **all** partial derivatives and hence momenta. Our approach is equivalent in the final form of $H_n^{\rm GIP}$, but perhaps clearer, as it independent of the background quantization.

References

Bardeen J M 1980 *Phys. Rev.* **D 22** 1882

Bardeen J M, Steinhardt P J and Turner M S 1983 *Phys. Rev.* **D 28** 679

Birrell N D and Davies P C W 1982 *Quantum fields in curved space* (Cambridge: CUP)

D'Eath P D and Halliwell J J 1987 *Phys. Rev.* **D 35** 1100

Deruelle N, Gundlach C and Langlois D 1992a *Phys. Rev.* **D 45** 3301

Deruelle N, Gundlach C and Langlois D 1992b 'The perturbations of extended chaotic inflation' to be published in *Phys. Rev.* **D**

Deruelle N, Gundlach C and Polarski D 1992 *Class. Quantum Grav.* **9** 137

Guth A and Pi S-Y 1982 *Phys. Rev. Lett.* **49** 1110

Halliwell J J and Hawking S W 1985 *Phys. Rev.* **D 31** 1777

Hawking S 1982 *Phys. Lett.* **115 B** 295

Hawking S W, Laflamme R and Lyons G W 1992 'Cosmological perturbations and the arrow of time' in preparation

Laflamme R 1987 *Phys. Lett.* **B 198** 156

Mukhanov V F 1988 *Zh. Eksp. Tor. Fiz.* **94** 1 (*Sov. Phys.* JETP 68 1297)

Mukhanov V F, Feldman H A and Brandenberger R 1990 'Theory of cosmological perturbations', preprints Brown-HET-780, 796 and 800

Salopek D S, Bond J R and Bardeen J M 1989 *Phys. Rev.* **D 40** 1753

Shirai I and Wada S 1988 *Nucl. Phys.* **B 303** 728

Starobinsky A A 1982 *Phys. Lett.* **117 B** 1982 175

Wada S 1986 *Nucl. Phys.* **B 276** 72