Chapter 3
Kinematics: Spatial Atoms

Quantum geometry determines properties of quantized space–time structures, which can be interpreted as providing an atomic understanding of space–time. A view results which is fascinating not only in its physical implications but also in its rich combination of aspects of geometry and quantum theory. Many relevant features can already be seen by analogy with quantized particles, then borne out by rigorous constructions of quantum space–time.

3.1 Quantized Particles

Different aspects seen already in single-particle quantum mechanics are important in the context of quantum space–time as well. First, we consider an ordinary free and non-relativistic particle. Its well-known solutions show that the wave function in general spreads out in time even if the particle is not moving. Clearly, there is more freedom in quantum compared to classical dynamics: quantum variables such as fluctuations usually change independently of what one classically considers as the degrees of freedom; even a particle which classically would stay at rest can have non-trivial quantum dynamics. The degree of spreading can easily be determined by solving an equation of motion for the position fluctuation: With the Hamiltonian $\hat{H} = \hat{p}^2/2m$ and the general identity

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\hat{O}\rangle = \frac{\langle[\hat{O},\hat{H}]\rangle}{i\hbar} \tag{3.1}$$

for an arbitrary operator $\hat{O}$, we have the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\langle\hat{q}^2\rangle - \langle\hat{q}\rangle^2\right) = \frac{\langle[\hat{q}^2,\hat{H}]\rangle - 2\langle\hat{q}\rangle\langle[\hat{q},\hat{H}]\rangle}{i\hbar} = \frac{1}{m}\langle\hat{q}\,\hat{p} + \hat{p}\hat{q}\rangle - \frac{2}{m}\langle\hat{q}\rangle\langle\hat{p}\rangle = \frac{2}{m}C_{qp}. \tag{3.2}$$

The fluctuation $(\Delta q)^2 = \langle\hat{q}^2\rangle - \langle\hat{q}\rangle^2$ is not guaranteed to remain constant in time; more precisely, its spreading is controlled by the covariance

M. Bojowald, *Quantum Cosmology*, Lecture Notes in Physics 835,
DOI: 10.1007/978-1-4419-8276-6_3, © Springer Science+Business Media, LLC 2011

$C_{qp} = \frac{1}{2}\langle \hat{q}\hat{p} + \hat{p}\hat{q}\rangle - \langle\hat{q}\rangle\langle\hat{p}\rangle$ of the state. A usual unsqueezed Gaussian state, for instance, which is often used for an initial profile and has the form $\psi(q) \propto \exp(-q^2/4\sigma^2)$ with a real variance σ, has vanishing covariance. Such an initial state would ensure that the initial spreading does not change momentarily. However, as a function of time the covariance must satisfy another equation of motion:

$$\frac{\mathrm{d}}{\mathrm{d}t}C_{qp} = \frac{2}{m}(\Delta p)^2 \tag{3.3}$$

again derived using (3.1). The covariance could be constant only if the momentum fluctuation Δp vanishes, which cannot be the case for a normalizable state. On the other hand, the equation of motion for Δp itself, derived by the same methods, tells us that it is a constant in time. We can thus solve (3.3) for $C_{qp}(t) = 2(\Delta p)^2 t/m + C_{qp}^{(0)}$ in terms of its initial value $C_{qp}^{(0)}$. This solution, in (3.2), gives

$$\Delta q(t) = \sqrt{\frac{2}{m^2}(\Delta p)^2 t^2 + \frac{2}{m}C_{qp}^{(0)} t + \Delta q^{(0)}} \tag{3.4}$$

as the well-known result showing the spreading of a free-particle state in time.

We have discussed this familiar example, already encountered in Sect. 2.3, at some length because it illustrates useful methods which we will come back to later, and because it provides important lessons. As seen clearly in this simple example, while one is always free to choose an initial state and make it as simple as possible, quantum dynamics in general changes its properties as time goes on. Here, we have seen that a vanishing covariance cannot be maintained in time; states tend to get "squeezed" and develop non-vanishing correlations. This is true even in situations which one would consider as semiclassical, and here correlations are even an integral part of decoherence scenarios [1].

To visualize the meaning of correlations, we use the second-order moments $(\Delta q)^2$, C_{qp} and $(\Delta p)^2$ of a state to define the family of ellipses

$$q^2(\Delta p)^2 + 2qpC_{qp} + p^2(\Delta q)^2 = \text{const} \tag{3.5}$$

around the origin in the $q - p$-plane. These ellipses demonstrate the amount of quantum fluctuations: for $C_{qp} = 0$, for instance, we have an ellipse of semimajor axes Δq along the q-axis and Δp along the p-axis. For $C_{qp} \neq 0$, these ellipses are rotated such that certain linear combinations of q and p show the maximal and minimal fluctuations. A distribution function with these properties can be computed from the wave function: the Wigner function

$$W(q,p) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} \psi^*\left(q + \frac{1}{2}\alpha\right)\psi\left(q - \frac{1}{2}\alpha\right)e^{-\mathrm{i}p\alpha/\hbar}\mathrm{d}\alpha. \tag{3.6}$$

The factor of $1/2\pi\hbar$ ensures that $W(q,p)$ and the marginal distributions it provides by integrating over q or p, respectively, are normalized: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(q,p)\mathrm{d}q\mathrm{d}p = 1$.

(A probability distribution in a strict sense, that is a non-negative function, is obtained if and only if $\psi(q)$ is Gaussian.)

For a Gaussian state

$$\psi(q) = \exp\left(-\tfrac{1}{4}(\sigma_{\mathrm{R}}^{-2} + \mathrm{i}\sigma_{\mathrm{I}}^{-2})q^2\right) \tag{3.7}$$

of arbitrary squeezing, we obtain the Wigner function

$$W(q,\, p) = \frac{1}{\pi\hbar}\exp\left(-\tfrac{1}{2}\left(\sigma_{\mathrm{R}}^{-2} + \sigma_{\mathrm{R}}^2/\sigma_{\mathrm{I}}^4\right)q^2 + 2\sigma_{\mathrm{R}}^2\sigma_{\mathrm{I}}^{-2}qp/\hbar - 2\sigma_{\mathrm{R}}^2p^2/\hbar^2\right). \tag{3.8}$$

In terms of fluctuations and the covariance, related to $\sigma_{R/I}$ via $(\Delta q)^2 = \sigma_R^2$. $(\Delta p)^2 = \hbar^2/4\sigma_R^2 + \hbar^2\sigma_R^2/4\sigma_I^4$ and $C_{qp} = -\hbar\sigma_R^2/2\sigma_I^2$ (see for instance (13.6)), we can write the exponent as

$$E := -2\hbar^{-2}\left(q^2(\Delta p)^2 + 2qpC_{qp} + p^2(\Delta q)^2\right). \tag{3.9}$$

Constant-level lines of the Gaussian Wigner function in phase space are thus ellipses. In order to determine their proportions, we choose reference values q_0 and p_0 with the dimensions of q and p. respectively, and work with dimensionless ratios q/q_0 and p/p_0. In the absence of a ground state or other specific features of states, no distinguished values for q_0 or p_0 can be provided (unlike, for instance, if one could refer to the harmonic-oscillator ground state with fixed fluctuations of the correct dimensions). After dividing E by $q_0^2p_0^2$, we express all terms by dimensionless variables, in which we find the major axis of the ellipse rotated against the q-axis by an amount

$$\tan(2\alpha) = \frac{q_0p_0C_{qp}}{q_0^2(\Delta p)^2 - p_0^2(\Delta q)^2}. \tag{3.10}$$

An interpretation of the covariance is thus as the rotation of the likelihood ellipse in phase space. The axes lengths of the ellipse are

$$\frac{p_0^2(\Delta q)^2 + q_0^2(\Delta p)^2 \pm \sqrt{(p_0^2(\Delta q)^2 - q_0^2(\Delta p)^2)^2 + 4q_0^2p_0^2C_{qp}^2}}{2q_0^2p_0^2}$$

which thanks to $(\Delta q)^2(\Delta p)^2 - C_{qp}^2 = \hbar^2/4$ (a Gaussian state saturates the uncertainty relation) can be written as

$$\frac{p_0^2(\Delta q)^2 + q_0^2(\Delta p)^2 \pm \sqrt{(p_0^2(\Delta q)^2 + q_0^2(\Delta p)^2)^2 - q_0^2p_0^2\hbar^2}}{2q_0^2p_0^2}.$$

Since there are no distinguished values for q_0 and p_0 in general, the only available meaning of the squeezing of states is by non-vanishing correlations, rotating the likelihood ellipse. Changing position and momentum fluctuations while keeping the uncertainty relation saturated at vanishing correlations provides a meaningful sense of squeezing only if one can refer to a distinguished state, such as the harmonic-oscillator ground state. In quantum cosmology, no ground state is available to define squeezing in the absence of correlations. From now on, we will use only the general sense of squeezing as determined by $C_{qp} \neq 0$.

If $(\Delta p)^2$ is constant in time (as for the free particle), $(\Delta q)^2$ must be large for large C_{qp}. The major axis of the ellipse then has a length approximately given by Δq, while the minor axis is very small. The ellipse is stretched out in one direction, along a linear combination of q and p determined by the covariance or the angle α in (3.10). While this linear combination has large fluctuations, the orthogonal one has very small ones and thus behaves rather

classical. In this way the emergence of a classical degree of freedom via decoherence can be seen, whose precise form is determined by the underlying dynamics. Squeezed states with large covariance automatically arise in the process, and play an important role for the nearly classical behavior. In general quantum systems deviations from Gaussian form more general than squeezing arise.

As far as semiclassical regimes are concerned, we arrive at our

First Principle State dynamics is important and to be derived. In particular, the form of appropriate semiclassical states cannot always be guessed, or assumed to be unsqueezed Gaussians.

From elementary particle physics, for which perturbations around the free vacuum state are often sufficient, one is used to Gaussian states to play a central role. But the form of the vacuum is a dynamical question, and general situations in quantum cosmology may not even allow a distinguished vacuum or ground state. More general classes of states and methods to deal with them must be used. In Sect. 5.4.1.3 we will see an example of states rapidly moving away from Gaussian form, then settling into a new, better preserved shape as determined by its moments [2]. Other example systems have dynamical coherent states of exactly preserved shape, but these systems, such as the harmonic oscillator, are very special and rarely realistic. More generally, states of "stable" shape exist [3], but this is a mathematical rather than physical generalization of the desired properties of dynamical coherent states. In fact, in these more general states the shape does change: As time goes on, the entire state evolution is determined by the change of as many parameters as there are classical degrees of freedom; however, unlike in the harmonic-oscillator case, these parameters are not in one-to-one correspondence with expectation values. They partially affect fluctuations and other moments of the state as well, and thus the state's shape evolves. In particular, semiclassicality may be lost as the states evolve.

Considerations of the free particle in quantum mechanics offer another observation. If the particle is very massive or macroscopic with large m in (3.4), it takes a long time for the wave function to spread out from some tightly peaked initial state. In a naive interpretation, this would suggest that macroscopic bodies do not show quantum effects at all. This conclusion can, of course, not be true since there are important properties even in macroscopic situations, such as conductivity, which rely on quantum aspects of their constituents. At this stage, the composite, atomic nature of matter becomes important: microscopic building blocks are much smaller, and they almost always behave very quantum. This is an obvious statement for condensed-matter physics, but it shows that quantum cosmology, where the dominant view is usually one of a large macroscopic and homogeneous universe, must properly take into account the underlying atomic nature of space–time if it is to describe all quantum phases of the universe reliably.

Second Principle Microscopic physics is important. In cosmology, even homogeneous models must include the proper small-scale quantum behavior. They constitute many-body systems when seen in quantum gravity; the large "number of particles", corresponding to $\mathcal{N}(t)$ of the preceding chapter, may lead to characteristic effects.

For instance regarding the singularity problem, a massive homogeneous "blob" universe, which contains all its matter smeared-out, may be non-singular in some models. For instance, examples exist in which effective violations of energy conditions can trigger a "bounce" where the isotropic universe volume is minimal [4]. But if the microscopic dynamics of its quantum building blocks remains singular, which is a question more complicated to address, the theory is still in danger of breaking down.

For the dynamics of many-particle systems at high energies, quantum field theory rather than particle quantum mechanics is required. Here, one starts with fields on a given space–time and applies quantization techniques. For gravity and cosmology, however, it is the space–time itself that is to be quantized. Familiar techniques, which always implicitly assume the availability of a background space–time, then fail. Canonical quantization of a scalar field on Minkowski space–time, for instance, might make use of a mode expansion

$$\phi(\boldsymbol{x}) = \int \frac{\mathrm{d}^3 k}{\sqrt{2\omega_{\boldsymbol{k}}}} \left(\hat{a}_{\boldsymbol{k}} e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} + \hat{a}^\dagger_{\boldsymbol{k}} e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\right). \tag{3.11}$$

to define annihilation operators $\hat{a}_{\boldsymbol{k}}$ and distinguish the vacuum state $|0\rangle$ as the state annihilated by all $\hat{a}_{\boldsymbol{k}}$. Many-particle states are obtained by acting with creation operators, the adjoints of annihilation operators, on the vacuum:

$$|\boldsymbol{k}_1, n_1; \ldots; \boldsymbol{k}_i, n_i\rangle = (\hat{a}^\dagger_{\boldsymbol{k}_1})^{n_1} \cdots (\hat{a}^\dagger_{\boldsymbol{k}_i})^{n_i} |0\rangle. \tag{3.12}$$

In such states, the total normal-ordered energy as the eigenvalue in

$$\hat{E}|\boldsymbol{k}_1, n_1; \ldots; \boldsymbol{k}_i, n_i\rangle = \sum_{j=1}^{i} \hbar n_j \omega(\boldsymbol{k}_j) |\boldsymbol{k}_1, n_1; \ldots; \boldsymbol{k}_i, n_i\rangle \tag{3.13}$$

is non-zero.

The mode decomposition, however, requires space–time equipped with a background metric to be defined: at least the integration measure $\mathrm{d}^3 x\sqrt{\det h}$ must be known in order to integrate the original field and obtain its modes (or use preferred Cartesian coordinates in which the spatial metric is $h_{ab} = \delta_{ab}$). In (3.11), the scalar product $\boldsymbol{k}\cdot\boldsymbol{x}$ refers to a flat background. Without the modes, we cannot even define the vacuum state. This consideration finally provides the third principle:

Third Principle Tools of quantum field theory must be appropriately adapted to deal with quantum geometry in a background-independent way. While the simplest cosmological models are homogeneous and of finitely many degrees of freedom, allowing straightforward quantizations of geometrical variables, significant changes in the quantization methods due to the generally covariant nature of the underlying field theory must also be reflected in quantum cosmology.

Quantum cosmology must take into account the lessons learned in attempted constructions of quantum gravity. Constructing quantum gravity or even deriving

quantum cosmology from it remains a formidable challenge, but important features can nevertheless be implemented and explored in sufficiently general formulations of quantum cosmological models. Sufficient generality is important for reliable conclusions and for stringent tests of the full framework, even if it comes at the expense of additional ambiguities.

All three principles will be revisited throughout this book; they are important for conceptual properties, for instance regarding singularities, and observational ones. We will begin by reviewing the reformulation proposed by loop quantum gravity to incorporate the Third Principle. The first two principles will have to be faced once we deal with the dynamics.

3.2 Quantized Space–Time

In general relativity, the dynamical object is the space–time metric, now to be quantized. As mentioned in the introduction, we do not require a viewpoint of general relativity being fundamental, but rather take a more general one: even if there is a more fundamental underlying theory, which may eventually be arrived at by the quantization procedure, there must be a consistent way of endowing the metric with fluctuations and uncertainty. There is a tried-and-true traditional method to unravel quantum properties of hitherto classical theories: canonical quantization, a procedure that takes a classical phase space and returns a non-commuting algebra of observables, such as pairs of basic operators on whose representation quantization can be built.

Coordinates are often used to describe space–times, but this is only superficially related to expressing point-particle dynamics by coordinates that become operators $\hat{q}$. Unlike the positions of point particles, space–time coordinates are not measurable; they can play no role in the final algebra of quantum observables. What can be measured is only the geometry and dynamics of space–time, which requires extended objects. Fully coordinate-independent observables, on the other hand, are infamously difficult to construct: only very few general expressions are known, and even approximate constructions become very tedious in anything but the simplest models. Instead of trying to quantize classical observables, a two-step procedure looks more promising: one first considers just spatial quantum geometry at a fixed time (referring to objects such as lengths, areas, volumes which for given regions are coordinate independent), and then imposes additional constraints to make sure that kinematical objects are combined suitably to space–time observables. Once complete, the results will show the quantum dynamics of space–time; but even before all constraints are implemented, spatial quantum geometry already provides interesting results.

3.2.1 Scaling

Kinematically, we consider objects such as the volume $V_R = \int_R \mathrm{d}^3x\sqrt{\det h}$ of some spatial region R, where h_{ab} is the metric induced by a space–time metric on a spatial slice $t = \text{const}$ with respect to some time coordinate. Volumes are surely sufficient to probe isotropic quantum geometries, in which case the classical phase space is small: as seen before, there is a single canonical pair (a, p_a). But if we allow all possible regions, spatial observables even in this simple geometry provide infinitely many numbers $V_R = \mathscr{V}(R)a^3$ for any given a, where $\mathscr{V}(R)$ is the coordinate-dependent, co-moving volume of the region measured just with the non-dynamical unit line element $d\sigma_k$ of constant curvature, (2.2). With a single dynamical degree of freedom, however, quantization can give wave functions only in one variable, such as $\psi(a)$ as used in a Wheeler–DeWitt quantization. It must then be ensured that wave functions have the correct scaling behavior under changing coordinates or $\mathscr{V}$ so that observables are invariant; otherwise one's quantization would not capture pure quantum geometry.

For a compact space (for instance the closed model), the total space would be a convenient choice to define $\mathscr{V}_R$. But one may still capture all isotropic degrees of freedom completely by any subspace, with a smaller value of $\mathscr{V}$. Moreover, in the closed model in its most general formulation one is not required to use the unit sphere as the total space multiplied by the scale factor, although it is certainly convenient. If a sphere of non-unit radius is used for the spatial coordinates, the coordinate volume changes to $\mathscr{V} = \lambda^3\mathscr{V}_{\text{unit}}$. One can obtain the new coordinate system by the transformation $r \mapsto \lambda r$ while the angular coordinates do not change. The parameter $k \mapsto \lambda^{-2}k$, obeying the scaling law of curvature, remains positive but now differs from one. The line element is invariant with the usual transformation $a \mapsto \lambda^{-1}a$ of the scale factor under rescaling the coordinates, and so the curvature term k/a^2 in the Friedmann equation is invariant. Notice that this rescaling of coordinates and the sphere is not the same as changing a smaller integration region within the unit sphere even if $\lambda < 1$, if one chooses a smaller integration region within the sphere, k and a do not change. Just as in the flat model, also in the closed model rescaling the coordinates and choosing different integration regions is allowed by independent choices. The parameter $\mathscr{V}$ retains a free value and is not fixed. Sometimes, $\mathscr{V}$ is called a "regulator" and then argued to require the limit $\mathscr{V} \to \infty$ (or the maximum value in a compact space) rather than full $\mathscr{V}$-independence of wave functions. However, $\mathscr{V}$ is not a regulator because the classical predictions do not depend on the chosen value. It is simply a parameter that labels different but equivalent formulations of the symmetry reduction within one and the same model. While one may have reasons to restrict all attention to a specific simple value, in doing so one loses access to an important consistency check. Especially in isotropic and homogeneous models, in which the anomaly problem, to be discussed later, trivializes, making sure that the proper $\mathscr{V}$ -behavior is realized is the only remaining test of consistency.

The scaling issue turns out to be a manifestation of a more general problem: How do we form a complete set of coordinate-independent measures of spatial geometry (intrinsic as well as extrinsic for the whole phase space) while retaining an algebra simple enough for further constructions? This problem, as of now, has not been solved in Wheeler–DeWitt-type quantizations, which when applied beyond homogeneous models remain formal. But at the level of spatial quantum geometry it can be solved after a change of phase-space variables.

3.2.2 Canonical Gravity

For a Hamiltonian formulation of general relativity (see [5]), one first brings the space–time metric in a form

$$\mathrm{d}s^2 = -N^2\mathrm{d}t^2 + h_{ab}(\mathrm{d}x^a + N^a\mathrm{d}t)(\mathrm{d}x^b + N^b\mathrm{d}t) \tag{3.14}$$

where h_{ab}, the only part contributing on a spatial slice $t = \text{const}$, is the spatial metric. The lapse function N and shift vector N^a provide the remaining components of a space–time metric, and can be seen to contain information about the spatial foliation in space–time: The unit normal vector n^a to a spatial slice $t = \text{const}$ satisfies $Nn^a = (\partial/\partial t)^a - N^a$, and the inverse space–time metric is $g^{ab} = h^{ab} + n^a n^b$.

The spatial metric h_{ab} can serve as a set of configuration variables, and a suitable geometric notion of its "velocities" is the extrinsic curvature

$$K_{ab} = \frac{1}{2N}(\dot{h}_{ab} - D_a N_b - D_b N_a). \tag{3.15}$$

It does indeed have a time derivative of the spatial metric, denoted by the dot, and extra contributions if the shift vector is not constant and thus the spatial slice is deformed as seen from the time coordinate t: the normal vector is not proportional to $(\partial/\partial t)^a$. The symbol D_a denotes the spatial covariant derivative operator compatible with the metric h_{ab}.

3.2.2.1 Action and Constraints

In these variables, the Einstein–Hilbert action of general relativity (ignoring boundary terms) can be expressed as

$$\begin{aligned} L_{\mathrm{grav}} &= \frac{1}{16\pi G}\int \mathrm{d}^3x\sqrt{-\det g}R \\ &= \frac{1}{16\pi G}\int \mathrm{d}^3x N\sqrt{\det h}\left({}^{(3)}R + K_{ab}K^{ab} - (K_a{}^a)^2\right) \end{aligned} \tag{3.16}$$

with the three dimensional Ricci scalar ${}^{(3)}R$ computed from h_{ab}. In this way, the action looks like a complicated version of the general form known from classical field theories, with a kinetic term quadratic in the velocities K_{ab} and a potential depending only on configuration variables as well as their spatial derivatives, here given by the spatial Ricci scalar. From the kinetic term, we first compute the momentum

$$p^{ab}(x) = \frac{\delta L_{\mathrm{grav}}}{\delta\dot{h}_{ab}(x)} = \frac{1}{2N}\frac{\delta L_{\mathrm{grav}}}{\delta K_{ab}} = \frac{\sqrt{\det h}}{16\pi G}(K^{ab} - K^c_c q^{ab}) \tag{3.17}$$

conjugate to h_{ab} : In this field-theoretical context, we have Poisson brackets

$$\{h_{ab}(x),\, p^{cd}(y)\} = \delta^c_{(a}\delta^d_{b)}\delta(x,y). \tag{3.18}$$

From the action we obtain the Hamiltonian

$$\begin{aligned} H_{\rm grav} &= \int \mathrm{d}^3x \left(\dot{h}_{ab}p^{ab} - L_{\rm grav}\right) \\ &= \int \mathrm{d}^3x \left(\frac{16\pi G N}{\sqrt{\det h}}\left(p_{ab}p^{ab} - \frac{1}{2}(p^c_c)^2\right) + 2p^{ab}D_aN_b - \frac{N\sqrt{\det q}}{16\pi G}{}^{(3)}R\right) \\ &=: \int \mathrm{d}^3x (NC^{\rm grav} + N^aC^{\rm grav}_a). \end{aligned} \tag{3.19}$$

All terms in the Hamiltonian are linear in the lapse function N and the shift vector N^a, but time derivatives of these fields do not appear. As components of the space–time metric, the action is to be extremized with respect to them, too, not just with respect to the spatial metric h_{ab}. While variation with respect to h_{ab} and p^{ab} provides equations of motion due to the canonical piece $\dot{h}_{ab}p^{ab}$, variation by lapse and shift leads to constraints on the phase-space variables: the diffeomorphism constraint

$$C^{\rm grav}_a = -2D_bp^b_a \tag{3.20}$$

and the Hamiltonian constraint

$$C^{\rm grav} = \frac{16\pi G}{\sqrt{\det h}}\left(p_{ab}p^{ab} - \frac{1}{2}(p^a_a)^2\right) - \frac{\sqrt{\det h}}{16\pi G}{}^{(3)}R. \tag{3.21}$$

Matter terms will contribute extra pieces, which is why we denote the pure gravitational terms with the superscript "grav". In vacuum, $C^{\rm grav}_a = 0$ and $C^{\rm grav} = 0$.

If these constraints are solved and objects invariant under the Hamiltonian flow they generate are considered, we are dealing with space–time observables independent of any coordinate choices. As already mentioned, completing such a program is extremely difficult; we thus postpone a discussion of the constraints at the classical level, trying to represent the tensorial objects h_{ab} and p^{ab} as operators. For this, we first need extra structures to get rid of the indices and integrate to scalars, for only those can directly be operators. (Otherwise, an appropriate behavior under complicated tensor transformations in a quantum algebra of non-scalar objects must be ensured, which, as experience shows, is prone to becoming anomalous.) One possibility to remove indices is by contraction with other geometrical objects, not containing the dynamical fields so as to retain linear structures. For instance, one may associate the length $\ell_e[h_{ab}] = \int_e \mathrm{d}t\sqrt{\dot{e}^a\dot{e}^bh_{ab}}$ with any differentiable curve e in space. For any given curve, this is a scalar object not changing under coordinate transformations. But it is not linear in phase-space variables due to the square root.

3.2.2.2 Ashtekar–Barbero Variables

No linear scalar representation of the full phase space of general relativity is known in terms of metric variables. But such a formulation does exist in terms of a new

set, formed by the densitized triad E^a_i together with the Ashtekar–Barbero connection A^i_a [6, 7]. The densitized triad replaces the spatial metric and is defined in several steps: Instead of using the metric we first introduce the co-triad e^i_a via $h_{ab} = e^i_a e^i_b$. (The index i does not refer to the tangent space but simply labels the three co-triad co-vectors. Its position, unlike the one of $a, b, \ldots$, is thus not relevant and we will freely move it up or down as convenient. No spatial metric is required to do so. For repeated indices, as in the defining relation, we will understand the summation convention unless stated otherwise.) A given metric does not uniquely define a co-triad, which can be redefined by any orthogonal transformation $e^i_a \mapsto R^i_j e^j_a$, $R^i_j R^k_i = \delta^k_j$, without changing h_{ab}. There is thus more freedom than in metric formulations, which later on will be removed via additional constraints. From the co-triad, one then obtains the triad e^a_i as its matrix inverse: $e^a_i e^j_a = \delta^j_i$. Equivalently, the triad, as suggested by the notation, is obtained by raising the index of e^i_a using the inverse metric h^{ab}. (The co-triad and the triad form dual orthonormal bases of the co-tangent and tangent space, respectively.) Finally, we densitize the triad by multiplying it with the scalar density $\sqrt{\det h} = |\det(e^i_a)|$:

$$E^a_i := |\det(e^j_b)| e^a_i . \tag{3.22}$$

Notice the absolute value at this stage. Even after factoring out the rotational freedom of a triad, it does have more information than a metric. Unless it is degenerate, the triplet of triad vectors can be left- or right-handed, meaning that the determinant of e^i_a seen as a 3×3-matrix can take both signs. Changing the sign corresponds to a large gauge transformation in O(3)/SO(3) not connected to the unit. It is thus not removed by factoring out the flow generated by a constraint, and remains relevant for geometry. Its meaning is that of the orientation of space, which will become important later in quantum cosmology.

Similarly, we manipulate extrinsic curvature K_{ab} by first contracting with the triad on one index, defining $K^i_a := e^b_i K_{ab}$. This expression turns out to be canonically conjugate to the densitized triad,

$$\{K^i_a(x), E^b_j(y)\} = 8\pi G \delta^b_a \delta^i_j \delta(x, y). \tag{3.23}$$

To define scalar objects to be quantized, it is useful to do one final step and combine extrinsic curvature with the spin connection Γ^i_a that is compatible with the triad: $D_a e^b_i = \partial_a e^b_i + \Gamma^b_{ac} e^c_i - \varepsilon^{ijk} \Gamma^j_a e^b_k = 0$, solved by

$$\Gamma^i_a = -\varepsilon^{ijk} e^b_j \left(\partial_{[a} e^k_{b]} + \frac{1}{2} e^c_k e^l_a \partial_{[c} e^l_{b]} \right). \tag{3.24}$$

We then have the Asthekar–Barbero connection

$$A^i_a = \Gamma^i_a + \gamma K^i_a \tag{3.25}$$

with the Barbero–Immirzi parameter $\gamma > 0$ [8]. This provides the final canonical pair with

$$\{A_a^i(x), E_j^b(y)\} = 8\pi\gamma G\delta_a^b\delta_j^i\delta(x,y), \tag{3.26}$$

a relation which directly follows from (3.23) since Γ_a^i is a functional of the triad and thus Poisson-commutes with it. These canonical variables have a structure analogous to what is known in gauge theories: a connection and a densitized vector field. They are subject to a Gauss constraint

$$D_a^{(A)}E_i^a := \partial_a E_i^a + \varepsilon^{ijk}A_a^j E_k^a = D_a E_i^a + \gamma\varepsilon^{ijk}K_a^j E_k^a = 0 \tag{3.27}$$

of the usual form. (In the last reformulation we used the fact that the covariant divergence of a densitized vector field equals the coordinate divergence. Since the spin connection is compatible with the triad, the Gauss constraint is equivalent to $\varepsilon^{ijk}K_a^j E_k^a = 0$. This relation implies that $K_{ab} = K_a^i e_b^i$ is a symmetric tensor.) Compared to those gauge theories that occur in particle physics, the gravitational diffeomorphism constraint, which now reads

$$C_a^{\text{grav}} = F_{ab}^i E_i^b \tag{3.28}$$

with the Yang–Mills curvature

$$F_{ab}^i = \partial_a A_b^i - \partial_b A_a^i - \varepsilon^{ijk}A_a^j A_b^k \tag{3.29}$$

is new, and the Hamiltonian is quite different from the Yang–Mills form and now a constraint contribution

$$C^{\text{grav}} = \left(\varepsilon^{ijk}F_{ab}^i - 2(1+\gamma^{-2})(A_a^i - \Gamma_a^i)(A_b^j - \Gamma_b^j)\right)\frac{E_j^{[a}E_k^{b]}}{\sqrt{|\det E|}}. \tag{3.30}$$

3.2.2.3 Holonomy-Flux Algebra

Since we are ignoring the constraints for now, their specific form does not play a role; we can thus define and use objects which are well-known from general gauge theories: holonomies

$$h_e[A_a^i] = \mathscr{P}\exp\int_e \dot{e}^a A_a^i \tau_i \mathrm{d}\lambda \tag{3.31}$$

for curves e in space, and fluxes

$$F_S^{(f)}[E_i^a] = \int_S n_a E_i^a f^i \mathrm{d}^2 y \tag{3.32}$$

through surfaces S in space with smearing functions f^i taking values in the internal space. (The co-normal $n_a = \frac{1}{2}\varepsilon_{abc}\varepsilon^{uv}\frac{\partial x^b}{\partial y^u}\frac{\partial x^c}{\partial y^v}$ of $S\colon\ y \mapsto x(y)$ is independent of any metric.)

For our purpose of finding a set of linear scalar quantities, these variables turn out to be immensely useful. Fluxes are already spatial scalars linear in one of the canonical variables, E^a_i. And while holonomies are not linear in A^i_a, they do form a linear algebra together with the fluxes:

$$\{h_e[A], F_S^{(f)}[E]\} = 8\pi\gamma G\eta(e, S)\mathscr{O}^{(f)}_{e,S}(\tau_i h_e[A]) \tag{3.33}$$

with a topological number $\eta(e, S)$ which determines how and how often the curve e and the surface S intersect, and $\mathscr{O}$ denoting a suitable ordering of $\tau_i \in \mathrm{su}(2)$ and $h_e[A] \in \mathrm{SU}(2)$ depending on where on the curve e its intersections with S lie. For instance, if the intersection is at the endpoint of e, we have $\mathscr{O}^{(f)}_{e,S}(\tau_i h_e[A]) = h_e[A]\tau_i f^i(e(1))$, and $\mathscr{O}^{(f)}_{e,S}(\tau_i h_e[A]) = f^i(e(0))\tau_i h_e[A]$ if the intersection is at the starting point. (Unless stated otherwise, we will always assume curves to be defined on the interval [0, 1] of their parameter.) In general, we have

$$\mathscr{O}^{(f)}_{e,S} = \sum_{p\in e\cap S} f^i(p) h_{e\to p}[A]\tau_i h_{p\leftarrow e}[A] \tag{3.34}$$

with $h_{e\to p}$ the holonomy along the starting piece of e up to p, and $h_{p\leftarrow e}$ along the ending piece from p onwards. See also [9] for detailed calculations of the holonomy-flux algebra.

The algebra (3.33) can explicitly be represented as operators on a Hilbert space [10]. All quantities are coordinate independent, and they do not make use of any extra structures except their labels e, S and f. (They certainly make use of standard structures such as topological or differentiable ones of the underlying manifold. But no extra metric, for instance, is introduced which would make the construction background dependent.) The algebra of $F_S^{(f)}[E]$ and $h_e[A]$ replaces the algebra of annihilation and creation operators $\hat{a}_{\boldsymbol{k}}$ and $\hat{a}^\dagger_{\boldsymbol{k}}$ for quantum field theories on a background.

The specific form of the algebra suggests to view holonomies as creation operators, raising the excitation level of fluxes: a state annihilated by some $F_S[E]$ would, after being acted on by a holonomy along a curve intersecting S, have a non-vanishing flux through S due to $\hat{F}_S(\hat{h}_e|\psi\rangle) = \hat{h}_e\hat{F}_S|\psi\rangle + [\hat{F}_S, \hat{h}_e]|\psi\rangle = [\hat{F}_S, \hat{h}_e]|\psi\rangle \neq 0$ with a non-vanishing commutator. What is needed for a construction of all possible excited states in this way is also a state to start with, from which holonomies can then generate new excitations. In common quantum field theories, this state would be the vacuum devoid of particles; here it would be a state where not even fluxes, and thus the densitized triad or spatial metric, would be present. It is a state in which geometry itself is highly quantum and only lowly excited, unlike any classical geometry. Such a state is extremely difficult to imagine physically, but it has a very simple mathematical expression: if we choose the connection representation of states $\psi[A^i_a]$, it is a mere constant. Then indeed, fluxes which would be derivative operators in such a representation, all vanish.

Let us thus define this state as the quantum geometrical "vacuum", $\psi_0(A^i_a) = 1$. Since holonomies only depend on the connection, they will become multiplication

operators, directly showing their action on the state. More precisely, as basic operators we should allow all matrix elements of holonomies, which are in SU(2). Multiplying several ones of them can be tedious due to the group structure, but what is relevant now is already illustrated by the simpler case of holonomies taking values in the Abelian group U(1). (This group would be obtained in a loop quantization of electromagnetism [11].) Then, each holomony is a simple connection-dependent phase factor $h_e[A_a] = \exp(\mathrm{i}\int_e \mathrm{d}\lambda \dot{e}^a A_a)$, and excited states can be written as

$$\psi_{e_1,n_{e_1};\ldots;e_i,n_{e_i}} = \hat{h}_{e_1}^{n_{e_1}} \cdots \hat{h}_{e_i}^{n_{e_i}} \psi_0. \tag{3.35}$$

As functionals, these states look like

$$\psi_{g,n}(A_a) = \prod_{e\in g} h_e(A_a)^{n_e} = \prod_{e\in g} \exp(\mathrm{i}n_e \int_e \mathrm{d}\lambda \dot{e}^a A_a) \tag{3.36}$$

In this notation, each occurrence of a curve e_i in space signals that geometry is excited along that curve: fluxes through surfaces intersected by the curve will be non-zero. Moreover, each curve e_i can be excited several times, as indicated by the integer n_{e_i}. Thus, curves and the integers technically play roles analogous to particle wave numbers and occupation numbers in quantum field theories on a background space–time.

These constructions are used in a more general setting in the diverse models of loop quantum gravity. We now assume that we have a d-dimensional spatial manifold Σ (which in the concrete applications of symmetry-reduced models will be $d < 3$, but formally the dimension could be larger than three). Furthermore, we assume a compact structure group G, and as fields (i) a G-connection A_a^i and a densitized $\mathscr{L}G$-valued vector field E_i^a forming the gauge part of the theory (with E_i^a dual to an $\mathscr{L}G$-valued $(d-1)$-form $\Sigma^i_{a_1\ldots a_{d-1}} = \varepsilon_{a_1\ldots a_d} E_i^{a_d}$), and (ii) scalars $\phi_I : \Sigma \to \mathbb{R}$ with densitized momenta $p^I : \Sigma \to \mathbb{R}$ forming the "matter" part of the theory. In the actual models, the scalars may arise as some of the components of the full gravitational connection, rather than playing the role of physical matter. The fields form canonical variables

$$\Theta_{\text{gauge}} = \frac{1}{\kappa}\int_\Sigma \mathrm{d}^3x \dot{A}_a^i E_i^a \longrightarrow \{A_a^i(x), E_j^b(y)\} = \kappa\delta_j^i\delta_a^b\delta(x,y)$$

$$\Theta_{\text{scalar}} = \int_\Sigma \mathrm{d}^3x \dot{\phi}_I p^I \longrightarrow \{\phi_I(x), p^J(y)\} = \delta_I^J\delta(x,y)$$

with a coupling constant κ, and are subject to certain constraints $C_\alpha[A_a^i, E_j^b, \phi_I, p^J] = 0$. Examples for this general setting of fields are Yang–Mills theory, where (A_a^i, E_j^b) are subject to the Gauss law $\mathscr{G}_i := \partial_a E_i^a + \varepsilon_{ij}{}^k A_a^j E_k^a = 0$ (for $G = \mathrm{SU}(2)$), or general relativity in Ashtekar–Barbero variables. Later chapters will provide a large set of further examples in which scalars arise from symmetry reduction. In this context, we start with a gauge theory (A_a^i, E_i^a) in 3+1 dimensions and impose invariance under some symmetry group S acting on the principal fiber bundle $P \to \Sigma$ that underlies the gauge theory. An example which will be discussed in more detail later is spherically symmetric gravity; see also Sect. 9.1. It turns out that spherically symmetric SU(2)-connections and densitized triads can always be written as

$$\begin{aligned}
{}^{3D}A_a^i\tau_i\mathrm{d}x^a &= A_x(x)\tau_3\mathrm{d}x + A_\varphi\bar{\Lambda}_\varphi^A\mathrm{d}\vartheta + A_\varphi(x)\Lambda_\varphi^A\sin\vartheta\mathrm{d}\varphi + \tau_3\cos\vartheta\mathrm{d}\varphi\\
{}^{3D}E_i^a\tau^i\frac{\partial}{\partial x^a} &= E^x(x)\tau_3\sin\vartheta\frac{\partial}{\partial x} + E^\varphi\bar{\Lambda}_E^\varphi\sin\vartheta\frac{\partial}{\partial\vartheta} + E^\varphi(x)\Lambda_E^\varphi\frac{\partial}{\partial\varphi}
\end{aligned}$$

with a U(1)-connection A_x, a densitized triad E^x, real-valued scalars A_φ, E^φ, and the angles α and β in $\Lambda_\varphi^A = \cos\beta(x)\tau_1 + \sin\beta(x)\tau_2$, $\Lambda_E^\varphi = \cos(\alpha(x)+\beta(x))\tau_1 + \sin(\alpha(x)+\beta(x))\tau_2$. The remaining fields are fixed by the conditions $\mathrm{tr}(\bar{\Lambda}_\varphi^A \Lambda_\varphi^A) = 0 = \mathrm{tr}(\bar{\Lambda}_E^\varphi \Lambda_E^\varphi)$. All free fields depend only on the radial coordinate x, and are thus defined on a 1-dimensional manifold.

Loop quantization then presents a specific way of canonical quantization, turning the Poisson algebra of basic fields into an operator algebra. Any such quantization requires smearing for field theories to remove delta-functions in the elementary Poisson brackets, usually done using a background metric, as in $\int_\Sigma \mathrm{d}^d x \sqrt{\det h}\phi(x)$ for a scalar field. But such a procedure is not suitable if the metric itself (or a densitized triad) is to be quantized: no linear algebra of basic smeared objects would result. The advantage of connection variables is that they provide a natural smearing without having to make use of a fixed metric: holonomies (3.31) along curves e in space, fluxes (3.32), in general through surfaces of $\mathrm{codim}(S, \Sigma) = 1$, scalar values $\phi_I(x)$, and integrated momenta $\int_R p^J(y)\mathrm{d}^d y$ can all be defined without an extra metric, and the integrations they contain remove all delta-functions from their Poisson brackets.

Constructing a Hilbert-space representation leads to states in a space of square-integrable functions $L^2(\bar{\mathscr{A}} \times \bar{\Phi}, \mathrm{d}\mu_{\mathrm{AL}})$ with a compact space $\bar{\mathscr{A}} \times \bar{\Phi}$ of generalized connections and scalars [12]. For (finite analytical) graphs $g \subset \Sigma$ with edge set $E(g)$ and vertex set $V(g)$, we define spaces of g-connections

$$\mathscr{A}_g = \{A_g : E(g) \to G\} \qquad \text{(holonomies along the graph g)}$$

and g-scalars

$$\Phi_g = \{\phi_g : V(g) \to \bar{\mathbb{R}}_{\mathrm{Bohr}}\} \qquad \text{(vertex values on g)}$$

taking values in the Bohr compactification of the real line (see below). For $g \subset g'$, projections $\pi_g^{\mathscr{A}} : \mathscr{A}_{g'} \to \mathscr{A}_g$ and $\pi_g^{\Phi} : \Phi_{g'} \to \Phi_g$ are defined by restriction, and they allow the definition of the full space of generalized connections and scalars by projective limits to arrive at $\bar{\mathscr{A}}$ and $\bar{\Phi}$ as the spaces of fields "on an arbitrarily fine graph". To define generalized scalars we use a certain compactification of the real line, the so-called Bohr compactification $R \subset \bar{\mathbb{R}}_{\mathrm{Bohr}}$. In this way, generalized scalars, just like generalized connections, take values in a compact set. This feature will allow us to provide a consistent definition of the inner product on $\bar{\mathscr{A}} \times \bar{\Phi}$. The Bohr compactification is a topological space such that all continuous functions are the almost-periodic ones:

$$f(\phi) = \sum_{\mu \in \mathscr{I} \subset \mathbb{R}\ \text{countable}} f_\mu \exp(\mathrm{i}\mu\phi)$$

The set of almost-periodic functions forms an Abelian C^*-algebra, and as a consequence the space $\bar{\mathbb{R}}_{\mathrm{Bohr}}$ on which these functions are defined (the Gel'fand spectrum) is compact. The Bohr compactification also inherits an Abelian group structure from $\mathbb{R}$, allowing us to introduce the Haar measure

$$\int_{\bar{\mathbb{R}}_{\mathrm{Bohr}}} \mathrm{d}\mu_{\mathrm{Haar}}(\phi) f(\phi) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}\phi f(\phi).$$

An orthonormal basis with respect to this measure is given by $\{\phi \mapsto \exp(\mathrm{i}\mu\phi) : \mu \in \mathbb{R}\}$; the Hilbert space $L^2(\bar{\mathbb{R}}_{\mathrm{Bohr}}, \mathrm{d}\mu_{\mathrm{Haar}})$ is non-separable. For more information on $\bar{\mathbb{R}}_{\mathrm{Bohr}}$, see Sect. 3.2.3.4. With these constructions we proceed to defining the inner product on our states. We focus on the dense set of cylindrical states: the projection $\pi_g : \bar{\mathscr{A}} \times \bar{\Phi} \to \mathscr{A}_g \times \Phi_g$, obtained by combining $\pi_g^{\mathscr{A}}$ and π_g^{Φ}, lifts any $f_g : \mathscr{A}_g \times \Phi_g \to \mathbb{C}$ to the cylindrical state $\psi = f_g \circ \pi_g$ such that

$$\psi(A,\phi) = f_g(A(e_1),\ldots,A(e_n),\phi(v_1),\ldots,\phi(v_m)).$$

On these states, the inner product is obtained from the Haar measures on G and $\bar{\mathbb{R}}_{\text{Bohr}}$ [13]: If $\psi^{(1)}$ and $\psi^{(2)}$ are cylindrical with respect to the same graph g,

$$\langle\psi^{(1)},\psi^{(2)}\rangle = \int\limits_{G^n\times\bar{\mathbb{R}}^m_{\text{Bohr}}} \prod_{i=1}^{|E(g)|} \mathrm{d}\mu_{\text{Haar}}(h_i) \prod_{j=1}^{|V(g)|} \mathrm{d}\mu_{\text{Haar}}(\phi_j) \times f_g^{(1)}(h_1,\ldots h_n,\phi_1,\ldots,\phi_m)^* f_g^{(2)}(h_1,\ldots h_n,\phi_1,\ldots,\phi_m).$$

If states are not based on the same graph, one can embed both graphs in a larger one by subdivision, or by an extension of the graphs by "dummy" edges without connection dependence. From properties of the Haar measure one quickly sees that states are orthogonal if they are cylindrical with respect to graphs such that there is an edge e for which $\psi^{(1)}$ depends non-trivially on $A(e)$ while $\psi^{(2)}$ does not. On the resulting Hilbert space holonomies act by multiplication, fluxes as derivative operators measuring the excitation level of geometry.

How exactly flux values are increased by excitations of geometry is derived from the action of the flux operator. We already know the states, and a flux, which is linear in the densitized triad, becomes a simple functional derivative operator by the connection. Again in the U(1)-example (with $\{A_a(x), E^b(y)\} = 8\pi\gamma G\delta_a^b\delta(x,y)$),

$$\begin{aligned}\hat{F}_S\psi_{g,n} &= \frac{8\pi\gamma G\hbar}{\mathrm{i}}\int\limits_S \mathrm{d}^2y n_a \frac{\delta\psi_{g,n}}{\delta A_a(y)} = \frac{8\pi\gamma\ell_{\mathrm{P}}^2}{\mathrm{i}}\sum_{e\in g}\int\limits_S \mathrm{d}^2y n_a \frac{\delta h_e}{\delta A_a(y)}\frac{\partial\psi_{g,n}}{\partial h_e}\\ &= 8\pi\gamma\ell_{\mathrm{P}}^2\sum_{e\in g} n_e \int\limits_S \mathrm{d}^2y \int\limits_e \mathrm{d}t n_a\dot{e}^a\delta(y,e(t))h_e\frac{\partial\psi_{g,n}}{\partial h_e} = 8\pi\gamma\ell_{\mathrm{P}}^2\sum_{e\in g} n_e\eta(e,S)\psi_{g,n}\end{aligned} \tag{3.37}$$

with the intersection number $\eta(e,S)$. Since such a state is reproduced after acting with a flux operator, we can directly read off the flux eigenvalues, which are proportional to sums over integers. Thus, the flux spectrum is discrete, providing a detailed realization of discrete spatial geometry [10].

Returning to SU(2)-valued variables, as required for general relativity, we have slightly more complicated expressions for states and operators. Instead of multiplying phase factors, as elements of irreducible U(1)-representations which all are 1-dimensional, we now multiply all possible matrix elements of SU(2)-holonomies along a set of edges. Such states can conveniently be expressed in terms of spin network states [14]

$$\psi_{g,j,C}(A_a^i) = \prod_{v\in g} C_v \prod_{e\in g}\rho_{j_e}(h_e[A]) \tag{3.38}$$

where g is a graph in space, labelled with spins j_e on its edges for irreducible SU(2)-representations ρ_j and with projection matrices C_v in the vertices of the graph which tell us how to pick and combine matrix elements of the holonomies used.

If we consider the example of an n-valent vertex v in which n edges $e_1,\ldots,e_n$ meet and, to be specific, all have the vertex as their endpoint, a suitable projection matrix C_v has n indices

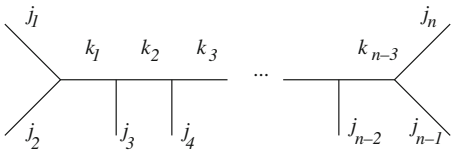

Fig. 3.1 Splitting an n-valent vertex into unique trivalent vertex contractions

such that the incoming holonomies $(h_{e_i})^{A_i}{}_{B_i}$ are multiplied to $C_{v,A_1,\ldots,A_n}\rho_{j_1}(h_{e_1})^{A_1}{}_{B_1}\cdots\rho_{j_n}(h_{e_n})^{A_n}{}_{B_n}$. (The remaining indices B_i will be contracted with projection matrices in the vertices corresponding to starting points of the edges incoming at v.) For a gauge-invariant state, the projection matrices have to satisfy certain conditions. A gauge transformation maps internal vectors v^A at a point to $g^A{}_B v^B$ with $g \in \mathrm{SU}(2)$. Holonomies along a curve $e\colon [0,1] \to \Sigma$ transform as $h_e \mapsto g(e(1))h_e g(e(0))^{-1}$ such that $(h_e v)^A = (h_e)^A{}_B v^B$ transforms as an internal vector at $e(1)$ if v^A is an internal vector at $e(0)$. The spin-network vertex v considered here, with $v = e_i(1)$ for all edges, thus receives a gauge transformation $C_{v,A_1,\ldots,A_n}\rho_{j_1}(g(v))^{A_1}{}_{C_1}\cdots\rho_{j_n}(g(v))^{A_n}{}_{C_n}$ by moving gauge factors from the incoming holonomies to the projection matrix. For the spin network to be gauge invariant, $C_{v,A_1,\ldots,A_n}\rho_{j_1}(g(v))^{A_1}{}_{C_1}\cdots\rho_{j_n}(g(v))^{A_n}{}_{C_n} = C_{v,C_1,\ldots,C_n}$ must hold, which can be realized only if the trivial representation is contained in the tensor product of the ρ_{j_n}. For a trivalent vertex, for instance, a gauge-invariant contraction exists if there is an integer $0 \leq k \leq 2j_1$ such that $j_3 = j_2 - j_1 + k$, where we assume $j_1 \leq j_2$. If this condition is satisfied, there is a unique gauge-invariant contraction. Higher-valent vertices do not have unique contractions. One can parameterize spaces of contraction matrices by integer spins, splitting the n-valent vertex into subsequent trivalent contractions as illustrated in Fig. 3.1. All intermediate spins k_i can take values only in finite ranges, and spaces of contraction matrices are finite dimensional.

Holonomies then act by contributions of new factors, changing some of the labels j_e in an original state by tensor-product decomposition. Fluxes become intersection sums of derivative operators on SU(2), of the well-known angular-momentum form: By analogy with (3.37), functional derivatives by the connection can be written in terms of

$$\hat{J}^i_e = \mathrm{tr}\left((h_e\tau^i)^T\,\partial/\partial h_e\right) \tag{3.39}$$

or using invariant derivative operators on SU(2). Since angular-momentum operators have discrete spectra and we now sum finitely many such contributions over intersections of the graph and a surface, SU(2)-fluxes have discrete spectra, too. (If an edge lies entirely on the surface, thus having infinitely many intersections, the flux vanishes thanks to a product $n_a\dot{e}^a = 0$ in (3.37).) Also the action of holonomies shows a key feature: While holonomy operators are well-defined, one cannot extract a connection operator from them. Trying to do so, for instance by applying the classical identity

$$\dot{e}^a(p)A^i_a(p)\tau_i = \lim_{t\to 0}\frac{\mathrm{d}h_e|_{[0,t]}}{\mathrm{d}t}$$

in the limit where e approaches $e(0) = p$, fails because $h_e|_{[0,t_1]}\psi$ and $h_e|_{[0,t_2]}\psi$ are orthogonal for $t_1 \neq t_2$: they are cylindrical with respect to different graphs.

From the elementary fluxes one can construct more familiar geometrical objects [15–17], such as the area $A_S[E] = \int_S \mathrm{d}^2 y \sqrt{n_a E^a_i n_b E^b_i}$ of a surface S or the volume $V_R[E] = \int_R \mathrm{d}^3 x \sqrt{|\det E|}$ of a region R. Area eigenvalues, just like fluxes, depend on spin labels on curves in the graph intersecting the surface; volume eigenvalues depend on the contraction in vertices within the region. Also these operators have discrete spectra, which in the case of area follow easily from the square of derivative operators. For volume, the spectrum is more difficult to compute since the determinant of E^a_i involves products of three factors of triad components, resulting in couplings of different SU(2)-representations. Nevertheless, recoupling theory allows one to derive matrix elements [18, 19], and powerful computer codes now exist to analyze the eigenvalues [20, 21]. This is expected to be of particular importance for quantum cosmology since the volume spectrum can show how a discrete growing universe must refine its structure as it expands. We will come back to refinement in more detail once we have introduced the dynamics.

3.2.3 Isotropic Models

Many constructions characteristic of canonical quantum gravity can conveniently be illustrated and explicitly be evaluated in symmetric models. The simplest case is that of isotropy, where the spatial geometry is determined by a single number: the scale factor a. Quantum cosmology has traditionally been formulated in this context, and also much work in loop quantum cosmology has been done in an isotropic setting.

3.2.3.1 Symmetry Reduction

Classical symmetry reduction is performed by restricting fields to those left invariant by a set of symmetries (possibly up to gauge transformations only). For invariant connections, for instance, one is looking for the general form of 1-forms ω on a principal fiber bundle $P = (\Sigma, G, \pi)$ with structure group G and base manifold Σ such that $s^*\omega = \omega$ for any element $s \in S$ of a symmetry group S acting on P.

This general definition has two important consequences:

1. An action on the principal fiber bundle P is required, while one usually starts with a desired symmetry on the base manifold, such as isotropy on the space Σ. Lifts of the symmetry action to the whole bundle must thus be found, which are often non-unique. (They are classified by conjugacy classes of homomorphisms $\lambda : F \to G$, where F is the isotropy subgroup of S and G the structure group.) This lifting procedure gives rise to different inequivalent classes of invariant connections, which in physical terms can be classified by topological charges. (An example is magnetic charge as the quantity characterizing topologically inequivalent spherically symmetric U(1)-connections; see the following example.)

2. In terms of local connection 1-forms, invariance implies that a connection may change under a symmetry transformation, but only by a gauge transformation: $s^*A = g(s)^{-1}Ag(s)+g(s)^{-1}\mathrm{d}g(s)$ where $g\colon S \to G$ is a mapping between the symmetry and structure groups. This mapping is not a group homomorphism, but must satisfy certain other conditions. Solving these conditions is equivalent to determining the possible lifts of symmetry actions to the bundle. In the bundle language, invariant connections are given by (i) a connection $A_{S/F}$ on a reduced bundle Q over Σ/S whose structure group $Z_G(\lambda(F))$ is the centralizer of $\lambda(F)$ in G, and (ii) scalar fields $\phi\colon Q \times \mathscr{L}F_\perp \to \mathscr{L}G$ subject to $\phi(\mathrm{Ad}_f X) = \mathrm{Ad}_{\lambda(f)}(\phi(X))$ for $f \in F$, $X \in \mathscr{L}S$. For more details of the bundle formulation, see [5].

Example 3.1 (Magnetic charge) Magnetic monopoles are spherically symmetric configurations of the magnetic field, which can be described by a U(1)-connection A_a, the vector potential. We are thus interested in the general form of spherically symmetric connections on U(1)-principal fiber bundles, on which the symmetry group SU(2) is acting. In general, inequivalent lifts of a symmetry group S from the base manifold, where its action is given and may have an isotropy subgroup F, to a principal fiber bundle with structure group G over the same base manifold, are classified by conjugacy classes of group homomorphisms $\lambda\colon F \to G$. They describe the twisting along fibers when the symmetry action is lifted from the base manifold to the bundle. In this example, we have $F \cong \mathrm{U}(1) \cong G$, all conjugacy classes are labelled by an integer $k \in \mathbb{Z}$ and can be represented as $\lambda_k(\exp(t\tau_3)) = \exp(\mathrm{i}kt)$. For every k, invariant U(1)-connections must have the form of an arbitrary radial U(1)-connection $A_r\mathrm{d}r$ plus a contribution of $\Lambda\colon \mathscr{L}F \to \mathscr{L}G$ with $\Lambda|_{\mathscr{L}F_\perp} = 0$ for U(1), $\Lambda|_{\mathscr{L}F} = \mathrm{d}\lambda_k\colon \tau_3 \mapsto \mathrm{i}k$ applied to the pull-back of the Maurer–Cartan form on S under an embedding of S/F in S. For spherical symmetry with $S = \mathrm{SU}(2)$ and $F = \mathrm{U}(1)$, this pulled-back form can be expressed as

$$A_{S/F} = (\tau_2 \sin\vartheta + \tau_3 \cos\vartheta)\mathrm{d}\varphi + \tau_1 \mathrm{d}\vartheta. \tag{3.40}$$

With these ingredients, we obtain generic spherically symmetric U(1)-connections of the form

$$A = A_r\mathrm{d}r + k\cos\vartheta\,\mathrm{d}\varphi.$$

We have a radial (densitized) magnetic field with $B^r = -k\sin\vartheta$, which implies a magnetic charge

$$Q = \frac{1}{4\pi}\int_{S^2} B^a n_a \mathrm{d}\vartheta\,\mathrm{d}\varphi = -k.$$

For a reduction of the full phase space, one must determine invariant forms of densitized vector fields as well. Once the form of invariant connections is known, for which a richer basis of mathematical results exists, the general form of invariant fields

can uniquely be read off from the symplectic structure they must imply. For every free field A_I in the symmetric form of connections A^i_a, there is a conjugate field E^I in the invariant form of densitized triads E^a_i such that $\int_\Sigma \mathrm{d}^3x \dot{A}^i_a E^a_i = \int_{\Sigma/S} \dot{A}_I E^I$ when evaluated on invariant fields. For gravitational variables one must take into account an extra condition to ensure that momenta of A^i_a can be non-degenerate. This condition in most of the standard cases leads to a unique sector with no topological charge. Examples will be provided in Sect. 9.1 and the next section.

3.2.3.2 Isotropic Configurations

For isotropic connections invariant under arbitrary translations $\mathbb{R}^3$ and rotations, combining to the Euclidean group, one can always choose a gauge where they take the form

$$A^i_a = \tilde{c}\delta^i_a. \tag{3.41}$$

The single component $\tilde{c}$ is spatially constant but for general solutions to the equations of motion depends on time. A densitized triad of the same symmetry type is of the form

$$E^a_i = \tilde{p}\delta^a_i. \tag{3.42}$$

The reduced symplectic potential

$$\frac{1}{8\pi\gamma G}\int_R \mathrm{d}^3x E^a_i \delta A^i_a = \frac{3\mathcal{V}}{8\pi\gamma G}\tilde{p}\delta\tilde{c}, \tag{3.43}$$

where δ denotes a derivative on phase space as opposed to space Σ, then shows that $\tilde{c}$ and $\tilde{p}$ form a canonical pair,

$$\{\tilde{c}, \tilde{p}\} = \frac{8\pi\gamma G}{3\mathcal{V}}. \tag{3.44}$$

In this derivation, as before, we have selected a bounded region $R \subset \Sigma$ to make the spatial integration of homogeneous fields well-defined. If we are considering a cosmological model with compact spatial manifolds, we could choose $R = \Sigma$, but this is not possible for unbounded spatial manifolds. And even for compact spaces, we may as well choose a smaller region as long as it is non-empty; no information about homogeneous configurations is lost provided we just know them in an arbitrarily small neighborhood. The definition of our variables then depends on the choice of the region, and its coordinate size $\mathcal{V} = \int_R \mathrm{d}^3x$. Physical results, of course, must be independent of the choice.

In metric variables, isotropic models are formulated on the phase space with coordinates (a, p_a); triad variables differ from this description by a canonical transformation as well as an extension of the configuration space. From the general relationships

between (A_a^i, E_j^b) and metric variables, one can directly derive the relation between $(\tilde{c}, \tilde{p})$ and the scale factor. For the triad component, we obtain $|\tilde{p}| = \frac{1}{4}\tilde{a}^2$, where the factor of $1/4$ can be seen to arise from matching variables of a closed model, not subject to arbitrary rescaling freedom of coordinates, at a fixed curvature scale [5, 22]. Computing the spin connection and extrinsic curvature for an isotropic metric, we combine them to obtain $\tilde{c} = \frac{1}{2}(k+\gamma\dot{a})$. For flat models, one may rescale the scale factor $\tilde{a}$ so as to eliminate the factor of $1/4$ in $\tilde{p}$, as often done. We will denote the rescaled parameter as $a = \frac{1}{2}\tilde{a}$, such that $|\tilde{p}| = a^2$, $\tilde{c} = \frac{1}{2}k+\gamma\dot{a}$. When a is rescaled, also coordinates and thus $\mathcal{V}$ are rescaled such that $a^3\mathcal{V}$ remains unchanged. Taking this into account, the Poisson-bracket relation (3.44) remains unchanged under any rescaling.

Before we describe possible quantizations of these variables, turning the Poisson bracket (3.44) into a commutator relationship, we should properly deal with the factor of $\mathcal{V}$. It merely multiplies the constant result for the Poisson bracket, but it is coordinate dependent. No such factors can be represented on a Hilbert space, which is defined independently of any coordinates chosen on space. We thus redefine our basic variables to absorb $\mathcal{V}$:

$$c := \mathcal{V}^{1/3}\tilde{c}, \qquad p := \mathcal{V}^{2/3}\tilde{p}. \tag{3.45}$$

The particular powers of $\mathcal{V}$ will turn out to be suitable later on in the context of a loop quantization. Moreover, they make the basic variables coordinate independent since $\tilde{p}$ and $\mathcal{V}^{2/3}$ change exactly in opposite ways when coordinates are rescaled, leaving the product invariant. Our new basic variables (c, p), being coordinate independent, should thus be representable on a Hilbert space. They do, however, depend on the size of the region R chosen which affects $\mathcal{V}$ but not $\tilde{p}$ or $\tilde{c}$. Care is then still needed in interpretations of our quantizations once they are formulated. In particular, although there is no explicit $\mathcal{V}$-dependence in the symplectic form

$$\Omega = \frac{3}{8\pi\gamma G}\mathrm{d}c \wedge \mathrm{d}p, \tag{3.46}$$

it must be rescaled proportionally by λ if the region R is enlarged to change $\mathcal{V}$ to $\lambda\mathcal{V}$. This rescaling will require a corresponding transformation on the resulting Hilbert-space representation.

3.2.3.3 Quantum Representation

Given just a pair of canonical variables allowed to take all real values, one possible quantum representation is a standard Schrödinger one as in quantum mechanics. Following this procedure will essentially be a Wheeler–DeWitt quantization, where we may choose either the triad representation with wave functions $\psi(p)$ or the connection representation with wave functions $\psi(c)$, required to be square integrable to make up a kinematical Hilbert space. Compared to the earlier choice of wave

functions $\psi(a)$ in Chap. 2, this formulation has two minor advantages: (i) we use one of the coordinate-independent (but integration-region dependent) variables, and (ii) any of the two basic variables takes the full real line as its range, such that no boundary conditions are required, in contrast to $a > 0$ which can make adjointness conditions of operators difficult.

But this representation cannot be the final one since we know that a full quantization in inhomogeneous situations does not allow quantum representations of connection components directly, but only of their holonomies. If an isotropic model is to grasp any of these characteristic features, it should be based on variables analogous to holonomies. For an isotropic connection, it suffices to consider segments of straight lines (along generators of the homogeneity group). Only the length of the segment matters, but not its position or direction. A single parameter ℓ_0 can then be used to label all such straight-edged holonomies. For curves along integral vector fields with tangent $\dot{e}^a = X^a$, normalized with respect to a metric δ_{ab} on the homogeneous space, we have holonomies

$$\mathscr{P}\exp\int_e \tilde{c}\delta^i_a X^a \tau_i = \exp(\ell_0 \tau_i X^i \tilde{c}) = \cos\left(\tfrac{1}{2}\ell_0\tilde{c}\right) + 2\tau_i X^i \sin\left(\tfrac{1}{2}\ell_0\tilde{c}\right). \tag{3.47}$$

Since $\tilde{c}$ (in contrast to X^i) is the only dynamical variable, we can express all relevant functions by the U(1)-holonomies $h_{\ell_0}(c) = \exp(\frac{1}{2}\mathrm{i}\ell_0\tilde{c}) = \exp(\frac{1}{2}\mathrm{i}\ell_0 c/\mathscr{V}^{1/3})$, where the length parameter first multiplies the original connection component $\tilde{c}$, which is then expressed in terms of the new c. Similarly, fluxes are integrated densitized triads, which for an isotropic configuration and a square surface of edge length given by the same parameter ℓ_0 is of the form $F_{\ell_0}(p) = \ell_0^2\tilde{p} = \ell_0^2 p/\mathscr{V}^{2/3}$. As is clear from the definitions, all these quantities are independent of coordinates, and they are independent of the region of size $\mathscr{V}$ chosen. In addition to the classical geometry given by a phase-space point $(\tilde{c}, \tilde{p})$, they only depend on the label ℓ_0 which alone now plays the role of all the curves and surfaces used in the full representation.

We could have chosen different parameters as labels for holonomies and fluxes at different places, instead of a single ℓ_0. The reason for not doing so is the understanding that an isotropic quantum configuration should require a rather regular graph, made of straight edges roughly of the same length ℓ_0. Such a graph also provides natural choices for similar-sized square surfaces filling a whole plane, such that each of them is transversal to one edge and intersects other surfaces at most at their boundaries. Regularity then requires all these surfaces to have the size ℓ_0^2, which we have used above.

In this picture, we have the added benefit of bringing in the number of discrete sites $\mathscr{N}$ in a natural way, allowing us to incorporate the Second Principle of Sect. 3.1: If there are $\mathscr{N}$ sites of linear dimension ℓ_0 in a region of size $\mathscr{V}$, then $\mathscr{N} = \mathscr{V}/\ell_0^3$ (Fig. 3.2). The geometrical size of each discrete site as measured by the metric to be quantized, moreover, is $\ell_0^3\tilde{p}^{3/2} = \ell_0^3 a^3$, which we can identify with the elementary size v in the refinement picture used before. Indeed, with the relations written here we identically satisfy (2.16). As we will see in the next chapter about dynamics, the

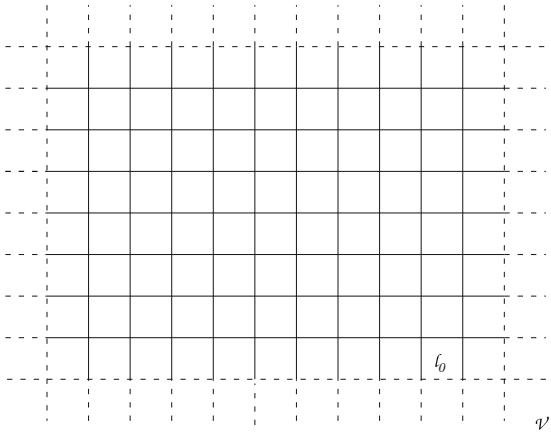

Fig. 3.2 A region of size $\mathcal{V}$, built from patches of linear size ℓ_0

parameter ℓ_0, once it is allowed to become phase-space dependent, is in fact directly related to refinement.

We thus choose the 1-parameter family $\{\exp(\frac{1}{2}\mathrm{i}\mu c), p\}|_{\mu\in\mathbb{R}}$ as our basic set of objects to construct the quantum representation. Here, we have defined $\mu = \ell_0/\mathcal{V}^{1/3}$, and have dropped the μ^2-factor of p since it would just be a multiplicative constant. The specific form of μ and its relationship with ℓ_0 and $\mathcal{V}$ are not relevant for basic operators, for which we can treat μ simply as a real-valued parameter. But the relationship to underlying discrete structures will become important for composite operators such as inverse triads or the Hamiltonian constraint.

As in the full theory, we construct the state space by starting from a basic state ψ_0, given in the connection representation by a mere constant, and generate all other states by multiplying with "holonomies" as creation operators. The result is a space of states all having the general form

$$\psi(c) = \sum_{\mu\in\mathscr{I}} \psi_\mu \exp\left(\tfrac{1}{2}\mathrm{i}\mu c\right) \tag{3.48}$$

where $\mathscr{I} \subset \mathbb{R}$ is a countable index set. As already encountered in Sect. 3.2.2.3, all these functions are called almost periodic, forming a subset of all continuous functions on the real line. Since the space of these functions forms a C^*-algebra, there is a compact space such that almost-periodic functions give the set of *all* continuous functions on that space. This space is compact because its set of continuous functions is a unital C^*-algebra. It contains the real line because almost-periodic functions are functions of a real variable. The real line allows many more continuous functions, and so the space on which almost-periodic functions are all the continuous ones must be larger, with additional points making continuity conditions more restrictive. This larger space contains the real line as a dense subset; it is called the Bohr compactification $\overline{\mathbb{R}}_{\mathrm{Bohr}}$ of $\mathbb{R}$.

Having based isotropic loop quantization on the space of almost-periodic functions, the quantum configuration space will be the Bohr compactification $\overline{\mathbb{R}}_{\mathrm{Bohr}}$ rather

than the real line itself. The usual integration on $\mathbb{R}$ also extends to $\overline{\mathbb{R}}_{\text{Bohr}}$, which is an Abelian group and thus has a unique Haar measure up to a constant factor. It can be written explicitly as

$$\int_{\overline{\mathbb{R}}_{\text{Bohr}}} \mathrm{d}\mu \, \bar{f}(c) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}c f(c) \tag{3.49}$$

where for any continuous function $\bar{f}$ on $\overline{\mathbb{R}}_{\text{Bohr}}$, f is its restriction to the dense subset $\mathbb{R}$ on which the ordinary integration measure is used.

As usual, holonomies then act on states simply by multiplication. We pick a basis given by the uncountable set $\{|\mu\rangle\}_{\mu\in\mathbb{R}}$ where

$$\langle c|\mu\rangle = \exp(\tfrac{1}{2}\mathrm{i}\mu c) \tag{3.50}$$

in the connection representation. It is clear that these states span all states (3.48), and with the inner product based on (3.49) they are orthonormal. On the basis, holonomies act by shifting the labels:

$$\widehat{\exp(\tfrac{1}{2}\mathrm{i}\delta c)}|\mu\rangle = |\mu+\delta\rangle. \tag{3.51}$$

To compare with basic holonomy operators in the full theory, one can think of this action as being analogous to changing the spin of an SU(2)-representation by coupling the edge spin of a spin network state with the spin of the holonomy used for the operator; or one can think of it as a holonomy operator extending an already existing edge, thus making the length parameter larger. In an isotropic context, these two interpretations (or any mixtures thereof) cannot be separated—a degeneracy which has to be taken into account for proper interpretations of operator actions.

The flux operator $\hat{p}$ can be expressed directly as a derivative operator

$$\hat{p} = \frac{8\pi\gamma\ell_{\text{P}}^2}{3\mathrm{i}} \frac{\mathrm{d}}{\mathrm{d}c} \tag{3.52}$$

taking into account the factor in the symplectic structure (3.46) and introducing the Planck length $\ell_{\text{P}} = \sqrt{G\hbar}$. Its action on the basis states follows directly as

$$\hat{p}|\mu\rangle = \frac{4\pi\gamma\ell_{\text{P}}^2}{3}\mu|\mu\rangle \tag{3.53}$$

which shows that we picked the flux eigenbasis with (3.50). Just as in the full case, (3.37), the flux operator has a discrete spectrum: all its eigenstates are normalizable. Unlike with the full spectrum, however, every real number is an eigenvalue. These two properties are consistent with each other in the present case of a non-separable Hilbert space. As we will see later, the mathematical definition of a discrete spectrum via the normalizability of eigenstates turns out to be the appropriate one here, too,

because it shows that crucial features of the full theory are realized in isotropic models as well. Also our isotropic quantum geometry is thus atomic in the sense of discrete flux spectra.

Now having the basic representation of the isotropic holonomy-flux algebra at our disposal, we can analyze its rescaling behavior. We had absorbed factors of $\mathcal{V}$ in the original canonical variables $\tilde{c}$ and $\tilde{p}$ in order to deal with coordinate-invariant quantities. But then the variables c and p, as well as their resulting quantum theory, become dependent on the volume $\mathcal{V}$ of a region chosen arbitrarily. Changing the region must result in a well-defined transformation between the quantum representations obtained for different values of $\mathcal{V}$; otherwise there would be no way of eventually testing whether observables are indendent of the rescaling.

To analyze this, let us rescale $\mathcal{V}$ to $\lambda\mathcal{V}$ with a positive real number λ. Then, c and p change to $\lambda^{1/3}c$ and $\lambda^{2/3}p$, respectively. Also the symplectic structure changes to $\lambda\Omega$ which indeed follows from the classical rescaling. Thus, the rescaling transformation is not canonical. Since the symplectic structure of basic variables is rescaled, also the quantum representation must change: commutators of basic operators, following from the basic Poisson brackets, are to be divided by λ. Instead of modifying commutators, which are universally defined, we can formally implement this rescaling of the representation by changing Planck's constant to $\lambda\hbar$ in all equations where it enters. (Of course, the physical value of $\hbar$ is fixed. The scaling can be normalized to the correct value only if the underlying discreteness scale of quantum gravity is taken into account. Such a scale is not explicitly available in minisuperspace models, but will be included at a later stage once we discuss inhomogeneity; Sect. 10.1.) This indeed provides the correct rescaling relationship between c and p. For instance, in (3.52) we have $\mathrm{d}/\mathrm{d}c \mapsto \lambda^{-1/3}\mathrm{d}/\mathrm{d}c$, while $\ell_{\mathrm{P}}^2 \mapsto \lambda\ell_{\mathrm{P}}^2$, combining to the correct $\hat{p} \mapsto \lambda^{2/3}\hat{p}$. From (3.53) we then read off that the state parameter μ must rescale to $\lambda^{-1/3}\mu$. (In particular, μ rescales like the classical c^{-1}, not like the flux p whose eigenvalues it provides after multiplying it with $4\pi\gamma\ell_{\mathrm{P}}^2/3$.) Also this is consistent, for (3.51) tells us that μ and the holonomy coefficient $\delta \propto \ell_0/\mathcal{V}^{1/3}$ must rescale in the same way for $|\mu+\delta\rangle$ to have a meaningful rescaling.

In quantum cosmology we are thus dealing with a family of models $(\hat{c}, \hat{p}, [\cdot,\cdot])_{\mathcal{V}}$. Classically, changing $\mathcal{V}$ is not a canonical transformation; in quantum theory there is no unitary relationship between models of different $\mathcal{V}$ with the same Hilbert-space representation. Still, there are simple means to check the $\mathcal{V}$-independence of results, as we will use them in what follows.

As a final remark, we notice that we could have decided to fix ℓ_0 once and for all, as it was indeed done in the first formulations of loop quantum cosmology [23, 24]. There is a disadvantage of implying, essentially, that one makes the configuration space of connections strictly periodic. From holonomies of this type, we could not reconstruct $\tilde{c} \in \mathbb{R}$ completely, but only up to integer multiples of $4\pi/\ell_0$. Almost-periodic functions, on the other hand, do not require a periodification of the configuration space but rather compactify it by an enlargement. At the kinematical level, no freedom is lost by using the Bohr compactification. Still, it remains to be seen at this stage what the dynamics actually requires: it will also rely on holonomies as basic building blocks of the Hamiltonian; and if there is a choice of ℓ_0 to be made there, as it will turn out to be indeed the case, one could have made that restriction already at the kinematical level. Dynamics, in this case, would see only one sector of periodic

connections, and nothing would be lost by fixing attention to one periodic sector from the outset. To allow the full freedom of how this might turn out, we do not fix ℓ_0 for now, permitting all $\exp(\frac{1}{2}\mathrm{i}\ell_0\tilde{c})$ as operators. (In fact, lattice refinement especially for anisotropic models will require this general attitude, as we will see later.) Quantum dynamics will then be based on the Bohr compactification of the real line as the configuration space. Since we already indicated the relationship between ℓ_0 and lattice refinement, this issue will naturally be revisited in the chapter on dynamics.

3.2.3.4 More on Bohr

The Bohr compactification of the real line plays an important role in loop quantum cosmology to include all kinematical sectors of the models in one non-separable Hilbert space. The dynamics sometimes picks a separable sector, but since the one that is realized depends on the specific form of the dynamics and is subject to quantization ambiguities, as we will see, it is useful to consider all quantum configurations in one setting even though this information is only kinematical. In this context, several properties and characterizations of the Bohr compactification are of interest. For additional mathematical discussions, see [25, 26].

- The Bohr compactification of the real line contains $\mathbb{R}$ densely. It is thus not a periodification which would identify different points in $\mathbb{R}$ and not contain all the original ones. It is also different from the one-point compactification, which, too, contains the real line densely but adds just one point at infinity. The difference can be seen by the set of continuous functions on these two compact spaces. For the Bohr compactification, as used in the definition, this is the set of almost-periodic functions; for the one-point compactifications, this is the set of functions $f(c)$ for which $\lim_{c\to-\infty} f(c)$ and $\lim_{c\to\infty} f(c)$ exist and agree. The only functions continuous on the Bohr compactification as well as the one-point compactification are the constant ones.
- One can visualize the Bohr compactification by subsets in a torus $[0, 1]^2$. The whole real line can be embedded in the torus as a straight line $x = \omega y$ if ω is irrational. In the trace topology, the resulting subset of the compact torus can be completed to a compact space containing the real line (the original embedding) densely.
- The Fourier space of the Bohr compactification is the discrete real line whose open neighborhoods are arbitrary unions of single points. This is indeed the space of momenta as it arises from (3.48) with (3.53).
- If we contrast a periodification with the Bohr compactification, the configuration space in one case is a circle with Fourier space given by the integers $\mathbb{Z}$. The enlargement of the Bohr compactification as compared to the periodification allows more momenta filling all of $\mathbb{R}$. As a trace of the compactness of the configuration space, the momentum space is the real line with discrete topology.
- A Wigner function for states supported on $\bar{\mathbb{R}}_{\mathrm{Bohr}}$ can be defined as [25]

$$W(c,p) = \frac{1}{2\pi\ell_{\mathrm{P}}^2} \int_{\bar{\mathbb{R}}_{\mathrm{Bohr}}} \psi^*\left(p - \tfrac{1}{2}\alpha\right) \psi\left(p + \tfrac{1}{2}\alpha\right) h^{(2\alpha/\ell_{\mathrm{P}}^2)}(c) \mathrm{d}\alpha \tag{3.54}$$

with holonomies $h^{(\delta)}(c) = \exp(\mathrm{i}\delta c/2)$. The Wigner function of a triad eigenstate, for instance, is then a delta-function peaked at the triad eigenvalue, and independent of c. Further properties of the Bohr-Wigner function are discussed in [25].

3.2.3.5 Inverse-Triad Operators

For a first glimpse on the singularity issue we now have a look at suitable quantizations of a^{-1} or any other inverse power, which would diverge at the classical singularity of isotropic cosmology. In a Wheeler–DeWitt quantization, a^{-1} can easily be quantized as a multiplication operator acting on wave functions $\psi(a)$. It is densely defined and thus suitable. It certainly fails to be a bounded operator, but so does $\hat{a}$ itself. Kinematically, the classical singularity does not appear to be different in Wheeler–DeWitt quantum cosmology.

At first, the problem seems worse in loop quantum cosmology: we would now have to find an inverse of the triad operator $\hat{p}$ which has a discrete spectrum containing zero. No such operator has a densely defined inverse. One could define an inverse using multiplication with μ^{-1} on all states except $|0\rangle$, but since $|0\rangle$ is orthogonal to all states $|\mu\rangle$ with $\mu \neq 0$,[1] this procedure would not make the inverse densely defined. It thus does not correspond to an operator on the Hilbert space. This issue would not just be a problem indicating a singularity, it would even prevent us from quantizing Hamiltonians, including the gravitational one (3.30), which all contain some form of inverse-triad components.

Nevertheless, operators whose classical limit is a^{-1} do exist. To see this, we follow a construction which is also available in the full theory [27, 28] and which in the next chapter will allow us to quantize Hamiltonian constraints and matter Hamiltonians. Applied in an isotropic setting [29], we have the identity

$$\frac{2\mathrm{i}}{\delta} e^{\mathrm{i}\delta c/2} \{e^{-\mathrm{i}\delta c/2}, |p|^{r/2}\} = \{c, |p|^{r/2}\} = \frac{4\pi\gamma G r}{3} |p|^{r/2-1} \operatorname{sgn} p \tag{3.55}$$

If we choose the real parameter r to be in the range $0 < r < 2$, we have an inverse power of the triad component p on the right-hand side, while no inverse is required on the left. The left-hand side of this equation can easily be quantized by using the volume operator $\hat{V} = \hat{p}^{3/2}$, holonomy operators for $\exp(\mathrm{i}\delta c/2)$ and turning the Poisson bracket into a commutator divided by $\mathrm{i}\hbar$. A densely defined operator with an inverse of p as the classical limit results.

To be closer to operators of the full form we first replace the exponentials by SU(2)-holonomies evaluated in isotropic connections:

[1] Assuming a sequence of states in $|\psi_n\rangle \in \mathscr{H}/(\mathbb{C}|0\rangle)$ such that $\lim_{n\to\infty} |\psi_n\rangle = |0\rangle$, we obtain the contradiction $0 = \lim_{n\to\infty} \langle 0|\psi_n\rangle = \langle 0| \lim_{n\to\infty} \psi_n\rangle = 1$.

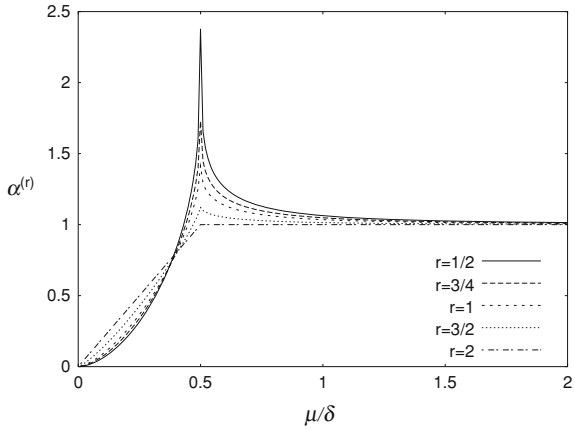

Fig. 3.3 Correction function (3.59) from inverse-triad corrections

$$\mathrm{tr}(\tau_3 \exp(\delta c\tau_3)[\exp(-\delta c\tau_3), \hat{V}^r]) = \sin(\delta c/2)\hat{V}^r \cos(\delta c/2) - \cos(\delta c/2)\hat{V}^r \sin(\delta c/2). \tag{3.56}$$

(The trace can be evaluated explicitly after inserting $\exp(A\tau_3) = \cos\left(\frac{1}{2}A\right) + 2\tau_3 \sin\left(\frac{1}{2}A\right)$.) Using the basic operators, it is easy to see that the resulting

$$\widehat{|p|^{r/2-1}\mathrm{sgn}(p)} = \frac{3}{4\pi\gamma\ell_{\mathrm{P}}^2\delta r}\left(\widehat{e^{\mathrm{i}\delta c/2}}\widehat{[e^{-\mathrm{i}\delta c/2}}, \widehat{|p|^{r/2}}] - \widehat{e^{-\mathrm{i}\delta c/2}}\widehat{[e^{\mathrm{i}\delta c/2}}, \widehat{|p|^{r/2}}]\right) \tag{3.57}$$

has the same eigenbasis $|\mu\rangle$ as the triad operator, with eigenvalues

$$\left(\widehat{|p|^{r/2-1}\mathrm{sgn}(p)}\right)_\mu = \frac{1}{\delta r}\left(\frac{4\pi\gamma\ell_{\mathrm{P}}^2}{3}\right)^{r/2-1}(|\mu+\delta/2|^{r/2} - |\mu-\delta/2|^{r/2}) \tag{3.58}$$

clearly well-defined even at $\mu = 0$. At $\mu = 0$, in fact, the eigenvalue vanishes instead of being divergent like the classical value. For large $|\mu| \gg \delta$, on the other hand, the classical expression $\mathrm{sgn}(\mu)|4\pi\gamma\ell_{\mathrm{P}}^2\mu/3|^{r/2-1}$ is approached. The difference gives rise to correction functions

$$\begin{aligned}\alpha^{(r)}(\mu) = \frac{\left(\widehat{p^{r/2-1}\mathrm{sgn}p}\right)_\mu}{p_\mu^{r/2-1}\mathrm{sgn}p_\mu} &= \left(\widehat{|p|^{r/2-1}\mathrm{sgn}(p)}\right)_\mu \left|\frac{4\pi\gamma\ell_{\mathrm{P}}^2}{3}\mu\right|^{1-r/2}\mathrm{sgn}(\mu) \\ &= \frac{1}{\delta r}|\mu|^{1-r/2}\left(|\mu+\delta/2|^{r/2} - |\mu-\delta/2|^{r/2}\right)\mathrm{sgn}(\mu) \neq 1\end{aligned} \tag{3.59}$$

in quantizations of expressions that classically contain inverses of densitized-triad components; see Fig. 3.3. (The δ-dependence is not explicitly noted as an ambiguity in $\alpha^{(r)}$ because it simply rescales μ.)

This calculation demonstrates that densely defined operators with the classical limit of an inverse power of p do exist; we will later use these constructions for

Hamiltonians. The classical divergence at the singularity implies that these operators cannot be inverse operators of $\hat{p}$:

$$\widehat{p^{r/2-1}}\,\hat{p}^{1-r/2} \neq 1;$$

they only approach the inverse in the classical limit. In this way, the initial problem of $\hat{p}$ having zero in its discrete spectrum is overcome.

Instead of being singular, the small-μ behavior is bounded and approaches zero at $\mu = 0$ as already seen. Around $\mu \sim \delta/2$, a peak is reached demarkating the strong quantum-geometrical behavior for small μ and the nearly classical behavior for large values. The position of the peak is not unique, but depends on quantization ambiguities. For instance, one can use different values of δ and also different r in the specified range without changing the crucial properties. As we will see later, the precise form of the function enters cosmological and other equations, such that ambiguities can in principle be fixed by phenomenology. The freedom is also reduced by considering the anomaly problem of quantum constraints, where inverse-triad operators enter, too; see Sect. 10.3.

For phenomenology, it will be important to consider the typical size of deviations of inverse-triad operators from the classical expectation. In an effective formulation, we would refer not to eigenvalues μ as in (3.59) but to an effective geometry reconstructed from μ via $p_\mu = 4\pi\gamma\ell_{\mathrm{P}}^2\mu/3$. This relationship leads to correction functions

$$\begin{aligned}\alpha^{(r)}(p_\mu) = \frac{1}{\delta r}\left(\frac{4\pi\gamma\ell_{\mathrm{P}}^2}{3}\right)^{-1} |p_\mu|^{1-r/2}\mathrm{sgn}(p_\mu)\\ \times\left(|p_\mu + 2\pi\gamma\delta\ell_{\mathrm{P}}^2/3|^{r/2} - |p_\mu - 2\pi\gamma\delta\ell_{\mathrm{P}}^2/3|^{r/2}\right)\end{aligned} \tag{3.60}$$

with strong corrections setting in at $p_\mu \sim 2\pi\gamma\delta\ell_{\mathrm{P}}^2/3$. At this stage lattice refinement again becomes relevant. If corrections are to arise from a general quantum state, we should use in expressions such as (3.56) not the total volume V of our region of coordinate size $\mathcal{V}$, but the elementary volume of coordinate size ℓ_0^3. (Otherwise, the expressions we obtain cannot be considered local. Further justification comes from the analogous operators in the full theory, where only local vertex terms touched by the holonomies contribute. See [30] for a calculation using kinematical coherent states in the full theory.) The conversion from p_μ to the scale factor then does not come from $|p_\mu| = \mathcal{V}^{2/3}a^2$, but from $|p_\mu| = |F_{\ell_0}(p)| = \ell_0^2 a^2$; see also Sect. 10.1. Once the replacement of the cell volume by the patch volume is made, we refer to almost-local phase-space variables; Planck's constant or the Planck length in commutators or other quantum formulas no longer rescale when $\mathcal{V}$ changes. In terms of a, we have a correction function

$$\alpha^{(r)}(a) = \frac{\ell_0^2}{\delta r}\left(\frac{4\pi\gamma\ell_{\mathrm{P}}^2}{3}\right)^{-1} a^{2-r}\left(|a^2 + 2\pi\gamma\delta\ell_{\mathrm{P}}^2/3\ell_0^2|^{r/2} - |a^2 - 2\pi\gamma\delta\ell_{\mathrm{P}}^2/3\ell_0^2|^{r/2}\right) \tag{3.61}$$

with strong corrections setting in at $a \sim a_* := \sqrt{2\pi\gamma\delta/3}\ell_{\mathrm{P}}/\ell_0$.

This refined procedure has two consequences: (i) While using the region of size $\mathcal{V}$ would make expressions dependent on $\mathcal{V}$, which is not allowed for observables, the elementary sizes F_{ℓ_0} are independent of $\mathcal{V}$ and instead refer to an underlying discrete state via the quantity ℓ_0; and (ii) with the region ℓ_0^3 being smaller than $\mathcal{V}$ we get deeper into the small-scale regime and inverse-triad corrections will become comparatively larger (a_* being proportional to ℓ_0^{-1}). When elementary sizes are used in expressions for correction functions, the latter peak for values of the discrete increment δ of about the elementary plaquette size $a^2\ell_0^2/\ell_{\mathrm{P}}^2$ relative to the Planck scale, not for $\delta \sim a^2\mathcal{V}^{2/3}/\ell_{\mathrm{P}}^2$ which could be huge, and is even subject to coordinate and other choices. These features must be taken into account for consistent formulations of models as well as reliable phenomenology, but also for a meaningful realization of inverse-triad corrections. We will provide an inhomogeneous calculation in Sect. 10.1.4.2, exhibiting these properties explicitly. As another consequence of the fuller treatment, the range of values for the ambiguity parameter δ will be restricted.

The explicit formulas provided here rely on the Abelianization of the full theory when it is reduced to isotropy. Several new features arise if one tries to construct inverse-triad operators in an SU(2)-setting and to evaluate the characteristic commutators [31]. First, the commutators quantizing $\mathrm{tr}(h\{h^{-1}, V\})$ no longer commute with the volume operator and it becomes less clear how to compare spectra when they refer to different eigenbases. Secondly, inverse-triad operators, though still densely defined, are no longer bounded [32]. The latter is a feature which is shared by some related operators in anisotropic models discussed later, and is thus not only a consequence of non-Abelian behavior. The non-commutativity of inverse-triad operators with the volume operator, on the other hand, is directly related to the full non-Abelian nature. It is probably the most serious issue that suggests some caution toward results obtained only in isotropic models.

References

1. Giulini, D., Kiefer, C., Joos, E., Kupsch, J., Stamatescu, I.O., Zeh, H.D.: Decoherence and the Appearance of a Classical World in Quantum Theory. Springer, Berlin (1996)
2. Bojowald, M., Brizuela, D., Hernandez, H.H., Koop, M.J., Morales-Técotl, H.A.: arXiv:1011.3022
3. Gazeau, J.P., Klauder, J.: J. Phys. A: Math. Gen. **32**, 123 (1999)
4. Novello, M., Bergliaffa, S.E.P.: Phys. Rep. **463**, 127 (2008)
5. Bojowald, M.: Canonical Gravity and Applications: Cosmology, Black Holes, and Quantum Gravity. Cambridge University Press, Cambridge (2010)
6. Ashtekar, A.: Phys. Rev. D **36**(6), 1587 (1987)
7. Barbero G.J.F., (1995) Phys. Rev. D 51(10): 5507
8. Immirzi, G.: Class. Quantum Grav. **14**, L177 (1997)
9. Lewandowski, J., Newman, E.T., Rovelli, C.: J. Math. Phys. **34**, 4646 (1993)
10. Rovelli, C., Smolin, L.: Nucl. Phys. B **331**, 80 (1990)
11. Corichi, A., Krasnov, K.: Mod. Phys. Lett. A **13**, 1339 (1998). hep-th/9703177
12. Ashtekar, A., Lewandowski, J., Marolf, D., Mour ao, J., Thiemann, T.: J. Math. Phys. **36**(11), 6456 (1995). gr-qc/9504018
13. Ashtekar, A., Lewandowski, J.: J. Math. Phys. **36**(5), 2170 (1995)
14. Rovelli, C., Smolin, L.: Phys. Rev. D **52**(10), 5743 (1995)

15. Rovelli, C., Smolin, L.: Nucl. Phys. B **442**, 593 (1995). gr-qc/9411005
16. Ashtekar, A., Lewandowski, J.: Class Quantum Grav. **14**, A 55 (1997). gr-qc/9602046
17. Ashtekar, A., Lewandowski, J.: Adv. Theor. Math. Phys. **1**, 388 (1998). gr-qc/9711031
18. De Pietri, R., Rovelli, C.: Phys. Rev. D **54**(4), 2664 (1996)
19. Brunnemann, J., Thiemann, T.: Class Quantum Grav. **23**, 1289 (2006). gr-qc/0405060
20. Brunnemann, J.., Rideout, D.: Class. Quant. Grav. **25**, 065001 (2008). arXiv:0706.0469
21. Brunnemann, J., Rideout, D.: Class. Quant. Grav. **25**, 065002 (2008). arXiv:0706.0382
22. Bojowald, M.: Living Rev. Relativity **11**, 4 (2008). gr-qc/0601085
23. Bojowald, M.: Class. Quantum Grav. **17**, 1489 (2000). gr-qc/9910103
24. Bojowald, M.: Class. Quantum Grav. **19**, 2717 (2002). gr-qc/0202077
25. Fewster, C., Sahlmann, H.: Class Quantum Grav **25**, 225015 (2008). arXiv:0804.2541
26. Velhinho, J.M.: Class. Quantum Grav. **24**, 3745 (2007). arXiv:0704.2397.
27. Thiemann, T.: Class Quantum Grav **15**, 839 (1998). gr-qc/9606089
28. Thiemann, T.: Class. Quantum Grav. **15**, 1281 (1998). gr-qc/9705019
29. Bojowald, M.: Phys. Rev. D **64**, 084018 (2001). gr-qc/0105067
30. Brunnemann, J., Thiemann, T.: Class. Quantum Grav. **23**, 1395 (2006). gr-qc/0505032
31. Bojowald, M.: Class. Quantum Grav. **23**, 987 (2006). gr-qc/0508118
32. Brunnemann, J., Thiemann, T.: Class. Quantum Grav. **23**, 1429 (2006). gr-qc/0505033