

TOPICAL REVIEW

Cosmological billiards

T Damour¹, M Henneaux^{2,3} and H Nicolai⁴

¹ Institut des Hautes Etudes Scientifiques, 35, Route de Chartres, F-91440 Bures-sur-Yvette, France

² Physique Théorique et Mathématique, Université Libre de Bruxelles, C.P. 231, B-1050, Bruxelles, Belgium

³ Centro de Estudios Científicos, Casilla 1469, Valdivia, Chile

⁴ Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Am Mühlenberg 1, D-14476 Golm, Germany

Received 14 January 2003

Published 14 April 2003

Online at stacks.iop.org/CQG/20/R145**Abstract**

It is shown in detail that the dynamics of the Einstein-dilaton- p -form system in the vicinity of a spacelike singularity can be asymptotically described, at a generic spatial point, as a billiard motion in a region of Lobachevskii space (realized as a hyperboloid in the space of logarithmic scale factors). This is done within the Hamiltonian formalism, and for an arbitrary number of spacetime dimensions $D \geq 4$. A key role in the derivation is played by the Iwasawa decomposition of the spatial metric, and by the fact that the off-diagonal degrees of freedom, as well as the p -form degrees of freedom, get ‘asymptotically frozen’ in this description. For those models admitting a Kac–Moody theoretic interpretation of the billiard dynamics, we outline how to set up an asymptotically equivalent description in terms of a one-dimensional nonlinear σ -model formally invariant under the corresponding Kac–Moody group.

PACS numbers: 04.65.+e, 04.50.+h, 11.10.Kk, 11.15.–q

1. Introduction

1.1. BKL analysis in spacetime dimension $D = 4$

The nonlinearities of the Einstein equations are notably known to prevent the construction of an exact general solution. Only peculiar solutions, corresponding to idealized situations, have been explicitly derived. The singularity theorems of [48] predict the generic appearance of spacetime singularities under certain conditions, but do not provide a detailed description of how the spacetime becomes singular. From this perspective, the work of Belinskii, Khalatnikov and Lifshitz [6, 8], also known as ‘BKL’, is quite remarkable as it gives a description of the generic asymptotic behaviour of the gravitational field in *four spacetime dimensions* in the

vicinity of a spacelike singularity. As argued by these authors, near the singularity the spatial points essentially decouple, in the sense that the dynamical evolution of the spatial metric at each spatial point is asymptotically governed by a set of second-order, nonlinear ordinary differential equations with respect to time.

These differential equations are the same, at each spatial point, as those that arise in some spatially homogeneous cosmological models, which therefore provide valuable insight into the qualitative features of the general solution. In the vacuum case, the spatially homogeneous models that capture the behaviour of the general solution are of Bianchi type IX or VIII (with homogeneity groups $SU(2)$ or $SL(2, \mathbb{R})$). The asymptotic evolution of the metric can then be pictured as an infinite sequence of ‘oscillations’ of the scale factors along independent spatial directions [6, 8]. This regime is called ‘oscillatory’ or ‘of mixmaster type’ [83], or also ‘chaotic’ because it exhibits strong chaotic features [16, 77]. The coupling to matter fields does not change the picture, except if one includes a massless scalar field (equivalent to a perfect fluid with ‘stiff’ equation of state $p = \rho$), in which case the chaotic evolution is replaced by a monotonic power law evolution of the scale factors [2, 5], which mimics the Kasner solution at each spatial point and is therefore called ‘Kasner-like’. In this case, the spatially homogeneous model that captures the behaviour of the general solution is the Bianchi I model (with the Abelian group of translations in \mathbb{R}^3 as homogeneity group).

1.2. BKL analysis in spacetime dimensions $D > 4$

The extension of the BKL analysis to higher dimensions was addressed within the context of pure gravity (with no symmetry assumption) in [32, 33], where it was shown that the general BKL approach remains valid: spatial points decouple as one approaches a spacelike singularity, i.e., the dynamical evolution at each spatial point of the scale factors is again governed by ordinary differential equations. The main result of [32, 33] was that, while the general behaviour of solutions of the vacuum Einstein equations remains oscillatory for spacetime dimensions $D \leq 10$, it ceases to be so for spacetime dimensions $D \geq 11$, where it becomes Kasner-like. Let us also note that, as in four spacetime dimensions, the coupling to a massless scalar field suppresses the chaotic behaviour in any number of spacetime dimensions and makes the solution monotonic (see, e.g., [29]).

The authors of [33] did not consider the inclusion of massless p -forms, which are part of the low-energy bosonic sector of superstring/M-theory models. This task was undertaken in [24, 25], with the finding that these p -forms play a crucial role and can reinstate chaos when it is otherwise suppressed. In particular, even though pure gravity is non-chaotic in 11 spacetime dimensions, the 3-form of $D = 11$ supergravity renders the system chaotic. Similarly, the bosonic sectors of all $D = 10$ supergravities related to string models define chaotic dynamical systems, thanks again to the p -forms, and in spite of the presence of a massless scalar dilaton. It is remarkable and significant that the (maximally supersymmetric) candidate models for a unified description of the fundamental forces not only have difficulties accommodating de Sitter-type spacetimes in any straightforward fashion [45, 80], but furthermore, and without exception, exhibit BKL chaos as one approaches the initial singularity.

1.3. Billiard description of BKL behaviour

An efficient way to grasp the asymptotic behaviour of the fields as one approaches a spacelike singularity is based on the qualitative Hamiltonian methods initiated by Misner [84] in the context of the Bianchi IX models (in four spacetime dimensions). The Hamiltonian approach naturally leads to a billiard description of the asymptotic evolution, in which the logarithms

of the spatial scale factors define (after projecting out the dynamics of the overall volume factor) a geodesic motion in a region of the Lobachevskii plane H_2 , interrupted by geometric reflections against the walls bounding this region [17, 85]. Chaos follows from the fact that the Bianchi IX billiard has finite volume⁵.

As pointed out in [26, 56, 57, 71, 73] this useful billiard description is quite general and can be extended to higher spacetime dimensions, with p -forms and dilaton. If $d \equiv D - 1$ is the number of spatial dimensions, and if there are n dilatons, the billiard is a region of hyperbolic space H_{d+n-1} , each dilaton being equivalent, in the Hamiltonian, to the logarithm of a new scale factor. Besides the dimension of the hyperbolic billiard, the other ingredients that enter its definition are the walls that bound it. These walls can be of different types [24, 26]: symmetry walls related to the off-diagonal components of the spatial metric, gravitational walls related to the spatial curvature and p -form walls (electric and magnetic) arising from the p -form energy density. All these walls are hyperplanar. The billiard is a convex polyhedron with finitely many vertices, some of which are at infinity. In many important cases, the billiard can be identified with the Weyl chamber of a Kac–Moody algebra, and the reflections against the billiard walls with the fundamental Weyl reflections [23, 26, 27]. This suggests deep connections with infinite symmetries.

The main purpose of this paper is to provide a self-contained derivation of the billiard picture, in the general context of inhomogeneous solutions in D dimensions, with dilaton and p -form gauge fields. In particular, we shall present a detailed derivation of the general results, announced and used in [24, 26], on the form of the various possible walls. For that purpose, we shall rely on the Iwasawa decomposition (see, e.g., [49]) of the spatial metric. This provides an efficient derivation of the symmetry walls in any number of spacetime dimensions, which we obtain by working out explicitly the Hamiltonian that governs the dynamics in the ‘BKL limit’ or ‘small volume limit’.

The present work is exclusively concerned with properties of solutions of Einstein’s equations and their generalizations in the vicinity of a *spacelike singularity*. Accordingly, our treatment does not apply to timelike singularities, for which no analogue of causal decoupling exists (the situation may be more subtle for the borderline case of a null singularity). Furthermore, we are not making the claim here that all spacelike singularities are necessarily and uniformly of the BKL type. Rather, the results presented here are intended to refine and to generalize the work of Belinskii *et al* [6, 7] by showing that, in higher dimensions and for many theoretically relevant matter sources, one can *self-consistently* describe the behaviour of all fields, in the vicinity of a spacelike singularity, and at a generic spatial point x , in terms of (i) a simple hyperbolic billiard description of the ‘angular’ dynamics (γ, π_γ) of the logarithmic scale factors (after projection of the ‘radial’ motion (ρ, π_ρ)) and (ii) an asymptotic ‘freezing’ of the other phase space variables. This self-consistent asymptotic solution is general in the sense that it involves as many arbitrary functions of space as the most general solution.

1.4. Organization of the paper

After fixing our conventions and notation, and defining the class of Lagrangians we shall consider in section 2, we discuss in section 3 the homogeneous and diagonal Kasner solution in D spacetime dimensions, with a dilaton (as usual, the term ‘homogeneous’ implies invariance under spatial translations). This solution plays a crucial role in the BKL approach because it

⁵ Throughout this paper, the word *billiard* used as a noun in the singular will denote the dynamical system consisting of a ball moving freely on a ‘table’ (region in some Riemannian space), with elastic bounces against the edges. *Billiard* will also sometimes mean the table itself.

describes the ‘free motion between collisions’ and allows us to develop an important tool of our approach: the (Minkowskian) geometry of the scale factors. We then introduce (in section 4) the other main tool of our investigation: the Iwasawa decomposition of the spatial metric. To gain some familiarity with it, we discuss the asymptotic behaviour of non-diagonal Kasner metrics.

In section 5, we explain the appearance of sharp potential walls in full generality, without imposing any homogeneity conditions on the metric and the matter fields. We then discuss in great detail the various walls that appear in physical models: symmetry walls, gravitational walls and p -form walls (section 6). The resulting geometry of the ‘billiard’ made from all these walls is analysed in section 7. Section 8 is devoted to the case when the billiard can be identified with the Weyl chamber of a Kac–Moody algebra. To deal with this case, we set up a Kac–Moody theoretic formulation of the billiard in terms of a nonlinear σ -model based on the relevant Kac–Moody group. As we will show, the asymptotic limit of these σ -models coincides with the asymptotic limit of the models discussed in the main part of this paper. An appendix illustrating by a toy model the asymptotic freezing of the off-diagonal degrees of freedom concludes this paper.

We should stress that our analysis is purely classical and accordingly, as it stands, is valid only up to the Planck (or string) scale. We shall also ignore fermionic fields throughout. Nevertheless, it is reasonable to expect that some of the ideas discussed here will remain relevant in a more general quantum mechanical context, at least qualitatively (see, e.g., [91] for some recent ideas in this direction). The subject of Hamiltonian cosmology has a long history in the context of four-dimensional, spatially homogeneous spacetimes and provides useful insight into the general discussion presented here. For reviews on this subject, with an extensive bibliography, see [61, 94, 95]; see also the topical review on multidimensional gravity [59].

2. Models and conventions

2.1. The models

We consider models of the general form

$$S[G_{MN}, \phi, A^{(p)}] = \int d^D x \sqrt{-G} \left[R - \partial_M \phi \partial^M \phi - \frac{1}{2} \sum_p \frac{1}{(p+1)!} e^{\lambda_p \phi} F_{M_1 \dots M_{p+1}}^{(p)} F^{(p) M_1 \dots M_{p+1}} \right] + \dots \quad (2.1)$$

where units are chosen such that $16\pi G_N = 1$ (where G_N is Newton’s constant) and the spacetime dimension $D \equiv d + 1$ is left unspecified. Besides the standard Einstein–Hilbert term, the above Lagrangian contains a dilaton field ϕ and a number of p -form fields $A_{M_1 \dots M_p}^{(p)}$ (for $p \geq 0$). As a convenient common formulation, we adopt the Einstein conformal frame and normalize the kinetic term of the dilaton ϕ with weight 1 w.r.t. the Ricci scalar. The Einstein metric G_{MN} has Lorentz signature $(- + \dots +)$ and is used to lower or raise the indices; its determinant is denoted by G . The p -form field strengths $F^{(p)} = dA^{(p)}$ are normalized as

$$F_{M_1 \dots M_{p+1}}^{(p)} = (p+1) \partial_{[M_1} A_{M_2 \dots M_{p+1}}^{(p)} \equiv \partial_{M_1} A_{M_2 \dots M_{p+1}}^{(p)} \pm p \text{ permutations}. \quad (2.2)$$

The dots in the action (2.1) indicate possible modifications of the field strength by additional Yang–Mills or Chapline–Manton-type couplings [11, 15], such as $F_C = dC^{(2)} - C^{(0)} dB^{(2)}$ for

two 2-forms $C^{(2)}$ and $B^{(2)}$ and a 0-form $C^{(0)}$, as they occur in type IIB supergravity. Further modifications include Chern–Simons terms, as in the action for $D = 11$ supergravity [22]. The real parameter λ_p measures the strength of the coupling of $A^{(p)}$ to the dilaton. When $p = 0$, we assume that $\lambda_0 \neq 0$ so that there is only one dilaton. This is done mostly for notational convenience. If there were other dilatons among the 0-forms, these should be separated off from the p -forms because they play a distinct role. They would define additional spacelike directions in the space of the (logarithmic) scale factors and would correspondingly increase the dimension of the relevant hyperbolic billiard.

The metric G_{MN} , the dilaton field(s) ϕ and the p -form fields $A_{M_1 \dots M_p}^{(p)}$ are *a priori* arbitrary functions of both space and time, on which *no symmetry conditions are imposed*. Nevertheless, it will turn out that the evolution equations near the singularity will be asymptotically the same as those of certain homogeneous cosmological models. It is important to keep in mind that this simplification does not follow from imposing extra dimensional reduction conditions but emerges as a direct consequence of the general dynamics.

2.2. Gauge conditions

Our analysis applies both to past and future singularities, and in particular to Schwarzschild-type singularities inside black holes. To follow historical usage, we shall assume for definiteness that the spacelike singularity lies in the past, at finite distance in proper time. More specifically, we shall adopt a spacetime slicing such that the singularity ‘occurs’ on a constant time slice ($t = 0$ in proper time). The slicing is built by using pseudo-Gaussian coordinates defined by vanishing lapse $N^i = 0$, with metric

$$ds^2 = -(N(x^0, x^i) dx^0)^2 + g_{ij}(x^0, x^i) dx^i dx^j. \quad (2.3)$$

In order to simplify various formulae later, we shall find it useful to introduce a rescaled lapse function

$$\tilde{N} \equiv N/\sqrt{g} \quad (2.4)$$

where $g \equiv \det g_{ij}$. We shall see that a useful gauge, within the Hamiltonian approach, is that defined by requiring

$$\tilde{N} = \rho^2, \quad (2.5)$$

where ρ^2 is a quadratic combination of the logarithms of the scale factors and the dilaton(s), which we will define below in terms of the Iwasawa decomposition. After fixing the time zero hypersurface, the only coordinate freedom left in the pseudo-Gaussian gauge (2.5) is that of making time-independent changes of spatial coordinates $x^i \rightarrow x'^i = f^i(x^j)$. Since the gauge condition (2.5) is not invariant under spatial coordinate transformations, such changes of coordinates have the unusual feature of also changing the slicing.

Throughout this paper, we will reserve the label t for the *proper time*

$$dt = -N dx^0 = -\tilde{N} \sqrt{g} dx^0, \quad (2.6)$$

whereas the time coordinate associated with the special gauge (2.5) will be designated by T , namely,

$$dT = -\frac{dt}{\rho^2 \sqrt{g}}. \quad (2.7)$$

Sometimes, it will also be useful to introduce the ‘intermediate’ time coordinate τ that would correspond to the gauge condition $\tilde{N} = 1$. It is explicitly defined by

$$d\tau = -\frac{dt}{\sqrt{g}} = \rho^2 dT. \quad (2.8)$$

At the singularity, the proper time t is assumed to remain finite and to decrease towards 0^+ . By contrast, the coordinates T and τ both increase towards $+\infty$, as ensured by the minus sign in (2.6). Irrespective of the choice of coordinates, the spatial volume density g is assumed to collapse to zero at each spatial point in this limit.

As for the p -form fields, we shall assume, throughout this paper, a generalized temporal gauge, namely,

$$A_{0i_2\dots i_p}^{(p)} = 0 \quad (2.9)$$

where small Latin letters i, j, \dots denote spatial indices from now on. This choice leaves the freedom of performing time-independent gauge transformations, and therefore fixes the gauge only partially.

3. Geometry of the space of the scale factors

3.1. Supermetric and Hamiltonian

To set the stage for our general Hamiltonian approach, it is useful to study first in detail the dynamics defined by considering *only the kinetic terms of the metric and of the dilaton(s)*. The corresponding reduced action is obtained from (2.1) by setting $A^{(p)} = 0$ and assuming that all the other fields depend only on time. In terms of a general time coordinate x^0 this reduced action reads

$$S[g_{ij}, \phi, \tilde{N}] = \int dx^0 \tilde{N}^{-1} \left[\frac{1}{4} (\text{Tr}(g^{-1}\dot{g})^2 - (\text{Tr} g^{-1}\dot{g})^2) + \dot{\phi}^2 \right] \quad (3.10)$$

where we have suppressed an integral $\int d^d x$ over the spatial volume for notational simplicity. Furthermore, we make use of the notation introduced in (2.3) and (2.4) with $\dot{F} \equiv dF/dx^0$, and adopt a matrix notation where $g(t) \in GL(d, \mathbb{R})$ stands for the matrix (g_{ij}) representing the spatial components of the metric at each spatial point.

The action (3.10) is the (quadratic-in-velocities) action for a massless free particle with coordinates (g_{ij}, ϕ) moving in a curved target space with metric

$$d\sigma^2 = \frac{1}{4} [\text{Tr}(g^{-1}dg)^2 - (\text{Tr} g^{-1}dg)^2] + d\phi^2. \quad (3.11)$$

We designate by $d\sigma^2$ the line element in this target ‘superspace’ to distinguish it from the line element in physical spacetime, which we denote by ds^2 . The first two terms in the rhs of (3.11) define the so-called DeWitt supermetric in the space of the metric coefficients g_{ij} [34]. If several dilatons ϕ^i (for $i = 1, \dots, n$) were present the term $d\phi^2$ in equation (3.11) would be replaced by $\sum_i (d\phi^i)^2$. That is, each dilaton adds a (flat) direction in the target superspace.

The rescaled lapse \tilde{N} plays the role of an ‘einbein’ in the geodesic action (3.10). As usual, extremizing over \tilde{N} yields the ‘zero-mass constraint’

$$\frac{1}{4} (\text{Tr}(g^{-1}\dot{g})^2 - (\text{Tr} g^{-1}\dot{g})^2) + \dot{\phi}^2 = 0. \quad (3.12)$$

Thus, the motion is given by a null geodesic of the metric (3.11). An affine parameter along those geodesics is $d\tau = +\tilde{N} dx^0 = -dt/\sqrt{g}$, cf equation (2.8). In terms of the parameter τ the equations of motion read

$$\frac{d}{d\tau} \left(g^{-1} \frac{dg}{d\tau} \right) = 0, \quad \frac{d^2}{d\tau^2} \phi = 0. \quad (3.13)$$

For *diagonal metrics*

$$g_K = \exp[\text{diag}(-2\beta)] \iff g_{ij}^K = \exp(-2\beta^i)\delta_{ij} \quad (3.14)$$

the supermetric (3.11) reduces to

$$\begin{aligned} d\sigma^2 &= \text{Tr } d\beta^2 - (\text{Tr } d\beta)^2 + d\phi^2 \\ &= \sum_{i=1}^d (d\beta^i)^2 - \left(\sum_{i=1}^d d\beta^i \right)^2 + d\phi^2 \equiv G_{\mu\nu} d\beta^\mu d\beta^\nu. \end{aligned} \quad (3.15)$$

Here we have introduced a $(d+1)$ -dimensional space with coordinates β^μ , with indices running over $\mu = 1, \dots, d+1$, such that the first d coordinates β^i correspond to the logarithms of the scale factors of the spatial metric (cf equation (3.14)), and the $(d+1)$ th coordinate $\beta^{d+1} \equiv \phi$ represents the dilaton⁶. The explicit form of the (flat) target space metric $G_{\mu\nu}$ can be read off directly from (3.15). The action for diagonal metrics is

$$S[\beta^\mu, \tilde{N}] = \int dx^0 \tilde{N}^{-1} G_{\mu\nu} \dot{\beta}^\mu \dot{\beta}^\nu. \quad (3.16)$$

In what follows, we shall refer to this space as the ‘extended space of (logarithmic) scale factors’ or only ‘the β -space’ for short.

Combining the scale factors and the dilaton(s) in a single space is natural because we know from Kaluza–Klein theory that the dilaton can be viewed as the logarithm of a scale factor in one extra spatial dimension. Independent of whether the original metric (2.3) has non-vanishing curvature or not, the metric (3.15) induced in the space of the scale factors (possibly including the dilaton) is flat. More precisely, the metric $G_{\mu\nu}$ of the (extended) space of scale factors is a Minkowski metric in \mathbb{R}^{d+1} with signature $(- + + \dots +)$. For instance, the direction in which only the dilaton varies (i.e., $d\beta^\mu \propto (0, \dots, 0, 1)$) is spacelike, while the direction in which only one scale factor varies (e.g., $d\beta^\mu \propto (1, 0, \dots, 0)$) is null. A timelike direction in this space is the direction $d\beta^\mu \propto (1, 1, \dots, 1, 0)$. This reflects the familiar fact that the gravitational action is not bounded from below (even with Euclidean signature): conformal transformations of the metric, in which the scale factors are all scaled in the same fashion, make $d\sigma^2$ negative. It is this characteristic feature of gravity which is responsible for the Lorentzian nature of the Kac–Moody algebras which emerge in the analysis of the billiard symmetries [26]. The Lorentzian signature of the metric in the space of the scale factors enables one to define the light cone through any point. We define the time orientation to be such that future-pointing vectors v^μ have $\sum_i v^i > 0$. Geometrically, small volumes (small g) are associated with large positive values of $\sum_i \beta^i$. Large volumes (large g), on the other hand, mean large negative values of $\sum_i \beta^i$. We are interested in the small volume limit, i.e., $\sum_i \beta^i \rightarrow +\infty$.

The Hamiltonian form of the action for the diagonal metric degrees of freedom and the dilaton is

$$S[\beta^\mu, \pi_\mu, \tilde{N}] = \int dx^0 \left[\pi_\mu \dot{\beta}^\mu - \frac{1}{4} \tilde{N} G^{\mu\nu} \pi_\mu \pi_\nu \right] \quad (3.17)$$

where $G^{\mu\nu}$ is the inverse of $G_{\mu\nu}$. Explicitly,

$$G^{\mu\nu} \pi_\mu \pi_\nu \equiv \sum_{i=1}^d \pi_i^2 - \frac{1}{d-1} \left(\sum_{i=1}^d \pi_i \right)^2 + \pi_\phi^2 \quad (3.18)$$

⁶ Obviously, the range of indices would be extended to $\mu, \nu = 1, \dots, d+n$ in the presence of n dilatons.

where $\pi_\mu \equiv (\pi_i, \pi_\phi)$ are the momenta conjugate to β^i and ϕ , respectively, i.e.,

$$\pi_\mu = 2\tilde{N}^{-1}G_{\mu\nu}\dot{\beta}^\nu = 2G_{\mu\nu}\frac{d\beta^\nu}{d\tau} \equiv 2G_{\mu\nu}v^\nu. \quad (3.19)$$

Here, the τ -parameter velocities have been designated by $v^\mu \equiv d\beta^\mu/d\tau$.

3.2. Diagonal Kasner solution

The Kasner solution (with or without dilaton) is now easily obtained by solving (3.12) and (3.13) in the diagonal case. Indeed, the equations of motion reduce to

$$\frac{d^2\beta^\mu}{d\tau^2} = 0. \quad (3.20)$$

They are solved by

$$\beta^\mu = v^\mu\tau + \beta_0^\mu \quad (3.21)$$

where v^μ and β_0^μ are constants of the motion. The ‘zero-mass constraint’ becomes

$$G_{\mu\nu}v^\mu v^\nu = 0. \quad (3.22)$$

One can transform the simple affine parameter solution (3.21) into the usual Kasner solution expressed in terms of the proper time by integrating the relation $dt = -\sqrt{g}d\tau$, with $\sqrt{g} = \exp(-\sum_i \beta^i)$, whence $t \propto \exp(-(\sum_i v^i)\tau)$, or

$$\tau = -\frac{1}{\sum_i v^i} \ln t + \text{const.} \quad (3.23)$$

We require $\sum_i v^i > 0$ to remain consistent with our convention that $\tau \rightarrow +\infty$ at $t \rightarrow 0^+$ near the singularity. This yields

$$ds^2 = -dt^2 + \sum_{i=1}^d A_i^2(t)(dx^i)^2, \quad A_i(t) = b_i t^{p_i} \quad (3.24)$$

$$\phi = -p_\phi \ln t + C_\phi \quad (3.25)$$

where $b_i \equiv e^{-\beta_0^i}$ and $C_\phi \equiv \beta_0^{d+1}$ are integration constants and the minus sign in front of p_ϕ in (3.25) is included for the sake of uniformity in the formulae below (if there is no dilaton one simply sets $p_\phi = C_\phi = 0$). By rescaling the spatial coordinates, one can set $b_i = 1$ and obtain the standard (proper time) form of the Kasner metric. The Kasner exponents $p_\mu = (p_i, p_\phi)$ are given in terms of the affine velocities $v^\mu \equiv d\beta^\mu/d\tau$ by

$$p_\mu = \frac{v^\mu}{\sum_i v^i}. \quad (3.26)$$

Note that the sum in the denominator does not include the dilaton.

They are subject to the quadratic constraint⁷

$$\sum_{i=1}^d p_i^2 - \left(\sum_{i=1}^d p_i \right)^2 + p_\phi^2 = 0 \quad (3.27)$$

coming from the ‘zero-mass condition’, and to the linear constraint

$$\sum_{i=1}^d p_i = 1 \quad (3.28)$$

coming from their definition above.

⁷ Contrary to the variables β^μ and the velocities v^μ , we do not assign any covariance properties to the standard Kasner exponents, but regard them simply as parameters, leaving their labels always in the lower position.

If there are no dilatons, it follows from the above equations that there is at least one Kasner exponent which is negative, so at least one of the scale factors $A_i(t)$ blows up as $t \rightarrow 0$. The scale factors associated with positive Kasner exponents contract to zero monotonically. By contrast, in the presence of a dilaton, all the Kasner exponents can be positive simultaneously. In both cases, there is an overall contraction of the spatial volume since the determinant g of the spatial metric tends to zero. Indeed, a consequence of the linear constraint above on the Kasner exponents is

$$g \propto t^2. \quad (3.29)$$

Note that relations (3.23) and (3.29) have been derived only for the exact (homogeneous) Kasner solution in the vacuum.

3.3. Hyperbolic space

Still as a preparation for dealing with the general inhomogeneous case, let us present an alternative way of solving the dynamics defined by the kinetic terms of diagonal metrics, i.e., the action (3.16), or its Hamiltonian form. This alternative way will turn out to be very useful for describing the asymptotic dynamics of general inhomogeneous metrics. It consists in decomposing the motion of the variables β^μ into two pieces, namely a radial part ρ and an angular one γ^μ . Here ‘radial’ and ‘angular’ refer to polar coordinates in the Minkowski space of the β^μ . More precisely, shifting, if necessary, the origin in β -space to arrange that $v_\mu \beta_0^\mu < 0$, the β^μ trajectories (3.21) will, for large enough values of τ , get inside the future light cone of the origin, i.e.,

$$\beta^\mu \beta_\mu = 2v_\mu \beta_0^\mu \tau + \beta_{0\mu} \beta_0^\mu < 0. \quad (3.30)$$

Let us then decompose β^μ in hyperbolic polar coordinates (ρ, γ^μ) , i.e.,

$$\beta^\mu = \rho \gamma^\mu, \quad (3.31)$$

where γ^μ are coordinates on the future sheet of the unit hyperboloid, which are constrained by

$$\gamma^\mu \gamma_\mu = -1 \quad (3.32)$$

and ρ is the timelike variable

$$\rho^2 \equiv -\beta^\mu \beta_\mu > 0. \quad (3.33)$$

This decomposition naturally introduces the unit hyperboloid (‘ γ -space’), which is a realization of the m -dimensional hyperbolic (Lobachevskii) space H_m , with $m = d - 1 + n$ if there are $n \geq 0$ dilatons.

In terms of the ‘polar’ coordinates ρ and γ^μ , the metric in β -space becomes

$$d\sigma^2 = -d\rho^2 + \rho^2 d\Sigma^2 \quad (3.34)$$

where $d\Sigma^2$ is the metric on the γ -space H_m . In these variables, the Hamiltonian reads

$$H_0 = \frac{\tilde{N}}{4} \left[-\pi_\rho^2 + \frac{1}{\rho^2} \pi_\gamma^2 \right] \quad (3.35)$$

where π_γ are the momenta conjugate to the constrained hyperbolic coordinates γ^μ . (It is straightforward to introduce angular coordinates on H_m to derive more explicit formulae for $d\Sigma^2$ and π_γ^2 .) The extra index 0 on H is to underline the fact that this Hamiltonian refers only to a small part of the total Hamiltonian which we shall study below, because it describes only the kinetic terms associated with the $(d + n)$ variables β^μ .

An equivalent expression is

$$H_0 = \frac{\tilde{N}}{4\rho^2}[-\pi_\lambda^2 + \pi_\gamma^2] \quad (3.36)$$

where we have introduced the new configuration variable

$$\lambda \equiv \ln \rho \equiv \frac{1}{2} \ln(-G_{\mu\nu}\beta^\mu\beta^\nu) \quad (3.37)$$

with conjugate momentum π_λ .

The form (3.36) of the Hamiltonian shows that, when using a radial projection on H_m , the motion becomes simplest in the gauge

$$\tilde{N} = \rho^2, \quad (3.38)$$

in terms of which (3.36) reduces to a free Hamiltonian on the pseudo-Riemannian space with metric $-\mathrm{d}\lambda^2 + \mathrm{d}\Sigma^2$. In the gauge (3.38) the logarithmic radial momentum $\pi_\lambda = \rho\pi_\rho$ is a constant of the motion. In this gauge, we see that the free motion of the β is projected onto a geodesic motion on H_m .

The coordinate time T associated with this gauge, see equation (2.7), is linked to the affine parameter τ via

$$\mathrm{d}T = \frac{\mathrm{d}\tau}{\rho^2}. \quad (3.39)$$

From (3.30) we get that ρ^2 varies linearly with τ ,

$$\rho^2 = -\beta_\mu\beta^\mu = -2v_\mu\beta_0^\mu\tau - \beta_0^\mu\beta_{0\mu} \quad (3.40)$$

which implies

$$T = -\frac{1}{2v_\mu\beta_0^\mu} \ln \tau + \text{const.} \quad (3.41)$$

Recalling that τ varies logarithmically with the proper time t , we see that $T \propto \ln |\ln t|$.

One can also use the configuration variable (3.37) as an intrinsic time variable to describe the dynamics. In view of (3.40) we have

$$\lambda = \frac{1}{2} \ln \tau + \text{const} = \frac{1}{2} \ln |\ln t| + \text{const.} \quad (3.42)$$

Note that the various links above between the different time scales are derived only for an exact Kasner solution. Similar relations will hold asymptotically in the general inhomogeneous case (see discussion in section 7).

4. Iwasawa decomposition and dynamics of non-diagonal Kasner metrics

4.1. Iwasawa decomposition

For homogeneous solutions *in vacuo*, the metric remains diagonal if the initial data are so. This is, in general, not true when matter (such as p -forms) or inhomogeneities are included, in which case off-diagonal components generically appear even if there are none initially. For this reason, it is important to understand the role of off-diagonal terms already in this simpler homogeneous context, by examining the evolution of initial data that are not diagonal. And, as for the simple diagonal metrics above, it is instructive to study the dynamics of non-diagonal metrics in several complementary ways.

A first way of dealing with the evolution of initially non-diagonal metrics is to perform a suitable linear transformation on the diagonal Kasner solution. If L is the linear transformation

needed to diagonalize the initial data, it is easy to see from (3.13) that the solution is given, in terms of the parameter τ , by

$$g(\tau) = L^T g_K(\tau) L \tag{4.1}$$

where superscript T denotes transposition and $g_K(\tau)$ is the diagonal Kasner solution (3.14), (3.21). The dilaton, being a scalar, is still given by the same expression as before, i.e. $\phi = \beta^{d+1} = v^{d+1} \tau + \beta_0^{d+1}$. Note the relation

$$\det g(\tau) = (\det L)^2 \det g_K(\tau), \tag{4.2}$$

so that relation (3.23) between τ and t still holds. Therefore $\det g$ still goes to zero like t^2 .

A second way of describing the evolution of non-diagonal metrics is to perform an Iwasawa decomposition of the metric,

$$g = \mathcal{N}^T \mathcal{A}^2 \mathcal{N} \tag{4.3}$$

where \mathcal{N} is an upper triangular matrix with 1's on the diagonal and \mathcal{A} is a diagonal matrix with positive entries, which we parametrize as

$$\mathcal{A} = e^{-\beta}, \quad \beta = \text{diag}(\beta^1, \beta^2, \dots, \beta^d). \tag{4.4}$$

The explicit form of (4.3) reads

$$g_{ij} = \sum_{a=1}^d e^{-2\beta^a} \mathcal{N}^a{}_i \mathcal{N}^a{}_j. \tag{4.5}$$

The evolution of the new configuration variables $\beta^a = f(g_{ij})$ defined by the Iwasawa decomposition differs, in general, from that of the quantities β^i entering the diagonal Kasner solution (3.21), i.e. $\beta^a \neq \beta^i$ except for the special case $\mathcal{N}^a{}_i = \delta_i^a$. Henceforth in this paper, the notation β will always refer to the logarithmic scale factors β^a w.r.t. the Iwasawa frame (4.5) (to be augmented, if needed, by the dilaton as a $(d + 1)$ th coordinate $\beta^{d+1} \equiv \phi$). Furthermore, we adopt the convention here and in the remainder that summation over the spatial coordinate indices i, j, \dots is always understood, whereas sums over the Iwasawa frame indices a, b, \dots will always be written out.

One can view the Iwasawa decomposition⁸ as the Gram–Schmidt orthogonalization of the initial coordinate coframe dx^i , which is indeed a triangular process,

$$g_{ij} dx^i dx^j = \sum_{a=1}^d e^{-2\beta^a} \theta^a \otimes \theta^a. \tag{4.6}$$

Starting with $\theta^d = dx^d$, one successively constructs the next θ by adding linear combinations of the dx^k (for $k > j$) in such a way that θ^{d-1} is orthogonal to θ^d , θ^{d-2} is orthogonal to both θ^d and θ^{d-1} , etc. Explicitly,

$$\begin{aligned} \theta^d &= dx^d, \\ \theta^{d-1} &= dx^{d-1} + \mathcal{N}^{d-1}{}_d dx^d, \\ \theta^{d-2} &= dx^{d-2} + \mathcal{N}^{d-2}{}_{d-1} dx^{d-1} + \mathcal{N}^{d-2}{}_d dx^d, \\ &\dots \end{aligned} \tag{4.7}$$

Equivalently,

$$\theta^a = \mathcal{N}^a{}_i dx^i. \tag{4.8}$$

⁸ The Iwasawa decomposition applies to general symmetric spaces (see, e.g., [49]). In our case, the relevant symmetric space is the coset space $SL(d, \mathbb{R})/SO(d)$ since the space of positive definite symmetric matrices can be identified with $GL(d, \mathbb{R})/O(d)$, which is isomorphic to $SL(d, \mathbb{R})/SO(d) \times \mathbb{R}^+$.

Thus \mathcal{N}^a_i vanishes for $a > i$, is equal to 1 for $a = i$, and has a non-trivial coordinate dependence only for $a < i$. It is easily seen that the determinant of the matrix \mathcal{N} is equal to 1. Therefore the sum of the β is directly related to the metric determinant (as in the diagonal case): $g = \det g = \det \mathcal{A}^2 = \exp(-2 \sum_a \beta^a)$.

In the following, we shall also need the vectorial frame $\{e_a\}$ dual to the coframe θ^a ,

$$e_a = \mathcal{N}^i_a \frac{\partial}{\partial x^i} \quad (4.9)$$

where the matrix \mathcal{N}^i_a is the inverse of \mathcal{N}^a_i , i.e., $\mathcal{N}^a_i \mathcal{N}^i_b = \delta^a_b$. It is again an upper triangular matrix with 1's on the diagonal.

4.2. Asymptotics of non-diagonal Kasner metrics

Later in this paper we will deal with the generic inhomogeneous, non-diagonal metric $g_{ij}(x^0, x^i)$ and show how to study its Hamiltonian dynamics directly in terms of the Iwasawa variables $(\beta^a, \mathcal{N}^a_i)$ and their conjugate momenta. But let us first take advantage of the known explicit solution (4.1) of the non-diagonal evolution in order to understand the qualitative behaviour of the Iwasawa variables $(\beta^a, \mathcal{N}^a_i)$ as $t \rightarrow 0$, i.e., $\tau \rightarrow +\infty$. To be concrete, let us explicitly consider the case of three spatial dimensions ($d = 3, D = 4$) using the results of [51] for the Iwasawa decomposition of three-dimensional metrics.

Setting

$$\mathcal{N} = \begin{pmatrix} 1 & n_1 & n_2 \\ 0 & 1 & n_3 \\ 0 & 0 & 1 \end{pmatrix} \quad (4.10)$$

together with

$$\mathcal{A} = \begin{pmatrix} e^{-\beta^1} & 0 & 0 \\ 0 & e^{-\beta^2} & 0 \\ 0 & 0 & e^{-\beta^3} \end{pmatrix}, \quad (4.11)$$

one finds

$$\begin{aligned} g_{11} &= e^{-2\beta^1}, & g_{12} &= n_1 e^{-2\beta^1}, & g_{13} &= n_2 e^{-2\beta^1}, \\ g_{22} &= n_1^2 e^{-2\beta^1} + e^{-2\beta^2}, & g_{23} &= n_1 n_2 e^{-2\beta^1} + n_3 e^{-2\beta^2}, \\ g_{33} &= n_2^2 e^{-2\beta^1} + n_3^2 e^{-2\beta^2} + e^{-2\beta^3} \end{aligned} \quad (4.12)$$

from which one obtains

$$\begin{aligned} \beta^1 &= -\frac{1}{2} \ln g_{11}, & \beta^2 &= -\frac{1}{2} \ln \left[\frac{g_{11} g_{22} - g_{12}^2}{g_{11}} \right], \\ \beta^3 &= -\frac{1}{2} \ln \left[\frac{g}{g_{11} g_{22} - g_{12}^2} \right], & n_1 &= \frac{g_{12}}{g_{11}}, \\ n_2 &= \frac{g_{13}}{g_{11}}, & n_3 &= \frac{g_{23} g_{11} - g_{12} g_{13}}{g_{11} g_{22} - g_{12}^2}. \end{aligned} \quad (4.13)$$

On the other hand, (4.1) and (3.24) yield

$$g_{ij}(t) = t^{2p_1} l_i l_j + t^{2p_2} m_i m_j + t^{2p_3} r_i r_j \quad (4.14)$$

with the constant matrix

$$L = \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \\ r_1 & r_2 & r_3 \end{pmatrix}, \tag{4.15}$$

where we have absorbed the b_i in l_i, m_i and r_i .

Combining these relations, we can deduce the explicit time dependence of the Iwasawa variables,

$$\begin{aligned} \beta^1(t) &= -\frac{1}{2} \ln X, & \beta^2(t) &= -\frac{1}{2} \ln \left[\frac{Y}{X} \right], \\ \beta^3(t) &= -\frac{1}{2} \ln \left[\frac{t^{2(p_1+p_2+p_3)} (\det L)^2}{Y} \right], \\ n_1(t) &= \frac{t^{2p_1} l_1 l_2 + t^{2p_2} m_1 m_2 + t^{2p_3} r_1 r_2}{X}, \\ n_2(t) &= \frac{t^{2p_1} l_1 l_3 + t^{2p_2} m_1 m_3 + t^{2p_3} r_1 r_3}{X}, & n_3(t) &= \frac{Z}{Y} \end{aligned} \tag{4.16}$$

with

$$\begin{aligned} X(t) &= t^{2p_1} l_1^2 + t^{2p_2} m_1^2 + t^{2p_3} r_1^2, \\ Y(t) &= t^{2p_1+2p_2} (l_1 m_2 - l_2 m_1)^2 + t^{2p_1+2p_3} (l_1 r_2 - l_2 r_1)^2 + t^{2p_2+2p_3} (m_1 r_2 - m_2 r_1)^2, \\ Z(t) &= t^{2p_1+2p_2} (l_1 m_2 - l_2 m_1)(l_1 m_3 - l_3 m_1) + t^{2p_1+2p_3} (l_1 r_2 - l_2 r_1)(l_1 r_3 - l_3 r_1) \\ &\quad + t^{2p_2+2p_3} (m_1 r_2 - m_2 r_1)(m_1 r_3 - m_3 r_1). \end{aligned} \tag{4.17}$$

These explicit formulae show that the evolution of the Iwasawa variables (β, \mathcal{N}) in the generic non-diagonal case is rather complicated. However, what will be important in the following is that they drastically simplify in the asymptotic limit $t \rightarrow 0$, i.e., $\tau \rightarrow +\infty$. Without loss of generality, one can assume $p_1 \leq p_2 \leq p_3$. If necessary, this can be achieved by multiplying L by an appropriate permutation matrix. We shall in fact only consider $p_1 < p_2 < p_3$, leaving the discussion of the limiting cases to the reader. For generic L , i.e., $l_1 \neq 0, r_1 \neq 0, l_1 m_2 - l_2 m_1 \neq 0$ and $m_1 r_2 - m_2 r_1 \neq 0$, one then finds the following asymptotic behaviour as $t \rightarrow 0$, i.e., $\tau \rightarrow +\infty$:

$$\begin{aligned} \tau \rightarrow +\infty: \quad \beta^1 &\sim v^1 \tau, & \beta^2 &\sim v_2 \tau, & \beta^3 &\sim v_3 \tau, \\ n_1 &\rightarrow \frac{l_2}{l_1}, & n_2 &\rightarrow \frac{l_3}{l_1}, & n_3 &\rightarrow \frac{l_1 m_3 - l_3 m_1}{l_1 m_2 - l_2 m_1}. \end{aligned} \tag{4.18}$$

Here we used the links derived above between (p_μ, t) and (v^μ, τ) .

The remarkable feature of these results is that, in the limit $\tau \rightarrow +\infty$, the evolution of the Iwasawa variables (β, \mathcal{N}) is essentially as simple as the evolution we obtained in the diagonal case. Namely, the diagonal degrees of freedom β in the Iwasawa decomposition behave linearly with τ , while the elements of the upper triangular matrix \mathcal{N} tend to constants. That the n_i should asymptotically tend to constants should be clear because they are homogeneous functions of degree zero in the metric coefficients—in fact, ratios of polynomials of degree one or two in the t^{2p_i} . It is more subtle that the scale factors $e^{-2\beta^i}$, which are homogeneous of degree one in the g_{ij} , are not all driven by the fastest growing (or least decreasing) term (t^{2p_1} for $t \rightarrow 0^+$). This is what happens for the first scale factor $e^{-2\beta^1}$. However, the second scale factor $e^{-2\beta^2}$ feels the subleading term t^{2p_2} because the leading term drops from its numerator, equal to the minor $g_{11} g_{22} - (g_{12})^2$. Similar cancellations occur for the last scale factor $e^{-2\beta^3}$, which feels only the smallest term t^{2p_3} as $t \rightarrow 0^+$.

The results (4.18) admit a simple generalization to d dimensions. By repeating the above explicit calculation, one can prove that, in any dimension d , the β become asymptotically linear functions of τ , as in the diagonal case, with coefficients that are given by a permutation of the underlying diagonal Kasner exponents. Furthermore, the \mathcal{N}^a_i tend to constants, a phenomenon which we will refer to as the ‘asymptotic freezing’ of the off-diagonal metric variables. More precisely, let $v^1 \leq v^2 \leq \dots \leq v^d$ be the ordered (unnormalized) underlying Kasner exponents; then, for generic L , we have

$$\beta^a \sim v^a \tau \quad \text{and} \quad \mathcal{N}^a_i \rightarrow \text{const} \quad (4.19)$$

as $\tau \rightarrow +\infty$ or, equivalently, $t \rightarrow 0^+$. An alternative derivation of these results will follow, as a particular case, from the general result we shall derive below concerning the qualitative evolution of generic, inhomogeneous, matter-driven solutions in Iwasawa variables (after projecting out the radial motion of the Iwasawa β variables).

Note also that, in the simple case of homogeneous non-diagonal metrics *in vacuo* (without curvature and/or matter) one can discuss not only the limit $t \rightarrow 0$, but also the limit $t \rightarrow +\infty$ (i.e., $\tau \rightarrow -\infty$). In this second limit, one finds again that the β become asymptotically linear functions of τ , with coefficients that are given by a permutation of the underlying Kasner exponents, and that the \mathcal{N}^a_i tend to constants. More precisely, for any d , $\beta^a \sim v^{(d-a)} \tau$, while the explicit results in $d = 3$ read

$$\begin{aligned} \tau \rightarrow -\infty: \quad \beta^1 &\sim p_3 \tau, & \beta^2 &\sim p_2 \tau, & \beta^3 &\sim p_1 \tau, \\ n_1 &\rightarrow \frac{r_2}{r_1}, & n_2 &\rightarrow \frac{r_3}{r_1}, & n_3 &\rightarrow \frac{m_1 r_3 - m_3 r_1}{m_1 r_2 - m_2 r_1}. \end{aligned} \quad (4.20)$$

Note that in *both* limits, one has $\beta^1 \leq \beta^2 \leq \dots \leq \beta^d$. The fact that the β are ordered in this way is due to our choice of an *upper triangular* \mathcal{N} ; had we taken \mathcal{N} to be *lower triangular* instead, these inequalities would have been reversed.

5. Asymptotic dynamics in the general case

Having warmed up with the simple diagonal and homogeneous Kasner solution in section 3, and having introduced the Iwasawa decomposition of the metric (section 4), we are now ready to apply these techniques to the description of the asymptotic dynamics in the limit $t \rightarrow 0$ to the general inhomogeneous, matter-driven case. The main ingredients (already introduced above for the homogeneous solutions) of our study are as follows:

- Use of the Hamiltonian formalism.
- Iwasawa decomposition of the metric, i.e., $g_{ij} \rightarrow (\beta^a, \mathcal{N}^a_i)$.
- Decomposition of $\beta^\mu = (\beta^a, \phi)$ into radial (ρ) and angular (γ^μ) parts.
- Use of the pseudo-Gaussian gauge (2.5), i.e., of the time coordinate T as the evolution parameter.

More explicitly, with the conventions already described before, we assume that in some spacetime patch, the metric is given by (2.3) (pseudo-Gaussian gauge), such that the local volume g collapses at each spatial point as $x^0 \rightarrow +\infty$, in such a way that the proper time t tends to 0^+ . We work in the Hamiltonian formalism, i.e., with first-order evolution equations in the phase space of the system. For instance, the gravitational degrees of freedom are initially described by the metric g_{ij} and its conjugate momentum π^{ij} . We systematically use the Iwasawa decomposition (4.5) of the metric to replace the $d(d+1)/2$ variables g_{ij} by the $d + d(d-1)/2$ variables $(\beta^a, \mathcal{N}^a_i)$. Note that $(\beta^a, \mathcal{N}^a_i)$ are *ultralocal* functions of g_{ij} , that is they depend, at each spacetime point, only on the value of g_{ij} at that point, not

on its derivatives. This would not have been the case if we had used a ‘Kasner frame’ (as defined below) instead of an Iwasawa one. The transformation $g \rightarrow (\beta, \mathcal{N})$ then defines a corresponding transformation of the conjugate momenta, as we will explain below. We then augment the definition of the β by adding the dilaton field, i.e., $\beta^\mu \equiv (\beta^a, \phi)$, and define the hyperbolic radial coordinate ρ as in (3.33). Note that ρ is also an ultralocal function of the configuration variables (g_{ij}, ϕ) . We assume that the hyperbolic coordinate ρ can be used everywhere in a given region of space near the singularity as a well-defined (real) quantity which tends to $+\infty$ as we approach the singularity. We then define the slicing of spacetime by imposing the gauge condition (2.5). The time coordinate corresponding to this gauge is called T as above (see (2.7)). Our aim is to study the asymptotic behaviour of all the dynamical variables $\beta(T), \mathcal{N}(T), \dots$ as $T \rightarrow +\infty$ (recall that this limit also corresponds to $t \rightarrow 0, \sqrt{g} \rightarrow 0, \rho \rightarrow +\infty$, with β^μ going to infinity inside the future light cone). Of course, we must also ascertain the self-consistency of this limit, which we shall refer to as the ‘BKL limit’.

5.1. Hamiltonian action

To focus on the features relevant to the billiard picture, we assume first that there are no Chern–Simons terms or couplings of the exterior form gauge fields through a modification of the curvatures $F^{(p)}$, which are thus taken to be Abelian, $F^{(p)} = dA^{(p)}$. We verify in section 6.5 that these interaction terms do not change the analysis. The Hamiltonian action in any pseudo-Gaussian gauge, and in the temporal gauge (2.9), reads

$$S[g_{ij}, \pi^{ij}, \phi, \pi_\phi, A_{j_1 \dots j_p}^{(p)}, \pi_{(p)}^{j_1 \dots j_p}] = \int dx^0 \int d^d x \left(\pi^{ij} g_{ij} + \pi_\phi \dot{\phi} + \frac{1}{p!} \sum_p \pi_{(p)}^{j_1 \dots j_p} \dot{A}_{j_1 \dots j_p}^{(p)} - H \right) \tag{5.1}$$

where the Hamiltonian density H is

$$H \equiv \tilde{N} \mathcal{H} \tag{5.2}$$

$$\mathcal{H} = \mathcal{K} + \mathcal{M} \tag{5.3}$$

$$\mathcal{K} = \pi^{ij} \pi_{ij} - \frac{1}{d-1} \pi^i{}_i \pi^j{}_j - \frac{1}{4} \pi_\phi^2 + \sum_p \frac{e^{-\lambda_p \phi}}{2p!} \pi_{(p)}^{j_1 \dots j_p} \pi_{(p)j_1 \dots j_p} \tag{5.4}$$

$$\mathcal{M} = -gR + g g^{ij} \partial_i \phi \partial_j \phi + \sum_p \frac{e^{\lambda_p \phi}}{2(p+1)!} g F_{j_1 \dots j_{p+1}}^{(p)} F^{(p)j_1 \dots j_{p+1}} \tag{5.5}$$

where R is the spatial curvature scalar. The dynamical equations of motion are obtained by varying the above action w.r.t. the spatial metric components, the dilaton, the spatial p -form components and their conjugate momenta. In addition, there are constraints on the dynamical variables,

$$\mathcal{H} \approx 0 \quad \text{‘Hamiltonian constraint’}, \tag{5.6}$$

$$\mathcal{H}_i \approx 0 \quad \text{‘momentum constraint’}, \tag{5.7}$$

$$\varphi_{(p)}^{j_1 \dots j_{p-1}} \approx 0 \quad \text{‘Gauss law’ for each } p\text{-form} \tag{5.8}$$

with

$$\mathcal{H}_i = -2\pi^j{}_{i|j} + \pi_\phi \partial_i \phi + \sum_p \frac{1}{p!} \pi_{(p)}^{j_1 \dots j_p} F_{ij_1 \dots j_p}^{(p)} \quad (5.9)$$

$$\varphi_{(p)}^{j_1 \dots j_{p-1}} = \pi_{(p)}^{j_1 \dots j_{p-1} j_p} \Big|_{j_p} \quad (5.10)$$

where the subscript $|j$ stands for the spatially covariant derivative.

Let us now see how the Hamiltonian action gets transformed when one performs, at each spatial point, the Iwasawa decomposition (4.3), (4.5) of the spatial metric. The ‘supermetric’ (3.11) giving the kinetic terms of the metric and of the dilaton then becomes

$$d\sigma^2 = \text{Tr} d\beta^2 - (\text{Tr} d\beta)^2 + d\phi^2 + \frac{1}{2} \text{Tr}[\mathcal{A}^2 (d\mathcal{N}\mathcal{N}^{-1}) \mathcal{A}^{-2} (d\mathcal{N}\mathcal{N}^{-1})^T] \quad (5.11)$$

i.e.,

$$d\sigma^2 = \sum_{a=1}^d (d\beta^a)^2 - \left(\sum_{a=1}^d d\beta^a \right)^2 + d\phi^2 + \frac{1}{2} \sum_{a<b} e^{2(\beta^b - \beta^a)} (d\mathcal{N}^a{}_i \mathcal{N}^i{}_b)^2 \quad (5.12)$$

where we recall that $\mathcal{N}^i{}_a$ denotes, as in (4.9), the inverse of the triangular matrix $\mathcal{N}^a{}_i$ appearing in the Iwasawa decomposition (4.5) of the spatial metric g_{ij} . For $d = 3$, this expression reduces to that of [51]. This change of variables corresponds to a point canonical transformation, which can be extended to the momenta in the standard way via

$$\pi^{ij} \dot{g}_{ij} \equiv \sum_a \pi_a \dot{\beta}^a + \sum_a P^i{}_a \dot{\mathcal{N}}^a{}_i. \quad (5.13)$$

Note that the momenta

$$P^i{}_a = \frac{\partial \mathcal{L}}{\partial \dot{\mathcal{N}}^a{}_i} = \sum_b e^{2(\beta^b - \beta^a)} \dot{\mathcal{N}}^a{}_j \mathcal{N}^j{}_b \mathcal{N}^i{}_b \quad (5.14)$$

conjugate to the non-constant off-diagonal Iwasawa components $\mathcal{N}^a{}_i$ are only defined for $a < i$; hence the second sum in (5.13) receives only contributions from $a < i$.

We next split the Hamiltonian density⁹ \mathcal{H} (5.2) in two parts, one denoted by \mathcal{H}_0 , which is the kinetic term for the local scale factors β^μ (including dilatons) already encountered in section 3, and a ‘potential density’ (of weight 2) denoted by \mathcal{V} , which contains everything else. Our analysis below will show why it makes sense to group the kinetic terms of both the off-diagonal metric components and the p -forms with the usual potential terms, i.e., the term \mathcal{M} in (5.2). (Remembering that, in a Kaluza–Klein reduction, the off-diagonal components of the metric in one dimension higher become a 1-form, it is not surprising that it might be useful to group together the off-diagonal components and the p -forms.) Thus, we write

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{V} \quad (5.15)$$

with the kinetic term of the β variables

$$\mathcal{H}_0 = \frac{1}{4} G^{\mu\nu} \pi_\mu \pi_\nu \quad (5.16)$$

where the rhs is that already defined in (3.18), with the replacement of the coordinate index i by the frame index a . The total (weight 2) potential density,

$$\mathcal{V} = \mathcal{V}_S + \mathcal{V}_G + \sum_p \mathcal{V}_p + \mathcal{V}_\phi, \quad (5.17)$$

⁹ We use the term ‘Hamiltonian density’ to denote both H and \mathcal{H} . Note that H is a usual spatial density (of weight 1, i.e., the same weight as \sqrt{g}), while $\mathcal{H} \equiv \sqrt{g}H/N$ is a density of weight 2 (like $g = (\sqrt{g})^2$). Note also that π^{ij} is of weight 1, while $\tilde{N} \equiv N/\sqrt{g}$ is of weight -1 .

is naturally split into a centrifugal part linked to the kinetic energy of the off-diagonal components (the index ‘ S ’ referring to ‘symmetry’, as discussed below)

$$\mathcal{V}_S = \frac{1}{2} \sum_{a < b} e^{-2(\beta^b - \beta^a)} (P^j{}_a \mathcal{N}^b{}_j)^2, \quad (5.18)$$

a ‘gravitational’ (or ‘curvature’) potential

$$\mathcal{V}_G = -gR, \quad (5.19)$$

and a term from the p -forms,

$$\mathcal{V}_{(p)} = \mathcal{V}_{(p)}^{\text{el}} + \mathcal{V}_{(p)}^{\text{magn}} \quad (5.20)$$

which is a sum of an ‘electric’ and a ‘magnetic’ contribution

$$\mathcal{V}_{(p)}^{\text{el}} = \frac{e^{-\lambda_p \phi}}{2p!} \pi_{(p)}^{j_1 \dots j_p} \pi_{(p)j_1 \dots j_p} \quad (5.21)$$

$$\mathcal{V}_{(p)}^{\text{magn}} = \frac{e^{\lambda_p \phi}}{2(p+1)!} g F_{j_1 \dots j_{p+1}}^{(p)} F^{(p)j_1 \dots j_{p+1}}. \quad (5.22)$$

Finally, there is a contribution to the potential linked to the spatial gradients of the dilaton:

$$\mathcal{V}_\phi = g g^{ij} \partial_i \phi \partial_j \phi. \quad (5.23)$$

We will analyse in detail these contributions to the potential, term by term, in section 6.

5.2. Appearance of sharp walls in the BKL limit

5.2.1. Derivation of central result. In the decomposition of the Hamiltonian given above, we have split off the kinetic terms of the scale factors β^a and of the dilaton $\beta^{d+1} \equiv \phi$ from the other variables, and assigned the off-diagonal metric components and the p -form fields to various potentials, each of which is a complicated function of β^μ , $\mathcal{N}^a{}_i$, $P^i{}_a$, $A_{j_1 \dots j_p}^{(p)}$, $\pi_{(p)}^{j_1 \dots j_p}$ and of some of their spatial gradients. The reason why this separation is useful is that, as we show, in the BKL limit, and in the special Iwasawa decomposition which we have adopted, the asymptotic dynamics is governed by the scale factors β^μ , whereas all other variables ‘freeze’, just like for the general Kasner solution discussed in section 3. Thus, in the asymptotic limit, we have schematically

$$\mathcal{V}(\beta^\mu, \mathcal{N}^a{}_i, P^i{}_a, A_{j_1 \dots j_p}^{(p)}, \pi_{(p)}^{j_1 \dots j_p}, \dots) \longrightarrow \mathcal{V}_\infty(\gamma^\mu) \quad (5.24)$$

where $\mathcal{V}_\infty(\gamma^\mu)$ stands for a sum of certain ‘sharp wall potentials’ which depend only on the angular hyperbolic coordinates $\gamma^\mu \equiv \beta^\mu / \rho$. As a consequence, the asymptotic dynamics can be described as a ‘billiard’ in the hyperbolic space of the γ^μ , whose walls (or ‘cushions’) are determined by the energy of the fields that are asymptotically frozen.

This reduction of the complicated potential to a much simpler ‘effective potential’ $\mathcal{V}_\infty(\gamma^\mu)$ follows essentially from the exponential dependence of \mathcal{V} on the diagonal Iwasawa variables β^μ , from its independence from the conjugate momenta of the β , and from the fact that the radial magnitude ρ of the β becomes infinitely large in the BKL limit.

To see the essence of this reduction, with a minimum of technical complications, let us consider a potential density (of weight 2) of the general form

$$\mathcal{V}(\beta, Q, P) = \sum_A c_A(Q, P) \exp(-2w_A(\beta)) \quad (5.25)$$

where (Q, P) denote the remaining phase space variables (that is, other than (β, π_β)). Here $w_A(\beta) = w_{A\mu} \beta^\mu$ are certain linear forms which depend only on the (extended) scale factors,

and whose precise form will be derived in the following section. Similarly, we shall discuss below the explicit form of the pre-factors c_A , which will be some complicated polynomial functions of the remaining fields, i.e., the off-diagonal components of the metric, the p -form fields and their respective conjugate momenta, and of some of their spatial gradients. The fact that the $w_A(\beta)$ depend *linearly* on the scale factors β^μ is an important property of the models under consideration. A second non-trivial fact is that, *for the leading contributions*, the pre-factors are always non-negative, i.e., $c_A^{\text{leading}} \geq 0$. Since the values of the fields for which $c_A = 0$ constitute a set of measure zero, we will usually make the ‘genericity assumption’ $c_A > 0$ for the leading terms in the potential \mathcal{V} .¹⁰ The third fact following from the detailed analysis of the walls that we shall exploit is that all the leading walls are *timelike*, i.e., their normal vectors (in the Minkowski β -space) are spacelike.

As shown in section 3.3, the part of the Hamiltonian describing the kinetic energy of the β , $H_0 = \tilde{N}\mathcal{H}_0$, takes the form (3.36) when parametrizing β^μ in terms of ρ and γ^μ , or equivalently, $\lambda \equiv \ln \rho$ and γ^μ (cf equation (3.37)). Choosing the gauge (2.5) to simplify the kinetic terms H_0 we end up with an Hamiltonian of the form

$$H(\lambda, \pi_\lambda, \gamma, \pi_\gamma, Q, P) = \tilde{N}\mathcal{H} = \frac{1}{4}[-\pi_\lambda^2 + \pi_\gamma^2] + \rho^2 \sum_A c_A(Q, P) \exp(-2\rho w_A(\gamma)) \quad (5.26)$$

where π_γ^2 is the kinetic energy of a particle moving on H_m . In (5.26) and below we shall regard λ as a primary dynamical variable (so that $\rho \equiv e^\lambda$).

The essential point now is that, in the BKL limit, $\lambda \rightarrow +\infty$, i.e., $\rho \rightarrow +\infty$, each term $\rho^2 \exp(-2\rho w_A(\gamma))$ becomes a sharp wall potential, i.e., a function of $w_A(\gamma)$ which is zero when $w_A(\gamma) > 0$, and $+\infty$ when $w_A(\gamma) < 0$. To formalize this behaviour, we define the sharp wall Θ -function¹¹ as

$$\Theta(x) := \begin{cases} 0 & \text{if } x < 0 \\ +\infty & \text{if } x > 0. \end{cases} \quad (5.27)$$

A basic formal property of this Θ -function is its invariance under multiplication by a positive quantity. With the above assumption checked below that all the relevant pre-factors $c_A(Q, P)$ are *positive* near each leading wall, we can formally write

$$\lim_{\rho \rightarrow \infty} [c_A(Q, P) \rho^2 \exp(-2\rho w_A(\gamma))] = c_A(Q, P) \Theta(-2w_A(\gamma)) \equiv \Theta(-2w_A(\gamma)). \quad (5.28)$$

Of course $\Theta(-2w_A(\gamma)) = \Theta(-w_A(\gamma))$, but we shall keep the extra factor of 2 to recall that the arguments of the exponentials, from which the Θ -functions originate, come with a well-defined normalization. Therefore, the limiting Hamiltonian density reads

$$H_\infty(\lambda, \pi_\lambda, \gamma, \pi_\gamma, Q, P) = \frac{1}{4}[-\pi_\lambda^2 + \pi_\gamma^2] + \sum_{A'} \Theta(-2w_{A'}(\gamma)), \quad (5.29)$$

where A' runs over the *dominant walls*. The set of dominant walls is defined as the *minimal set* of wall forms which suffice to define the billiard table, i.e., such that the restricted set of inequalities $\{w_{A'}(\gamma) \geq 0\}$ imply the full set $\{w_A(\gamma) \geq 0\}$. The concept of dominant wall will be illustrated below. (Note that the concept of ‘dominant’ wall is a refinement of the distinction, which will also enter our discussion, between a leading wall and a subleading one.)

¹⁰ Understanding the effects of the possible failure of this assumption is one of the subtle issues in establishing a rigorous proof of the BKL picture.

¹¹ One should more properly write $\Theta_\infty(x)$, but since this is the only step function encountered in this paper, we use the simpler notation $\Theta(x)$.

The crucial point is that the limiting Hamiltonian (5.29) no longer depends on λ , Q and P . Therefore the Hamiltonian equations of motion for λ , Q and P tell us that the corresponding conjugate momenta, i.e., π_λ , P and Q , respectively, all become *constants of the motion* in the limit $\lambda \rightarrow +\infty$. The total Hamiltonian density H_∞ is also a constant of the motion (which must be set to zero). The variable λ evolves according to $d\lambda/dT = -\frac{1}{2}\pi_\lambda$. Hence, in the limit, λ is a linear function of T . The only non-trivial dynamics resides in the evolution of (γ, π_γ) which reduces to the sum of a free (non-relativistic) kinetic term $\pi_\gamma^2/4$ and a sum of sharp wall potentials, such that the resulting motion of the γ indeed constitutes a billiard, with geodesic motion on the unit hyperboloid H_m interrupted by reflections on the walls defined by $w_A(\gamma) = 0$. These walls are hyperplanes (in the sense of hyperbolic geometry) because they are geometrically given by the intersection of the unit hyperboloid $\beta^\mu\beta_\mu = -1$ with the usual Minkowskian hyperplanes $w_A(\beta) = 0$.

We note in passing that an alternative route for reducing the dynamics to a billiard in γ -space would be to eliminate the variable λ by solving the Hamiltonian constraint for π_λ . This allows one to use λ as a time variable, and is similar to going from the quadratic form of the action of a relativistic particle to its ‘square-root form’,

$$\begin{aligned} H_\lambda &= \left(\pi_\gamma^2/4 + \sum_A c_A(Q, P)\rho^2 \exp(-2\rho w_A(\gamma)) \right)^{1/2} \longrightarrow \left(\pi_\gamma^2/4 + \sum_{A'} \Theta(-2w_{A'}(\gamma)) \right)^{1/2} \\ &\equiv \frac{1}{2}|\pi_\gamma| + \sum_{A'} \Theta(-2w_{A'}(\gamma)). \end{aligned} \quad (5.30)$$

As the above derivation of the asymptotic constancy of all the phase space variables (Q, P) may seem a bit formal, we study, in the appendix, a simplified model explicitly showing how this asymptotic constancy arises. This toy model also exemplifies the residual, asymptotically decaying variations of (Q, P) as $\rho \rightarrow \infty$.

5.2.2. Finite volume versus infinite volume. Geodesic motion in a billiard in hyperbolic space has been much studied. It is known that this motion is chaotic or non-chaotic depending on whether the billiard has finite or infinite volume [36, 52, 81, 102]. In the finite volume case, the generic evolution exhibits an infinite number of collisions with the walls with strong chaotic features (‘oscillating behaviour’).

By contrast, if the billiard has infinite volume, the evolution is non-chaotic. For a generic evolution, there are only finitely many collisions with the walls. The system generically settles after a finite time in a Kasner-like motion that lasts all the way to the singularity.

5.2.3. β -space description. The above derivation relied on the use of hyperbolic polar coordinates (ρ, γ) . This use is technically useful in that it represents the walls as being located at an asymptotically fixed position in hyperbolic space, namely $w_A(\gamma) = 0$. However, once one has derived the final result (5.29), one can reexpress it in terms of the original variables β^μ , which run over a linear (Minkowski) space. Owing to the linearity of $w_A(\beta) = \rho w_A(\gamma)$ in this Minkowskian picture, the asymptotic motion takes place in a ‘polywedge’, bounded by the hyperplanes $w_A(\beta) = 0$. The billiard motion then consists of free motions of β^μ on straight lightlike lines within this polywedge, which are interrupted by specular reflections off the walls. (See formula (7.3) for the explicit effect of these reflections on the components of the velocity vector of the β -particle.) Indeed, when going back to β -space (i.e., before

taking the BKL limit), the dynamics of the scale factors at each point of space is given by the Hamiltonian

$$H(\beta^\mu, \pi_\mu) = \tilde{N}\mathcal{H} = \tilde{N} \left[\frac{1}{4} G^{\mu\nu} \pi_\mu \pi_\nu + \sum_A c_A \exp(-2w_A(\beta)) \right]. \quad (5.31)$$

The β -space dynamics simplifies in the gauge $\tilde{N} = 1$, corresponding to the time coordinate τ . In the BKL limit, the Hamiltonian (5.31) takes the limiting form (in the gauge $\tilde{N} = 1$)

$$H_\infty(\beta^\mu, \pi_\mu) = \frac{1}{4} G^{\mu\nu} \pi_\mu \pi_\nu + \sum_{A'} \Theta(-2w_{A'}(\beta)) \quad (5.32)$$

where the sum is again only over the dominant walls. When taking equal time slices of this polywedge (e.g., slices on which $\sum_i \beta^i$ is constant), it is clear that with increasing time (i.e., increasing $\sum_i \beta^i$, or increasing τ or ρ) the walls recede from the observer. The β -space picture is useful for simplifying the mathematical representation of the dynamics of the scale factors which takes place in a linear space. However, it is inconvenient both for proving that the exponential walls of (5.31) do reduce, in the large $\sum_i \beta^i$ limit to sharp walls, and for dealing with the dynamics of the other phase space variables (Q, P) , whose appearance in the coefficients c_A has been suppressed in (5.31). Let us only mention that, in order to prove, in this picture, the freezing of the phase space variables (Q, P) one must consider in detail the accumulation of the ‘redshifts’ of the energy–momentum π_μ of the β -particle when it undergoes reflections on the receding walls, and the effect of the resulting decrease of the magnitudes of the components of π_μ on the evolution of (Q, P) . (In the notation of the appendix, it is important to take into account the fact that p_0 decreases with time.) By contrast, the γ -space picture that we used above allows for a more streamlined treatment of the effect of the limit $\rho \rightarrow +\infty$ on the sharpening of the walls and on the freezing of (Q, P) .

In summary, the dynamics simplifies enormously in the asymptotic limit. It becomes ultralocal in that it reduces to a continuous superposition of evolution systems (depending only on a time parameter) for the scale factors and the dilatons, at each spatial point, with asymptotic freezing of the off-diagonal and p -form variables. This ultralocal description of the dynamics is valid only asymptotically. It would make no sense to speak of a billiard motion prior to this limit, because one cannot replace the exponentials by Θ -functions. Prior to this limit, the evolution system for the scale factors involves the coefficients in front of the exponential terms, and the evolution of these coefficients depends on various spatial gradients of the other degrees of freedom. However, one may contemplate setting up an expansion in which the sharp wall model is replaced by a model with exponential (‘Toda-like’) potentials, and where the evolution of the quantities entering the coefficients of these ‘Toda walls’ is treated as a next to leading effect. See [28] for the definition of the first steps of such an expansion scheme for maximal supergravity in 11 dimensions.

5.3. Kasner frames versus Iwasawa frames

At this point, we want to clarify an issue that might at first sight seem paradoxical to the reader. In most of the BKL literature (notably in the original analysis of [6, 8]), one tries to construct the metric in frames equal (or close) to ‘Kasner frames’¹². These are defined as frames with respect to which both the spatial metric and the extrinsic curvature are diagonal.

¹² We know, however, no detailed development of the Hamiltonian formalism within such a frame whose definition involves *both* the metric variables and their conjugate momenta.

Once the time slicing has been fixed, this geometric frame is unique, up to *time-dependent rescalings* of each basis vector, when the eigenvalues of the extrinsic curvature are distinct and some definite ordering of the (time-independent) eigenvalues p_a has been adopted. It is not clear how to fix in a rigorous (and useful) manner the arbitrary time-dependent rescalings of each basis vector. However, this can be done (following [7]) in an approximate manner by considering the evolution of the geometry at each spatial point as a succession of free Kasner flights interrupted by collisions against the walls. More precisely, one can uniquely fix the normalization of the Kasner frame by imposing the two requirements that (i) the Kasner frame be *time independent* during each Kasner epoch, i.e., the Lie derivative of the frame vectors along $\partial/\partial t$ be zero, and (ii) the linear transformation between two successive Kasner frames has the special form given in (5.35) up to some time-independent rescalings.

More explicitly, if we consider the three-dimensional case for definiteness and denote the covariant components of the frame by $\{l_i, m_i, r_i\}$, one has (suppressing the x -dependence)

$$g_{ij}(t) = A_1^2(t)l_i l_j + A_2^2(t)m_i m_j + A_3^2(t)r_i r_j \tag{5.33}$$

during a certain Kasner epoch, and

$$g_{ij}(t) = A_1^2(t)l'_i l'_j + A_2^2(t)m'_i m'_j + A_3^2(t)r'_i r'_j \tag{5.34}$$

during the subsequent Kasner epoch. Here the two successive Kasner frames $\{l_i, m_i, r_i\}$, $\{l'_i, m'_i, r'_i\}$ are (in this approximation) independent of time, while the scale factors $A_a(t)$ vary like a power law during each Kasner free flight (say $A_a(t) \approx b_a t^{p_a}$ during the first epoch, and $A_a(t) \approx b'_a t^{p'_a}$ during the next, with some interpolating behaviour during the collision). Belinskii *et al* [7] have argued (by studying the effect of one collision in the ‘incoming’ frame $\{l_i, m_i, r_i\}$, and by re-diagonalizing the ‘outgoing’ metric) that the transformation between the two successive Kasner frames could be written as

$$l'_i = l_i, \quad m'_i = m_i + \sigma_m l_i, \quad r'_i = r_i + \sigma_r l_i. \tag{5.35}$$

Here, σ_m and σ_r are quantities which can *a priori* be of order unity.

The seeming paradox is that according to [7] the σ_m and σ_r do not get smaller for collisions closer to the singularity, so that the transformation (5.35) from the old Kasner axes l to the new Kasner axes l' is of order 1, no matter how close one gets to the singularity; therefore, the Kasner axes generically never ‘come to rest’ if there is an infinite number of collisions. On the other hand, as we saw, the Iwasawa frames become approximately time independent for asymptotic values of τ , and the extrinsic curvature approximately diagonal.

We now show that there is no contradiction between the oscillatory behaviour of the Kasner axes (in the generic inhomogeneous case) and the asymptotic freezing of the off-diagonal components \mathcal{N} in the Iwasawa frame. The key point is the very restricted form of the change (5.35) of the covariant components during a collision. Recall first that the covariant components of the Iwasawa frame are given by the three covectors appearing as the three lines of the matrix \mathcal{N} , i.e. $\theta^1 = dx^1 + n_1 dx^2 + n_2 dx^3$, $\theta^2 = dx^2 + n_3 dx^3$, $\theta^3 = dx^3$. (Note also in passing that, contrary to the Kasner frames, the Iwasawa frames are not geometrically uniquely defined since one can redefine the coordinates x^i .)

Let us now see what (5.35) implies for the change in the Iwasawa variables. For doing this we need to relate these to the l_i, m_i, n_i . This was done in section 4.2 for the simple exact Kasner solution and is easy to generalize to the case of an oscillatory metric. Indeed, when one is not in the ‘collision region’ (and this can be applied both *before* and *after* the specific collision under consideration in (5.35)), the coordinate components of the metric take the form (5.33) or (5.34) with $A_2^2 \ll A_1^2$ and $A_3^2 \ll A_2^2$. As the definition of the Iwasawa components n_a is purely algebraic, we can apply the formulae (4.16), (4.17) of section 4.2 to the present

case. It suffices to replace t^{p_a} by A_a everywhere. Then, it is easy to see that n_1, n_2 and n_3 are still given by the same final formula as above, i.e.,

$$n_1 = \frac{l_2}{l_1}, \quad n_2 = \frac{l_3}{l_1}, \quad n_3 = \frac{l_1 m_3 - l_3 m_1}{l_1 m_2 - l_2 m_1}, \quad (5.36)$$

(before the collision) and

$$n'_1 = \frac{l'_2}{l'_1}, \quad n'_2 = \frac{l'_3}{l'_1}, \quad n'_3 = \frac{l'_1 m'_3 - l'_3 m'_1}{l'_1 m'_2 - l'_2 m'_1}, \quad (5.37)$$

(after the collision). If we substitute in this second formula l'_i, m'_i and r'_i in terms of l_i, m_i and r_i according to (5.35), we obtain the *same* values for the Iwasawa off-diagonal variables n_1, n_2 and n_3 , before and after the collision, namely, $n'_1 = n_1, n'_2 = n_2, n'_3 = n_3$. There is thus no contradiction between the change of Kasner axes (5.35) and the freezing of the off-diagonal Iwasawa variables. The same conclusion holds for collisions against the other types of walls, where the Kasner axes ‘rotate’ as in (5.35).

5.4. Constraints

We have just seen that in the BKL limit, the evolution equations become ordinary differential equations with respect to time. Although the spatial points are decoupled in the evolution equations, they are, however, still coupled via the constraints. These constraints just restrict the initial data and need only be imposed at one time, since they are preserved by the dynamical equations of motion. Indeed, one easily finds that, in the BKL limit,

$$\dot{\mathcal{H}} = 0 \quad (5.38)$$

since $[\mathcal{H}(x), \mathcal{H}(x')] = 0$ in the ultralocal limit. This corresponds simply to the fact that the collisions preserve the lightlike character of the velocity vector. Furthermore, the gauge constraints (5.8) are also preserved in time since the Hamiltonian constraint is gauge invariant. In the BKL limit, the momentum constraint fulfils

$$\dot{\mathcal{H}}_k(x) = \partial_k \mathcal{H} \approx 0. \quad (5.39)$$

It is important that the restrictions on the initial data do not bring dangerous constraints on the coefficients of the walls in the sense that these may all take non-zero values. For instance, it is well known that it is consistent with the Gauss law to take non-vanishing electric and magnetic energy densities; thus the coefficients of the electric and magnetic walls are indeed generically non-vanishing. In fact, the constraints are essentially conditions on the spatial gradients of the variables entering the wall coefficients, not on these variables themselves. In some non-generic contexts, however, the constraints could force some of the wall coefficients to be zero; the corresponding walls would thus be absent. (For example, for vacuum gravity in four dimensions, the momentum constraints for some Bianchi homogeneous models force some symmetry wall coefficients to vanish. But this is peculiar to the homogeneous case.)

It is easy to see that the number of arbitrary physical functions involved in the solution of the asymptotic BKL equations of motion is the same as in the general solution of the complete Einstein-matter equations. Indeed, the number of constraints on the initial data and the residual gauge freedom are the same in both cases. Further discussion of the constraints in the BKL context may be found in [2, 29].

5.5. Consistency of BKL behaviour in spite of the increase of spatial gradients

The essential assumption in the BKL analysis, also made in the present paper, is the asymptotic dominance of time derivatives with respect to space derivatives near a spacelike singularity.

This assumption has been mathematically justified, in a rigorous manner, in the cases where the billiard is of infinite volume, i.e., in the (simple) cases where the asymptotic behaviour is not chaotic, but is monotonically Kasner-like [2, 29].

On the other hand, one might *a priori* worry that this assumption is self-contradictory in those cases where the billiard is of finite volume, when the asymptotic behaviour is chaotic, with an infinite number of oscillations. Indeed, it has been pointed out [4, 72] that the independence of the billiard evolution at each spatial point will have the effect of *infinitely increasing* the spatial gradients of various quantities, notably of the local values of the Kasner exponents $p_\mu(x)$. This increase of spatial gradients towards the singularity has been described as a kind of turbulent behaviour of the gravitational field, in which energy is pumped into shorter and shorter length scales [4, 72], and, if it were too violent, it would certainly work against the validity of the BKL assumption of asymptotic dominance of time derivatives. For instance, in our analysis of gravitational walls in the following section, we will encounter subleading walls, whose pre-factors depend on spatial gradients of the logarithmic scale factors β .

To address the question of consistency of the BKL assumption we need to know how fast the spatial gradients of β , and of similar quantities entering the pre-factors, grow near the singularity. Let us consider the spatial gradient of $\beta \equiv \rho\gamma$, which is

$$\partial_i \beta = \rho \partial_i \gamma + \rho \gamma \partial_i \lambda. \quad (5.40)$$

As we work here in the gauge (2.5), the spatial derivatives must be taken with fixed T . We know that, asymptotically, λ is a linear function of T , i.e., $\lambda = a(x)T + b(x)$ where $a(x) = -\frac{1}{2}\pi_\lambda(x)$ is linked to the (spatially dependent) conjugate momentum of λ . Therefore the spatial gradient $\partial_i \lambda = T \partial_i a(x) + \partial_i b(x)$ behaves linearly in T , so that $\partial_i \lambda \propto \lambda \equiv \ln \rho$. The second term in (5.40) consequently behaves as $\gamma \rho \ln \rho$ when $\rho \rightarrow \infty$. Let us now estimate the first term (which will turn out to dominate the sum).

To estimate $\partial_i \gamma$ we can use the standard results on billiards on hyperbolic space. Indeed, $dx^i \partial_i \gamma$ can be thought of as the infinitesimal deviation between two billiard trajectories. The chaotic behaviour of the billiard implies that this deviation will grow exponentially with T . In fact, because we work on H_m with curvature -1 the Liapunov exponent for billiard trajectories is equal to 1, when using geodesic length as time parameter. Remembering the Hamiltonian constraint $\pi_\lambda^2 = \pi_\gamma^2$, the geodesic length is simply equal to λ . This yields the growth estimate

$$\partial_i \gamma = \mathcal{O}(1) e^\lambda = \mathcal{O}(1) \rho \quad (5.41)$$

where the coefficient $\mathcal{O}(1)$ is a chaotically oscillating quantity. This estimate is not affected by the collisions against the walls because these preserve the angles made by neighbouring trajectories (the walls are hyperplanes). Inserting (5.41) in (5.40), we see that the first term indeed dominates the second. We conclude that the chaotic character of the billiard indeed implies an unlimited growth of the spatial gradients of β , but that this growth is only of *polynomial* order in ρ

$$\partial_i \beta = \mathcal{O}(1) \rho^2. \quad (5.42)$$

This polynomial growth of $\partial_i \beta$ (and of its second-order spatial derivatives) entails a polynomial growth of the pre-factors of the subdominant walls in section 6.2. Because it is polynomial (in ρ), this growth is, however, negligible compared to the *exponential* (in ρ) behaviour of the various potential terms. It does not jeopardize our reasoning based on keeping track of the various exponential behaviours. As we will see, the potentially dangerously growing terms that we have controlled here appear only in subdominant walls. The reasoning of the appendix shows that the pre-factors of the dominant walls are self-consistently predicted to evolve very little near the singularity.

We conclude that the unlimited growth of some of the spatial gradients does not affect the consistency of the BKL analysis done here. This does not mean, however, that it will be easy to promote our analysis to a rigorous mathematical proof. The main obstacle to such a proof appears to be the existence of exceptional points, where a pre-factor of a dominant wall happens to vanish, or points where a subdominant wall happens to be comparable to a dominant one. Though the set of such exceptional points is (generically) of measure zero, their density might increase near the singularity because of the increasing spatial gradients. This situation might be compared to the KAM (Kolmogorov–Arnold–Moser) one, where the ‘bad’ tori have a small measure, but are interspersed densely among the ‘good’ ones.

5.6. BKL limit versus other limits

The last issue which we wish to address in this section concerns the relation of the BKL limit to other limits considered in the literature, such as the strong coupling limit or the ‘small tension limit’, as well as the relation with asymptotically velocity dominated solutions.

It is sometimes useful to separate the time derivatives (conjugate momenta) in the Hamiltonian from the space derivatives, namely,

$$\mathcal{H} = \mathcal{K}' + \varepsilon \mathcal{V}' \quad (5.43)$$

where $\varepsilon = \pm 1$ according to whether the spacetime signature is Lorentzian ($\varepsilon = 1$) or Euclidean ($\varepsilon = -1$). Here,

$$\mathcal{K}' = \mathcal{H}_0 + \mathcal{V}_S + \mathcal{V}_{(p)}^{\text{el}} \quad (5.44)$$

contains all the kinetic terms, and

$$\mathcal{V}' = \mathcal{V}_G + \mathcal{V}_{(p)}^{\text{magn}} + \mathcal{V}_\phi \quad (5.45)$$

the terms with spatial derivatives. We stress that this split is different from that introduced in section 5.1, where only the kinetic terms of the scale factors and the dilatons were kept as such. The above split is useful because for some models the asymptotic dynamics is entirely controlled by \mathcal{K}' , i.e., by the limit $\varepsilon = 0$. This happens whenever the billiard that emerges in the BKL limit is defined by the symmetry and electric walls, as, for instance, for $D = 11$ supergravity [24] or the pure Einstein–Maxwell system in spacetime dimensions $D \geq 5$ [25, 74]. Curvature and magnetic walls are then subdominant, i.e., spatial gradients become negligible as one approaches the singularity.

If the curvature and magnetic walls can be neglected, the evolution equations are exactly the same as the equations of motion obtained by performing a direct torus reduction to $1+0$ dimensions. We stress, however, that no homogeneity assumption has to be made here. *The effective torus-dimensional reduction follows from the dynamics and is not imposed manually.*

The limit $\varepsilon = 0$ is known as the ‘zero signature limit’ [98] and lies half-way between spacetimes of Minkowskian or Lorentzian signature. It corresponds to a vanishing velocity of light (or vanishing ‘medium tension’); the underlying geometry is built on the Carroll contraction of the Lorentz group [50]. The terminology ‘strong coupling’ is also used [55] and stems from the fact that, with appropriate rescalings, \mathcal{H} can be rewritten as

$$\mathcal{H} = G_N \mathcal{K}' + G_N^{-1} \mathcal{V}' \quad (5.46)$$

such that the limit in question corresponds to large values of Newton’s constant G_N . The interest in this ultrarelativistic (‘Carrollian’) limit has recently been revived in [1, 30, 46, 78].

If there are only finitely many collisions with the walls (corresponding to a billiard of infinite volume) the dynamics in the vicinity of the singularity becomes even simpler. After the last collision, the asymptotic dynamics is controlled solely by the kinetic energy \mathcal{H}_0 of the

scale factors. This case where both spatial gradients and matter (here p -form) terms can be neglected, has been called ‘asymptotically velocity-dominated’ (or ‘AVD’) in [35] and allows a rigorous analysis of its asymptotic dynamics by means of Fuchsian techniques [2, 29]. By contrast, rigorous results are rare for the case of infinitely many collisions (see, however, the recent analytic advances in [92]). Besides the existing rigorous results, there also exists a wealth of numerical support for the BKL ideas [9, 10].

6. Walls

The decomposition (5.15) of the Hamiltonian gives rise to different types of walls, which we now discuss in turn. Specifically, we will derive explicit formulae for the linear forms $w_A(\beta)$ and the field dependence of the pre-factors c_A entering the various potentials.

6.1. Centrifugal (or symmetry) walls

We start by analysing the effects of the off-diagonal metric components which will give rise to the so-called ‘symmetry walls’. As they originate from the gravitational action they are always present. The relevant contribution to the potential is the centrifugal potential (5.18). When comparing (5.18) to the general form (5.25) analysed above, we see that firstly the summation index A must be interpreted as a double index (a, b) , with the restriction $a < b$, secondly that the corresponding pre-factor is $c_{ab} = (P^j_a \mathcal{N}^b_j)^2$ is automatically non-negative (in accordance with our genericity assumptions, we shall assume $c_{ab} > 0$). The centrifugal wall forms read

$$w^S_{(ab)}(\beta) \equiv w^S_{(ab)\mu} \beta^\mu \equiv \beta^b - \beta^a \quad (a < b). \tag{6.1}$$

We refer to these wall forms as the ‘symmetry walls’ for the following reason. When applying the general collision law (7.3) derived below to the case of the collision on the wall (6.1) one easily finds that its effect on the components of the velocity vector v^μ is simply to permute the components v^a and v^b , while leaving unchanged the other components $\mu \neq a, b$.

The hyperplanes $w^S_{(ab)}(\beta) = 0$ (i.e., the symmetry walls) are timelike since

$$G^{\mu\nu} w^S_{(ab)\mu} w^S_{(ab)\nu} = +2. \tag{6.2}$$

This ensures that the symmetry walls intersect the hyperboloid $G_{\mu\nu} \beta^\mu \beta^\nu = -1, \sum_a \beta^a \geq 0$. The symmetry billiard (in β -space) is defined to be the region of Minkowski space determined by the inequalities

$$w^S_{(ab)}(\beta) \geq 0, \tag{6.3}$$

with $\sum_a \beta^a \geq 0$ (i.e., by the region of β -space where the Θ functions are zero). Its projection on the hyperbolic space H_m is defined by the inequalities $w^S_{(ab)}(\gamma) \geq 0$.

The explicit expressions above of the symmetry wall forms also allow us to illustrate the notion of a ‘dominant wall’ defined in section 5.2.1. Indeed, the $d(d - 1)/2$ inequalities (6.3) already follow from the following *minimal* set of $d - 1$ inequalities:

$$\beta^2 - \beta^1 \geq 0, \beta^3 - \beta^2 \geq 0, \dots, \beta^d - \beta^{d-1} \geq 0. \tag{6.4}$$

More precisely, each linear form which must be positive in (6.3) can be written as a linear combination, with positive (in fact, integer) coefficients of the linear forms entering the subset (6.4). For instance, $\beta^3 - \beta^1 = (\beta^3 - \beta^2) + (\beta^2 - \beta^1)$, etc. In the last section, we will reinterpret this result by identifying the dominant linear forms entering (6.4) with the simple roots of $SL(n, \mathbb{R})$. Note that, in principle, the final set of dominant walls can only be decided when one starts from the complete list of all the dynamically relevant walls. In all the models we

examine the set of dominant symmetry walls (6.4) will, however, be part of the final minimal set of dominant walls defining the complete billiard table.

If only the symmetry walls were present in the Hamiltonian, one would easily deduce from the above the following picture of the dynamics: when the trajectory hits a wall $\beta^{a+1} = \beta^a$ from within the interior of the billiard $\beta^{a+1} - \beta^a > 0$ it undergoes a reflection which reorders v^{a+1} and v^a from the incident state $v^{a+1} < v^a$ (in which $\beta^{a+1} - \beta^a$ decreases towards zero) into an outgoing state $v^{a+1} > v^a$ (in which $\beta^{a+1} - \beta^a$ increases away from zero). Each collision reorders a pair of velocity components so that, after a finite number of collisions, they will get reordered in a stable configuration where $v^1 \leq v^2 \dots \leq v^d$ and $\beta^1 \leq \beta^2 \dots \leq \beta^d$. The motion would then continue freely, i.e., without collisions, after the last reordering reflection. The same conclusion was already reached in section 4 by a direct calculation of the dynamics of the Iwasawa scale factors of non-diagonal Kasner metrics (which is, indeed, a case where only symmetry walls are present).

If we still consider for a moment the simple case of homogeneous but non-diagonal metrics and recall that the metric can be diagonalized at all times by a time-independent coordinate transformation $x^i \rightarrow x'^i = L^i_j x^j$, it might appear that the symmetry walls, which are related to the off-diagonal components, are only a gauge artefact with no true physical content. This conclusion, however, would be incorrect. First, the transformation needed to diagonalize the metric may not be a globally well-defined coordinate transformation if the spatial sections have non-trivial topology, e.g., are tori, since it would conflict in general with periodicity conditions. Second, even if the spatial sections are homeomorphic to \mathbb{R}^d , the transformation $x^i \rightarrow x'^i = L^i_j x^j$, although a diffeomorphism, is not a proper gauge transformation in the sense that it is generated by a non-vanishing charge, and therefore two solutions that differ by such a transformation should be regarded as physically distinct (although related by a symmetry). Initial conditions, for which the metric is diagonal and hence the symmetry walls are absent, form a set of measure zero.

Let us finally mention the possibility of alternative treatments of the dynamics of off-diagonal metric components. We have just shown that the Iwasawa decomposition of the spatial metric leads to a projected description of the $GL(d, \mathbb{R})/SO(d)$ -geodesics as motions in the space of the scale factors with exponential ('Toda-like') potentials. An alternative description can be based on the decomposition $G = R^T A R$ of the spatial metric, where $R \in SO(d)$ and A is diagonal [70, 93]. One then obtains Calogero-like potentials $\propto \sinh^{-2}(\beta^a - \beta^b)$. In the BKL limit, these potentials can be replaced by sharp wall potentials but whether the system lies to the left or to the right of the wall $\beta^a - \beta^b = 0$ depends on the initial conditions in this alternative description.

6.2. Curvature (gravitational) walls

Next we analyse the gravitational potential, which requires a computation of curvature. To that end, one must explicitly express the spatial curvature in terms of the scale factors and the off-diagonal variables \mathcal{N}^a_i . Again, the calculation is most easily done in the Iwasawa frame (4.8), in which the metric assumes the form (4.6). We use the short-hand notation $A_a \equiv e^{-\beta^a}$ for the (Iwasawa) scale factors. Let $C^a_{bc}(x)$ be the structure functions of the Iwasawa basis $\{\theta^a\}$, namely,

$$d\theta^a = -\frac{1}{2}C^a_{bc}\theta^b \wedge \theta^c \quad (6.5)$$

where d is the spatial exterior differential. The structure functions obviously depend only on the off-diagonal components \mathcal{N}^a_i , but not on the scale factors. Using the Cartan formulae for the connection 1-form ω^a_b ,

$$d\theta^a + \sum_b \omega^a{}_b \wedge \theta^b = 0 \quad (6.6)$$

$$d\gamma_{ab} = \omega_{ab} + \omega_{ba} \quad (6.7)$$

where $\omega_{ab} \equiv \gamma_{ac}\omega^c{}_b$, and

$$\gamma_{ab} = \delta_{ab}A_a^2 \equiv e^{-2\beta^a} \delta_{ab} \quad (6.8)$$

is the metric in the frame $\{\theta^a\}$, one finds

$$\begin{aligned} \omega^c{}_d = \sum_b \frac{1}{2} \left(C^b{}_{cd} \frac{A_b^2}{A_c^2} + C^d{}_{cb} \frac{A_d^2}{A_c^2} - C^c{}_{db} \frac{A_c^2}{A_d^2} \right) \theta^b \\ + \sum_b \frac{1}{2A_c^2} [\delta_{cd}(A_c^2)_{,b} + \delta_{cb}(A_c^2)_{,d} - \delta_{db}(A_d^2)_{,c}] \theta^b. \end{aligned} \quad (6.9)$$

In the last bracket above, the commas denote the frame derivatives $\partial_a \equiv \mathcal{N}^i{}_a \partial_i$.

The Riemann tensor $R^c{}_{def}$, the Ricci tensor R_{de} and the scalar curvature R are obtained through

$$\Omega^a{}_b = d\omega^a{}_b + \sum_c \omega^a{}_c \wedge \omega^c{}_b \quad (6.10)$$

$$= \frac{1}{2} \sum_{e,f} R^a{}_{bef} \theta^e \wedge \theta^f \quad (6.11)$$

where $\Omega^a{}_b$ is the curvature 2-form and

$$R_{ab} = \sum_c R^c{}_{acb}, \quad R = \sum_a \frac{1}{A_a^2} R_{aa}. \quad (6.12)$$

Direct, but somewhat cumbersome, computations yield

$$R = -\frac{1}{4} \sum_{a,b,c} \frac{A_a^2}{A_b^2 A_c^2} (C^a{}_{bc})^2 + \sum_a \frac{1}{A_a^2} F_a(\partial^2 \beta, \partial \beta, \partial C, C) \quad (6.13)$$

where F_a is some complicated function of its arguments whose explicit form will not be needed here. The only property of F_a that will be of importance is that it is a polynomial of degree two in the derivatives $\partial \beta$ and of degree one in $\partial^2 \beta$. Thus, the exponential dependence on the β which determines the asymptotic behaviour in the BKL limit, occurs only through the A_a^2 -terms written explicitly in (6.13).

In (6.13) one obviously has $b \neq c$ because the structure functions $C^a{}_{bc}$ are antisymmetric in the pair $[bc]$. In addition to this restriction, we can assume, without loss of generality, that $a \neq b, c$ in the first sum on the right-hand side of (6.13). Indeed, the terms with either $a = b$ or $a = c$ can be absorbed into a redefinition of F_a . We can thus write the gravitational potential density (of weight 2) as

$$\mathcal{V}_G \equiv -gR = \frac{1}{4} \sum'_{a,b,c} e^{-2\alpha_{abc}(\beta)} (C^a{}_{bc})^2 - \sum_a e^{-2\mu_a(\beta)} F_a \quad (6.14)$$

where the prime on \sum indicates that the sum is to be performed only over unequal indices, i.e., $a \neq b, b \neq c, c \neq a$, and where the linear forms $\alpha_{abc}(\beta)$ and $\mu_a(\beta)$ are given by

$$\alpha_{abc}(\beta) = 2\beta^a + \sum_{e \neq a,b,c} \beta^e \quad (a \neq b, b \neq c, c \neq a) \quad (6.15)$$

and

$$\mu_a(\beta) = \sum_{c \neq a} \beta^c, \quad (6.16)$$

respectively. Note that α_{abc} is symmetric under the exchange of b with c , but the index a plays a special role.

Comparing the result (6.14) to the general form (5.25) we see that there are, *a priori*, two types of gravitational walls: the α -type and the μ -type. The α -type walls clearly come with positive pre-factors, proportional to the square of a structure function $C^a{}_{bc}$. The μ -type terms seem to pose a problem because they do not have a definite sign. It would therefore seem that, in the BKL limit, the gravitational potential would tend to

$$\lim_{\rho \rightarrow \infty} \mathcal{V}_G = \sum'_{a,b,c} \Theta[-2\alpha_{abc}(\beta)] + \sum_a (\pm \Theta[-2\mu_a(\beta)]). \quad (6.17)$$

However, the indefinite μ -type terms can generically be neglected in the BKL limit. This is most simply seen by noting that the inequalities $\alpha_{abc}(\beta) \geq 0$ imply $\mu_a(\beta) \geq 0$ because μ_a is a linear combination with positive coefficients of the α_{abc} . Indeed, we can write $\mu_c = (\alpha_{abc} + \alpha_{bca})/2$. Therefore the α -walls dominate the μ -ones.

In fact, one can establish a stronger result, namely $\mu_a(\beta) \geq 0$ within the entire future light cone of the β . For this purpose, we note first that each linear form $\mu_a(\beta)$ is *lightlike*, i.e., $G^{\mu\nu}(\mu_a)_\mu(\mu_a)_\nu = 0$. Therefore, each hyperplane $\mu_a(\beta) = 0$ is tangent to the light cone along some null generator. This means that the future light cone is entirely on one side of the hyperplane $\mu_a(\beta) = 0$ (i.e., either $\mu_a(\beta) > 0$ for all points inside the future light cone or $\mu_a(\beta) < 0$). Now, the point $\beta^1 = \beta^2 = \dots = \beta^d = 1$ is inside the future light cone and makes all the μ_a positive. Hence $\mu_a(\beta) > 0$ inside the future light cone for each a and $\Theta[-2\mu_a(\beta)] = 0$, and we really have

$$\lim_{\rho \rightarrow \infty} \mathcal{V}_G = \sum'_{a,b,c} \Theta[-2\alpha_{abc}(\beta)]. \quad (6.18)$$

Note, however, that the μ -type walls may make their existence felt in the exceptional case when β is close to the lightlike direction defined by μ . This is the case of ‘small oscillations’ considered by BKL in [6, 8], for which they verify (for $d = 3$) that the evolution is indeed controlled by the α_{abc} -terms even in that region.

From these considerations, we deduce the additional constraints

$$\alpha_{abc}(\beta) \geq 0 \quad (D > 3) \quad (6.19)$$

besides the symmetry inequalities (6.4). The hyperplanes $\alpha_{abc}(\beta) = 0$ are called the ‘curvature’ or ‘gravitational’ walls. Like the symmetry walls, they are timelike since

$$G^{\mu\nu}(\alpha_{abc})_\mu(\alpha_{abc})_\nu = +2. \quad (6.20)$$

The restriction $D > 3$ is due to the fact that in $D = 3$ spacetime dimensions, the gravitational walls $\alpha_{abc}(\beta) = 0$ are absent, simply because one cannot find three distinct spatial indices. In this case all gravitational walls are of subdominant type μ_a and thus, in the BKL limit,

$$\mathcal{V}_G \simeq \sum_a (\pm \Theta[-2\mu_a(\beta)]) \simeq 0 \quad (D = 3). \quad (6.21)$$

This is, of course, in agreement with expectations, because gravity in three spacetime dimensions has no propagating degrees of freedom (gravitational waves).

The fact that the gravitational potential becomes, in the BKL limit, a positive sum of sharp wall potentials, is remarkable for several reasons. First, the final form of the potential is

quite simple, even though the curvature is a rather complicated function of the metric and its derivatives. Secondly, the limiting expression of the potential is positive, even though there are subdominant terms in \mathcal{V}_G with indefinite sign. Thirdly, it is ultralocal in the scale factors, i.e., involves only the scale factors but not their derivatives. It is this fact that accounts for the decoupling of the various spatial points.

The coefficients of the dominant exponentials involve only the undifferentiated structure functions $C^a{}_{bc}$. Consequently, one can model the gravitational potential in leading order by *spatially homogeneous* cosmologies, which have constant structure functions, and by considering homogeneity groups that are ‘sufficiently non-Abelian’ so that none of the coefficients of the relevant exponentials vanish (Bianchi types VIII and IX for $d = 3$, other homogeneity groups for $d > 3$ —see [31]). By contrast, inspection of (6.13) and (6.14) reveals that the subleading terms not only lack manifest positivity, but also do depend on spatial inhomogeneities via the spatial gradient of the structure functions, and the first and second spatial derivatives of the scale factors. It is quite intriguing that the associated walls are *lightlike*, unlike the walls associated with the leading terms all of which are timelike. Terms involving lightlike walls will thus have to be taken into account in higher orders of the BKL expansion. A Kac–Moody theoretic interpretation of this fact was recently proposed in [28].

The computations in this subsection involve only the Cartan formulae. They remain valid if in (4.8) we replace the 1-forms dx^i by some anholonomic frame $f^i = f^i{}_j(x) dx^j$. This modifies the Iwasawa frame $\{\theta^a\}$, which has no intrinsic geometrical meaning. The structure functions $C^a{}_{bc}(x)$ of the new frame get extra contributions from the spatial derivatives acting on f^i . In fact, for $f^i = dx^i$ not all gravitational walls α_{abc} appear because from (4.7) we then have $C^d{}_{bc} = 0$ for the top component (since $\theta^d = dx^d$), and $C^{d-1}{}_{bc} = 0$ ($b, c \neq d$). Hence, the corresponding gravitational walls are absent. To obtain all the gravitational walls, one therefore needs an anholonomic frame f^i . However, the dominant gravitational wall α_{1d-1d} is always present, and this is the one relevant for the billiard, when gravitational walls are relevant at all.

6.3. p -form walls

While none of the wall forms considered so far involved the dilatons, the electric and magnetic ones do as we shall now show. To make the notation less cumbersome, we will omit the super-(or sub-)script (p) on the p -form fields in this subsection.

6.3.1. Electric walls. The electric potential density can be written as

$$\mathcal{V}_{(p)}^{\text{el}} = \frac{1}{2p!} \sum_{a_1, a_2, \dots, a_p} e^{-2e_{a_1 \dots a_p}(\beta)} (\mathcal{E}^{a_1 \dots a_p})^2 \quad (6.22)$$

where $\mathcal{E}^{a_1 \dots a_p}$ are the components of the electric field $\pi^{i_1 \dots i_p}$ in the basis $\{\theta^a\}$,

$$\mathcal{E}^{a_1 \dots a_p} \equiv \mathcal{N}^{a_1}{}_{j_1} \mathcal{N}^{a_2}{}_{j_2} \dots \mathcal{N}^{a_p}{}_{j_p} \pi^{j_1 \dots j_p} \quad (6.23)$$

(recall our summation conventions for spatial coordinate indices) and where $e_{a_1 \dots a_p}(\beta)$ are the electric wall forms,

$$e_{a_1 \dots a_p}(\beta) = \beta^{a_1} + \dots + \beta^{a_p} - \frac{\lambda_p}{2} \phi. \quad (6.24)$$

Here the indices a_j are all distinct because $\mathcal{E}^{a_1 \dots a_p}$ is completely antisymmetric. The variables $\mathcal{E}^{a_1 \dots a_p}$ do not depend on the β^μ . It is thus rather easy to take the BKL limit. The exponentials

in (6.22) are multiplied by positive factors which generically are different from zero. Thus, in the BKL limit, $\mathcal{V}_{(p)}^{\text{el}}$ becomes

$$\mathcal{V}_{(p)}^{\text{el}} \simeq \sum_{a_1 < a_2 < \dots < a_p} \Theta[-2e_{a_1 \dots a_p}(\beta)]. \quad (6.25)$$

The transformation from the variables $(\mathcal{N}^a_i, P^i_a, A_{j_1 \dots j_p}, \pi^{j_1 \dots j_p})$ to the variables $(\mathcal{N}^a_i, \mathcal{P}^i_a, \mathcal{A}_{a_1 \dots a_p}, \mathcal{E}^{a_1 \dots a_p})$ is a point canonical transformation whose explicit form is obtained from

$$\sum_a P^i_a \mathcal{N}^a_i + \sum_p \frac{1}{p!} \pi^{j_1 \dots j_p} \dot{A}_{j_1 \dots j_p} = \sum_a \mathcal{P}^i_a \mathcal{N}^a_i + \sum_p \sum_{a_1, \dots, a_p} \frac{1}{p!} \mathcal{E}^{a_1 \dots a_p} \dot{\mathcal{A}}_{a_1 \dots a_p}. \quad (6.26)$$

The new momenta \mathcal{P}^i_a conjugate to \mathcal{N}^a_i differ from the old ones P^i_a by terms involving \mathcal{E} , \mathcal{N} and \mathcal{A} since the components $\mathcal{A}_{a_1 \dots a_p}$ of the p -forms in the basis $\{\theta^a\}$ depend on the \mathcal{N} ,

$$\mathcal{A}_{a_1 \dots a_p} = \mathcal{N}^{j_1}_{a_1} \dots \mathcal{N}^{j_p}_{a_p} A_{j_1 \dots j_p}.$$

However, it is easy to see that these extra terms do not affect the symmetry walls in the BKL limit.

6.3.2. Magnetic walls. The magnetic potential is dealt with similarly. Expressing it in the $\{\theta^a\}$ -frame, one obtains

$$\mathcal{V}_{(p)}^{\text{magn}} = \frac{1}{2(p+1)!} \sum_{a_1, a_2, \dots, a_{p+1}} e^{-2m_{a_1 \dots a_{p+1}}(\beta)} (\mathcal{F}_{a_1 \dots a_{p+1}})^2 \quad (6.27)$$

where $\mathcal{F}_{a_1 \dots a_{p+1}}$ are the components of the magnetic field $F_{m_1 \dots m_{p+1}}$ in the basis $\{\theta^a\}$,

$$\mathcal{F}_{a_1 \dots a_{p+1}} = \mathcal{N}^{j_1}_{a_1} \dots \mathcal{N}^{j_{p+1}}_{a_{p+1}} F_{j_1 \dots j_{p+1}}. \quad (6.28)$$

The $m_{a_1 \dots a_{p+1}}(\beta)$ are the magnetic linear forms

$$m_{a_1 \dots a_{p+1}}(\beta) = \sum_{b \notin \{a_1, a_2, \dots, a_{p+1}\}} \beta^b + \frac{\lambda_p}{2} \phi \quad (6.29)$$

where again all a_j are distinct. One sometimes rewrites $m_{a_1 \dots a_{p+1}}(\beta)$ as $\tilde{m}_{a_{p+2} \dots a_d}$, where $\{a_{p+2}, a_{p+3}, \dots, a_d\}$ is the set complementary to $\{a_1, a_2, \dots, a_{p+1}\}$; e.g.,

$$\tilde{m}_{12 \dots d-p-1} = \beta^1 + \dots + \beta^{d-p-1} + \frac{\lambda_p}{2} \phi = m_{d-p \dots d}. \quad (6.30)$$

Of course, the components of the exterior derivative \mathcal{F} of \mathcal{A} in the non-holonomic frame $\{\theta^a\}$ involve the structure coefficients, i.e., $\mathcal{F}_{a_1 \dots a_{p+1}} = \partial_{[a_1} \mathcal{A}_{a_2 \dots a_{p+1}]} + C\mathcal{A}$ -terms where $\partial_a \equiv \mathcal{N}^i_a \partial_i$ is the frame derivative.

Again, the BKL limit is quite simple and yields (assuming generic magnetic fields)

$$\mathcal{V}_{(p)}^{\text{magn}} \simeq \sum_{a_1 < \dots < a_{d-p-1}} \Theta[-2b_{a_1 \dots a_{d-p-1}}(\beta)]. \quad (6.31)$$

As the off-diagonal variables, the electric and magnetic fields freeze in the BKL limit since the Hamiltonian no longer depends on the p -form variables. These drop out because one can rescale the coefficient of any Θ -function to be 1 (when it is not zero), thereby absorbing the dependence on the p -form variables.

The scale factors are therefore constrained by the further ‘billiard’ conditions

$$e_{a_1 \dots a_p}(\beta) \geq 0, \quad \tilde{m}_{a_1 \dots a_{d-p-1}}(\beta) \geq 0. \quad (6.32)$$

The hyperplanes $e_{a_1 \dots a_p}(\beta) = 0$ and $\tilde{m}_{a_1 \dots a_{d-p-1}}(\beta) = 0$ are called ‘electric’ and ‘magnetic’ walls, respectively. Both walls are timelike because their gradients are spacelike, with squared norm

$$\frac{p(d-p-1)}{d-1} + \left(\frac{\lambda_p}{2}\right)^2 > 0. \tag{6.33}$$

(For $D = 11$ supergravity, we have $d = 10$, $p = 3$ and $\lambda_p = 0$ and thus the norm is equal to +2.) This equality explicitly shows the invariance of the norms of the p -form walls under electric–magnetic duality.

6.4. Walls due to dilatons or to a cosmological constant

The fact that the leading walls originating from the centrifugal, gravitational and p -form potentials, are all timelike, is an important ingredient of the overall BKL picture. As we saw above, however, there exist subleading contributions in the gravitational potential whose associated wall forms are lightlike. Such is the case also for the dilaton contribution

$$\mathcal{V}_\phi = g g^{ij} \partial_i \phi \partial_j \phi, \tag{6.34}$$

which has the same form as the subleading gravitational walls, since the exponentials that control its asymptotic behaviour are easily seen to be $e^{-2\mu_a(\beta)}$. Consequently, at least to leading order, we can neglect \mathcal{V}_ϕ in the BKL limit.

The only example of a *spacelike* wall that we know is the cosmological constant term (in its weight -2 form)

$$\mathcal{V}_\Lambda = \Lambda g = \Lambda \exp \left[-2 \sum_a \beta^a \right]. \tag{6.35}$$

When Λ is positive (de Sitter sign) the spacelike wall (6.35) is repulsive. Depending on the initial conditions—which set the scale—this wall either prevents the system from reaching the BKL small volume regime or does not prevent the collapse. In the first case, the spacelike wall acts as a ‘barrier’ to the motion of the billiard ball, and the reflection against it forces the billiard ball to run ‘backwards in β -time’ in the direction of *increasing* spatial volume (this is analogous to the bounce in the (global) de Sitter solution, which is a hyperboloid—a sphere that first contracts and then expands). In the second case, the billiard ball is already ‘beyond the barrier’, and the presence of the spacelike wall has only a subdominant effect on its motion. In the BKL picture it quickly becomes negligible: the cosmological potential Λg tends to zero as g goes to zero. When Λ is negative (anti-de Sitter sign) the spacelike wall (3.37) is attractive and tends to favour collapse. It, however, quickly becomes negligible in the BKL limit.

6.5. Chern–Simons and Yang–Mills couplings

The addition of Chern–Simons terms, Yang–Mills or Chapline–Manton couplings does not bring in new (asymptotically relevant) walls. The only change in the asymptotic dynamics is a modification of the constraints.

For Yang–Mills couplings, the contribution to the energy density from the Yang–Mills field takes the same form as for a collection of Abelian 1-forms, with the replacement of the momenta π^i by the Yang–Mills momenta π_a^i (where $a = 1, \dots, N$ and N is the dimension of the internal Lie algebra) and of the magnetic fields by the corresponding non-Abelian field strengths. As their Abelian counterparts, these do not involve the scale factors β^μ . Because of this key property, the same analysis goes through. Each electric and magnetic 1-form wall

is simply repeated a number of times equal to the dimension of the Lie algebra. Gauss law is, however, modified and reads

$$\mathcal{D}_i \pi_a^i \equiv \nabla_i \pi_a^i + f^b{}_{ac} \pi_b^i A_i^c = 0. \quad (6.36)$$

Here, ∇_i is the standard metric covariant derivative. Similarly, the momentum constraints are modified and involve the non-Abelian field strengths.

The discussion of Chapline–Manton couplings or Chern–Simons terms proceeds in the same way. The energy density of the p -forms has the same dependence on the scale factors as in the absence of couplings, i.e., provides the same exponentials. The only difference is that the wall coefficients are different functions of the p -form canonical variables; but this difference is again washed out in the sharp wall limit, where the coefficients can be replaced by 1 (provided they are different from zero). The momentum and Gauss constraints are genuinely different and impose different conditions on the initial data.

Nevertheless, the Chern–Simons terms may play a more significant role in peculiar contexts when only specialized field configurations are considered. This occurs for instance in [58], where it is shown that the $D = 11$ supergravity Chern–Simons term for spatially homogeneous metrics and magnetic fields may constrain some electromagnetic walls to disappear ‘accidentally’. This changes the finite volume billiard to that of infinite volume.

7. Cosmological billiards

Let us summarize our findings. The dynamics in the vicinity of a spacelike singularity is governed by the scale factors, while the other variables (off-diagonal metric components, p -form fields) tend to become mere ‘spectators’ which get asymptotically frozen. This simple result is most easily derived in terms of the hyperbolic polar coordinates (ρ, γ) , and in the gauge (2.5). In this picture, the essential dynamics is carried by the angular variables γ which move on a fixed billiard table, with cushions defined by the dominant walls $w_{A'}(\gamma)$. However, it is often geometrically more illuminating to ‘unproject’ this billiard motion in the full Minkowski space of the extended scale factors β^μ . In that picture, the asymptotic evolution of the scale factors at each spatial point reduces to a zigzag of null straight lines w.r.t. the metric $G_{\mu\nu} d\beta^\mu d\beta^\nu$. The straight segments of this motion are interrupted by collisions against the sharp walls

$$w_A(\beta) \equiv w_{A\mu} \beta^\mu = 0 \quad (7.1)$$

defined by the symmetry, gravitational and p -form potentials, respectively. As we showed all these walls are *timelike*, i.e., they have spacelike gradients:

$$G^{\mu\nu} w_{A\mu} w_{Av} > 0 \quad (7.2)$$

Indeed, the gradients of the symmetry and gravitational wall forms have squared norm equal to +2, independent of the dimension d . By contrast, the norms of the electric and magnetic gradients, which are likewise positive, depend on the model. As we saw, there also exist subdominant walls, which can be neglected as they are located ‘behind’ the dominant walls.

In the β -space picture, the free motion before a collision is described by a null straight line of the type (3.21), with the constraint (3.22). The effect of a collision on a particular wall $w_A(\beta)$ is easily obtained by solving (5.31), or (5.32), with only one term in the sum. This dynamics is exactly integrable: it suffices to decompose the motion of the β -particle into two (linear) components: (i) the component parallel to the (timelike) wall hyperplane and (ii) the orthogonal component. One easily finds that the parallel motion is left unperturbed by the presence of the wall, while the orthogonal motion suffers a (one-dimensional) reflection, with

a change of the sign of the outgoing orthogonal velocity with respect to the ingoing one. The net effect of the collision on a certain wall $w(\beta)$ then is to change the ingoing velocity vector v^μ entering the ingoing free motion (3.21) into an outgoing velocity vector v'^μ given (in any linear frame) by the usual formula for a geometric reflection in the hyperplane $w(\beta) = 0$:

$$v'^\mu = v^\mu - 2 \frac{(w \cdot v)w^\mu}{(w \cdot w)}. \quad (7.3)$$

Here, all scalar products, and index raisings, are done with the β -space metric $G_{\mu\nu}$. Note that the collision law (7.3) leaves invariant the (Minkowski) length of the vector v^μ . Because the dominant walls are timelike, the geometric reflections that the velocities undergo during a collision, are elements of the orthochronous Lorentz group. Each reflection preserves the norm and the time orientation; hence, the velocity vector remains null and future oriented.

From this perspective, we can also better understand the relevance of walls which are *not* timelike. *Lightlike* walls (like some of the subleading gravitational walls) can never cause reflections because in order to hit them the billiard ball would have to move at superluminal speeds in violation of the Hamiltonian constraint. The effect of *spacelike* walls (like the cosmological constant wall) is again different: they are either irrelevant (if they are ‘behind the motion’), or otherwise they reverse the time orientation inducing a motion towards increasing spatial volume (‘bounce’).

The *hyperbolic billiard* is obtained from the β -space picture by a radial projection onto the unit hyperboloid of the piecewise straight motion in the polywedge defined by the walls. The straight motion thereby becomes a geodesic motion on hyperbolic space. The ‘cushions’ of the hyperbolic billiard table are the intersections of the hyperplanes (7.1) with the unit hyperboloid, such that the billiard motion is constrained to be in the region defined as the intersection of the half-spaces $w_A(\beta) \geq 0$ with the unit hyperboloid. As we already emphasized, not all walls are relevant since some of the inequalities $w_A(\beta) \geq 0$ are implied by others [26]. Only the dominant wall forms, in terms of which all the other wall forms can be expressed as linear combinations with non-negative coefficients, are relevant for determining the billiard. Usually, these are the minimal symmetry walls and some of the p -form walls. The billiard region, as a subset of hyperbolic space, is in general *non-compact* because the cushions meet at infinity (i.e., at a cusp); in terms of the original scale-factor variables β , this means that the corresponding hyperplanes intersect on the light cone. It is important that, even when the billiard is non-compact, the hyperbolic region can have finite volume.

Given the action (2.1) with definite spacetime dimension, menu of fields and dilaton couplings, one can determine the relevant wall forms and compute the billiards. For generic initial conditions, we have the following results, as to which of the models (2.1) exhibit oscillatory behaviour (finite volume billiard) or Kasner-like behaviour (infinite volume billiard):

- Pure gravity billiards have finite volume for spacetime dimension $D \leq 10$ and infinite volume for spacetime dimension $D \geq 11$ [33]. This can be understood in terms of the underlying Kac–Moody algebra [27]: as shown there, the system is chaotic precisely if the underlying indefinite Kac–Moody algebra is hyperbolic.
- The billiard of gravity coupled to a dilaton always has infinite volume, hence exhibits Kasner-like behaviour [2, 5, 29].
- If gravity is coupled to p -forms (with $p \neq 0$ and $p < D - 2$) *without a dilaton* the corresponding billiard has a finite volume [25]. The most prominent example in this class is $D = 11$ supergravity, whereas vacuum gravity in 11 dimensions is Kasner-like. The 3-form is crucial for closing the billiard. Similarly, the Einstein–Maxwell system in four (in fact any number of) dimensions has a finite-volume billiard (see [60, 75, 100] for

a discussion of four-dimensional homogeneous models with Maxwell fields exhibiting oscillatory behaviour).

- The volume of the mixed Einstein-dilaton- p -form system depends on the dilaton couplings. For a given spacetime dimension D and a given menu of p -forms, there exists a subcritical domain \mathcal{D} in the space of the dilaton couplings, i.e., an open neighbourhood of the origin $\lambda_p = 0$ such that (i) when the dilaton couplings λ_p belong to \mathcal{D} the general behaviour is Kasner-like, but (ii) when the λ_p do not belong to \mathcal{D} the behaviour is oscillatory [24, 29]. For all the superstring models, the dilaton couplings do not belong to the subcritical domain and the billiard has finite volume. Note, however, that the superstring dilaton couplings are precisely ‘critical’, i.e., on the borderline between the subcritical and the overcritical domain, for the heterotic and type I theories.

As a note of caution let us point out that some indicators of chaos must be used with care in general relativity, because of reparametrization invariance, and in particular redefinitions of the time coordinate; see [18, 54] for a discussion of the original Bianchi IX model.

We next discuss the link between the various time coordinates used in the analysis. When working in the gauge (2.5), the basic time coordinate is T . Let us see how the other time coordinates depend on T . First, we note that the dynamical variable λ is asymptotically a linear function of T . Indeed, its conjugate momentum $\pi_\lambda < 0$ is asymptotically constant, so that the integration of $d\lambda/dT = -\frac{1}{2}\pi_\lambda$ yields

$$\lambda = -\frac{1}{2}\pi_\lambda T + \text{const.} \quad (7.4)$$

Hence ρ is an exponential function of T :

$$\rho = e^\lambda \propto e^{-\frac{1}{2}\pi_\lambda T}. \quad (7.5)$$

From this behaviour, we infer the time dependence of the intermediate time coordinate τ by integrating its defining relation $d\tau = \rho^2 dT = -(2/\pi_\lambda)\rho^2 d\lambda = -(2/\pi_\lambda)\rho d\rho$. This yields

$$\tau = -(1/\pi_\lambda)\rho^2 + \text{const.} \quad (7.6)$$

At this stage, the asymptotic links between T , τ , λ and ρ are similar to the results derived above for a Kasner solution and do not depend on whether the system is chaotic or non-chaotic.

The situation is more subtle for the proper time t in the chaotic case (in the non-chaotic case where one settles in a Kasner regime after a finite number of collisions, nothing is changed, of course). The proper time is obtained by integrating

$$dt = \sqrt{g} d\tau = \sqrt{g}\rho^2 dT = -(2/\pi_\lambda)\sqrt{g}\rho d\rho. \quad (7.7)$$

The main term in the integrand is $\sqrt{g} = e^{-\rho\sigma}$, where $\sigma \equiv \sum_i \gamma^i$. The quantity σ oscillates chaotically and is difficult to control. As argued in [69], however, one nevertheless obtains (in the four-dimensional case) the usual Kasner relation $\sqrt{g} \sim t$ (and thus $\tau \sim \ln t$ and $T \sim \ln |\ln t|$) up to subdominant corrections. We refer the interested reader to [69] for the details.

Concerning the frequency of collisions, note that the billiard picture makes it clear that the typical time interval between two collisions is constant as a function of T . In other words, the number of collisions goes like $\lambda \approx \frac{1}{2} \ln \tau \sim \ln |\ln t|$.

The hyperbolic billiard description of the $(3+1)$ -dimensional homogeneous Bianchi IX system was first worked out by Chitre [17] and Misner [85]. It was subsequently generalized to inhomogeneous metrics in [56, 71]. The extension to higher dimensions with perfect fluid sources was considered in [73], without symmetry walls. Exterior p -form sources were investigated in [57, 58] for special classes of metric and p -form configurations. As far as we know, however, the uniform approach used in the present paper based on a systematic use of the Iwasawa decomposition of the spatial metric is new.

8. Kac–Moody theoretic formulation

Although the billiard description holds for all systems governed by the action (2.1), the billiard in general has no notable regularity property. In particular, the dihedral angles between the faces, which can depend on the (continuous) dilaton couplings, need not be integer submultiples of π . In some instances, however, the billiard can be identified with the fundamental Weyl chamber of a symmetrizable Kac–Moody (or KM) algebra of indefinite type¹³, with Lorentzian signature metric [23, 26, 27]. Such billiards are called ‘Kac–Moody billiards’. More specifically, in [26], superstring models were considered and the rank 10 KM algebras E_{10} and BE_{10} were shown to emerge, in line with earlier conjectures made in [62, 64].¹⁴ This result was further extended to pure gravity in any number of spacetime dimensions, for which the relevant KM algebra is AE_d , and it was understood that chaos (finite volume of the billiard) is equivalent to hyperbolicity of the underlying KM algebra [27]. For pure gravity in $D = 4$ the relevant algebra is the hyperbolic algebra AE_3 first investigated in [38]. Further examples of emergence of Lorentzian KM algebras, based on the models of [13, 21], are given in [23].

The main feature of the gravitational billiards that can be associated with KM algebras is that there exists a group theoretical interpretation of the billiard motion: the asymptotic BKL dynamics is equivalent (in a sense to be made precise below), at each spatial point, to the asymptotic dynamics of a one-dimensional nonlinear σ -model based on a certain infinite-dimensional coset space G/K , where the KM group G and its maximal compact subgroup K depend on the specific model. As we have seen, the walls that determine the billiards are the dominant walls. For KM billiards, they correspond to the simple roots of the KM algebra. Some of the subdominant walls also have an algebraic interpretation in terms of higher-height positive roots. This enables one to go beyond the BKL limit and to see the beginnings of a possible identification of the dynamics of the scale factors *and* of all the remaining variables with that of a nonlinear σ -model defined on the cosets of the KM group divided by its maximal compact subgroup [28].

The KM theoretic reformulation will not only enable us to give a unified group theoretical derivation of the different types of walls discussed in the preceding section, but also shows that *the β -space of logarithmic scale factors, in which the billiard motion takes place, can be identified with the Cartan subalgebra of the underlying indefinite KM algebra*. The various types of walls can thus be understood directly as arising from the large field limit of the corresponding σ -models. It is the presence of gravity, which comes with a metric in scale-factor space of Lorentzian signature, which forces us to consider *infinite-dimensional* groups if we want to recover all the walls found in our previous analysis, and this is the main reason why we need the theory of KM algebras. For finite-dimensional Lie algebras, we obtain only a subset of the walls: one of the cushions of the associated billiard is missing, and one always ends up with a monotonic Kasner-type behaviour in the limit $t \rightarrow 0^+$ (for instance, the non-diagonal Kasner solution discussed in section 4 has only symmetry walls corresponding to the finite-dimensional coset space $GL(d, \mathbb{R})/SO(d)$). The absence of chaotic oscillations for models based on finite-dimensional Lie groups is consistent with the classical integrability of these models. While they remain formally integrable for infinite-dimensional KM groups, one can understand the chaotic behaviour as resulting from the projection of a motion in an infinite-dimensional space onto a finite dimensional subspace.

¹³ Throughout this section, we will use the abbreviations ‘KM’ for ‘Kac–Moody’ and ‘CSA’ for ‘Cartan subalgebra’.

¹⁴ Note that the Weyl groups of the E -family have been discussed in a similar vein in the context of U -duality [3, 79, 89].

Before proceeding we would like to emphasize that the equivalence between the models discussed in the foregoing sections and the KM σ -models to be presented in this section has so far only been established, in the case of general models, for their asymptotic dynamics. A proposal relating the infinitely many off-diagonal degrees of freedom arising in the KM σ -model to spatial gradients of the metric and other fields, as well as possibly other degrees of freedom has recently been made in [28]. The first steps of this proposal have been explicitly checked for the relation between $D = 11$ supergravity and the E_{10} σ -model. The relevance of nonlinear σ -models for uncovering the symmetries of M-theory has also been discussed from a different, spacetime-covariant point of view in [96, 97, 101], but there it is E_{11} rather than E_{10} that has been proposed as a fundamental symmetry.

8.1. Some basic facts about KM algebras

We first summarize some basic results from the theory of KM algebras, referring the reader to [66, 86] for comprehensive treatments. As explained there, every KM algebra $\mathfrak{g} \equiv \mathfrak{g}(A)$ is defined by means of an integer-valued Cartan matrix A and a set of generators and relations [66, 86]. We shall assume that the Cartan matrix is symmetrizable since this is the case encountered for cosmological billiards. The Cartan matrix can then be written as $(i, j = 1, \dots, r)$, with r denoting the rank of $\mathfrak{g}(A)$

$$A_{ij} = \frac{2\langle \alpha_i | \alpha_j \rangle}{\langle \alpha_i | \alpha_i \rangle} \quad (8.1)$$

where $\{\alpha_i\}$ is a set of r simple roots, and where the angular brackets denote the invariant symmetric bilinear form of $\mathfrak{g}(A)$ [66]. Here the bilinear form acts on the roots, which are linear forms on the Cartan subalgebra (CSA) $\mathfrak{h} \subset \mathfrak{g}(A)$. The generators, which are also referred to as Chevalley–Serre generators, consist of triples $\{h_i, e_i, f_i\}$ with $i = 1, \dots, r$, and for each i form an $sl(2, \mathbb{R})$ subalgebra. The CSA \mathfrak{h} is then spanned by the elements h_i , so that

$$[h_i, h_j] = 0. \quad (8.2)$$

Furthermore,

$$[e_i, f_j] = \delta_{ij} h_j \quad (8.3)$$

and

$$[h_i, e_j] = A_{ij} e_j, \quad [h_i, f_j] = -A_{ij} f_j \quad (8.4)$$

so that the value of the linear form α_j , corresponding to the raising operator e_j , on the element h_i of the preferred basis $\{h_i\}$ of \mathfrak{h} is $\alpha_j(h_i) = A_{ij}$. More abstractly, and independent of the choice of any basis in the CSA, the roots appear as eigenvalues of the adjoint action of any element h of the CSA on the raising (e_i) or lowering (f_i) generators: $[h, e_i] = +\alpha_i(h)e_i$, $[h, f_i] = -\alpha_i(h)f_i$. Last but not least, we have the so-called Serre relations

$$\text{ad}(e_i)^{1-A_{ij}}(e_j) = 0, \quad \text{ad}(f_i)^{1-A_{ij}}(f_j) = 0 \quad (8.5)$$

Every KM algebra possesses the triangular decomposition

$$\mathfrak{g}(A) = \mathfrak{n}^- \oplus \mathfrak{h} \oplus \mathfrak{n}^+ \quad (8.6)$$

where \mathfrak{n}^+ and \mathfrak{n}^- , respectively, are spanned by the multiple commutators of the e_i and f_i which do not vanish on account of the Serre relations and the Jacobi identity. To be completely precise, \mathfrak{n}^+ is the quotient of the free Lie algebra generated by the e_i by the ideal generated by the Serre relations (*idem* for \mathfrak{n}^- and f_i). In more mundane terms, when the algebra is realized, in a suitable basis, by infinite-dimensional matrices, \mathfrak{n}^+ and \mathfrak{n}^- simply consist of the ‘nilpotent’ matrices with non-zero entries only above or below the diagonal. Exponentiating

them formally, one obtains infinite-dimensional matrices again with non-zero entries above or below the diagonal, and 1's on the diagonal—as exemplified in the finite-dimensional case by the matrices \mathcal{N} in section 4.

One of the main results of the theory is that, for positive definite A , one only recovers from these relations Cartan's list of finite-dimensional Lie algebras (see, e.g., [53] for a clear introduction). For non-positive definite A , on the other hand, the associated KM algebras are infinite dimensional. If A has only one zero eigenvalue (with all other eigenvalues strictly positive) one obtains the so-called affine algebras, whose structure and properties are rather well understood [47, 66]. For indefinite A (i.e., at least one negative eigenvalue of A), on the other hand, very little is known, and it remains an outstanding problem to find a manageable representation for them [66, 86] (see also [43] for a physicist's introduction). In particular, there is no single example of an indefinite KM algebra for which the root multiplicities, i.e., the number of Lie-algebra elements associated with a given root, are known in closed form. The scarcity of results is even more acute for the 'KM groups' obtained by formal exponentiation of the associated Lie algebras. In spite of these caveats, we will proceed formally, making sure that our definitions reduce to the standard formulae in the truncation to finite-dimensional Lie groups, and more generally remain well defined when we restrict the number of degrees of freedom to any finite subset.

In the remainder, we will thus assume A to be *Lorentzian*, i.e., non-degenerate and indefinite, with one negative eigenvalue. We shall see that this choice is physically motivated by the fact that the negative eigenvalue can be associated with the conformal factor which, as we saw, is the one degree of freedom making the reduced Einstein action unbounded from below. As a special, and important case, this class of Lorentzian KM algebras includes *hyperbolic* KM algebras whose Cartan matrices are such that the deletion of any node from the Dynkin diagram leaves either a finite or an affine subalgebra, or a disjoint union of them.

The 'maximal compact' subalgebra \mathfrak{k} is defined as the invariant subalgebra of $\mathfrak{g}(A)$ under the standard Chevalley involution, i.e.,

$$\theta(x) = x \quad \text{for all } x \in \mathfrak{k} \tag{8.7}$$

with

$$\theta(h_i) = -h_i, \quad \theta(e_i) = -f_i, \quad \theta(f_i) = -e_i. \tag{8.8}$$

More explicitly, it is the subalgebra generated by multiple commutators of $(e_i - f_i)$. For finite-dimensional $\mathfrak{g}(A)$, the inner product induced on the maximal compact subalgebra \mathfrak{k} is negative definite, and the orthogonal complement to \mathfrak{k} has a positive definite inner product. This is not so, however, for indefinite A (see the footnote on page 438 of [65]).

It will be convenient in the following to introduce the operation of *transposition* acting on any Lie-algebra element E as

$$E^T := -\theta(E). \tag{8.9}$$

In this notation, the relations above become $h_i^T = h_i$, $e_i^T = f_i$, $f_i^T = e_i$, and the subalgebra \mathfrak{k} is generated by the 'antisymmetric' elements satisfying $E^T = -E$. After exponentiation, the elements of the corresponding maximally compact subgroup K formally appear as 'orthogonal matrices' obeying $k^T = k^{-1}$.

Sometimes it is convenient to use the so-called Cartan–Weyl basis for $\mathfrak{g}(A)$. Using Greek indices μ, ν, \dots to label the root components corresponding to an arbitrary basis H_μ in the CSA, with the usual summation convention and a Lorentzian metric $G_{\mu\nu}$ for an indefinite \mathfrak{g} , we have $h_i := \alpha_i^\mu H_\mu$, where α_i^μ are the 'contravariant components', $G_{\mu\nu}\alpha_i^\nu \equiv \alpha_{i\mu}$, of

the simple roots α_i ($i = 1, \dots, r$), which are linear forms on the CSA, with usual ‘covariant components’ defined as $\alpha_{i\mu} \equiv \alpha_i(H_\mu)$.

To an arbitrary root α corresponds to a set of Lie-algebra generators $E_{\alpha,s}$, where $s = 1, \dots, \text{mult}(\alpha)$ labels (in general) the multiple Lie-algebra elements associated with α . The root multiplicity $\text{mult}(\alpha)$ is always 1 for finite-dimensional Lie algebras, and is also 1 for the positive norm roots $\alpha^2 \equiv \langle \alpha | \alpha \rangle > 0$ (i.e., the ‘real roots’) of general KM algebras (including, of course, the simple roots), but generically exhibits exponential growth as a function of $-\alpha^2$ for indefinite A . In this notation, the remaining Chevalley–Serre generators are given by $e_i := E_{\alpha_i}$ and $f_i := E_{-\alpha_i}$. Then we have

$$[H_\mu, E_{\alpha,s}] = \alpha_\mu E_{\alpha,s} \quad (8.10)$$

and

$$[E_{\alpha,s}, E_{\alpha',t}] = \sum_u c_{\alpha\alpha'}^{s,t,u} E_{\alpha+\alpha',u}. \quad (8.11)$$

The elements of the Cartan–Weyl basis are normalized such that

$$\langle H_\mu | H_\nu \rangle = G_{\mu\nu}, \quad \langle E_{\alpha,s} | E_{\beta,t} \rangle = \delta_{st} \delta_{\alpha+\beta,0} \quad (8.12)$$

where we have assumed that the basis satisfies $E_\alpha^T = E_{-\alpha}$. Let us finally recall that the Weyl group of a KM algebra is the discrete group generated by reflections in the hyperplanes orthogonal to the simple roots.

8.2. Decomposition of AE_3 into $SL(3)$ representations

As a special example, let us consider the hyperbolic KM algebra AE_3 , with Cartan matrix¹⁵

$$A_{ij} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -2 \\ 0 & -2 & 2 \end{pmatrix} \quad (8.13)$$

and work out the first few terms of its decomposition into $SL(3, \mathbb{R})$ representations. This decomposition refers to the *adjoint action* of the $sl(3, \mathbb{R})$ subalgebra defined below on the complete KM algebra.

From the mathematical point of view, the algebra (8.13) was first studied in [38] as it is the simplest hyperbolic KM algebra containing a non-trivial affine subalgebra (see also [68] and references therein for more recent work). From the physical perspective, this algebra is of special interest for pure gravity in four dimensions both because its regular subalgebras and its Chevalley–Serre generators can be physically identified with known symmetry groups arising in dimensional reductions of general relativity [87], and because its Weyl chamber is related to the original BKL billiard [27] (see footnote 16). First of all, the $SL(2, \mathbb{R})$ subgroup corresponding to the third diagonal entry of A_{ij} can be identified with the Ehlers group [37] which acts on solutions of Einstein’s equations with one Killing vector. The affine subgroup corresponding to the submatrix

$$\begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \quad (8.14)$$

is the Geroch group acting on solutions of Einstein’s equations with two commuting Killing vectors [44] (axisymmetric stationary or colliding plane wave solutions, respectively); see, e.g., [12, 62, 63, 88] and references therein. Both the Ehlers group and the Geroch group act in part by *nonlocal* transformations on the metric (or vierbein) components.

¹⁵ This algebra is also called \mathcal{F} [38], H_3 [66], $HA_1^{(1)}$ [68] or A_1^{\wedge} [27].

A crucial role, in our analysis, is played by the $SL(3, \mathbb{R})$ subgroup generated by (e_1, f_1, h_1) and (e_2, f_2, h_2) , corresponding to the submatrix

$$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}. \tag{8.15}$$

This group can be realized as acting on the spatial components of the metric (or vierbein), as an extension of the so-called Matzner–Misner $SL(2, \mathbb{R})$ group for solutions with two commuting Killing vectors [82].

The billiard picture for pure gravity in four dimensions can be readily understood in terms of the Weyl group of AE_3 [27], whose even part is just the modular group $PGL(2, \mathbb{Z})$ [38]. For $SL(3, \mathbb{R})$, which has two simple roots, the Weyl group is the permutation group on three objects. The two hyperplanes orthogonal to the simple roots of $SL(3, \mathbb{R})$ can be identified with the symmetry walls encountered in section 6.1. The third simple root extending (8.15) to the full rank 3 algebra (8.13), which ‘closes off’ the billiard, can then be identified with curvature wall orthogonal to α_{123} [27]. Readers may indeed check that the scalar products between the two symmetry wall forms and the curvature wall form reproduce the above Cartan matrix¹⁶.

Let us now expand the nilpotent subalgebra \mathfrak{n}^+ in terms of representations of the $A_2 \equiv \mathfrak{sl}(3, \mathbb{R})$ subalgebra of AE_3 exhibited in (8.15). For this purpose, given any root α , we define its $\mathfrak{sl}(3, \mathbb{R})$ level ℓ to be the number of times the root α_3 appears in it, to wit

$$\alpha = m\alpha_1 + n\alpha_2 + \ell\alpha_3. \tag{8.16}$$

Note that this notion of level is different from the affine level ($\equiv m$) which counts the number of appearances of the overextended root α_1 [38]. The hyperbolic algebra is thus decomposed into an infinite tower of irreducible representations of its $\mathfrak{sl}(3, \mathbb{R})$ subalgebra. Such a decomposition is simpler than that in terms of the affine subalgebra, whose representation theory is far more complicated (and, unlike that of $\mathfrak{sl}(3, \mathbb{R})$, only incompletely understood).

After inclusion of the third Cartan generator h_3 , the level $\ell = 0$ sector is only a $\mathfrak{gl}(3, \mathbb{R})$ subalgebra with generators K^i_j (where $i, j = 1, 2, 3$) and commutation relations

$$[K^i_j, K^k_l] = \delta_j^k K^i_l - \delta_l^i K^k_j. \tag{8.17}$$

These Lie-algebra elements will be seen to generate the $GL(3, \mathbb{R})$ group acting on the spatial components of the vierbein. The restriction of the AE_3 -invariant bilinear form to the level-0 sector is

$$\langle K^i_j | K^k_l \rangle = \delta_j^k \delta_l^i - \delta_j^i \delta_l^k. \tag{8.18}$$

The identification with the Chevalley–Serre generators is

$$\begin{aligned} e_1 &= K^1_2, & f_1 &= K^2_1, & h_1 &= K^1_1 - K^2_2 \\ e_2 &= K^2_3, & f_2 &= K^3_2, & h_2 &= K^2_2 - K^3_3 \\ h_3 &= -K^1_1 - K^2_2 + K^3_3 \end{aligned} \tag{8.19}$$

¹⁶ Note that in the original analysis of [6, 8, 17, 85], the symmetry walls are not included; the KM algebra that arises has the 3×3 Cartan matrix

$$A_{ij} = \begin{pmatrix} 2 & -2 & -2 \\ -2 & 2 & -2 \\ -2 & -2 & 2 \end{pmatrix}$$

and its fundamental Weyl chamber (radially projected on the hyperbolic plane H_2) is the ideal equilateral triangle having its three vertices at infinity.

manifestly showing how the overextended CSA generator h_3 enlarges the original $sl(3, \mathbb{R})$ generated by (e_1, f_1, h_1) and (e_2, f_2, h_2) to the Lie algebra $gl(3, \mathbb{R})$. The CSA generators are related to the ‘central charge’ generator c by

$$c = h_2 + h_3 = -K^1_1. \quad (8.20)$$

This is indeed the expected result: in the reduction of gravity to *two* dimensions (i.e., with two commuting Killing vectors), the central charge does not act on the internal degrees of freedom, but as a scaling on the conformal factor [12, 63, 88] (here realized as the 1–1 component of the vierbein). The affine level counting operator d is given by [38]

$$d = h_1 + h_2 + h_3 = -K^2_2 \quad (8.21)$$

and in contradistinction to c it does act on the internal volume in the reduction to two dimensions (i.e., the ‘dilaton’ ρ in the notation of [12, 88]). The operator d may be viewed as the L_0 operator of a full Witt–Virasoro algebra enlarging the Geroch group via a semidirect product [65].

The irreducible representations of $SL(3, \mathbb{R})$ are most conveniently characterized by their Dynkin labels. Let us recall that these labels are non-negative integers which characterize any highest-weight representation of any finite-dimensional Lie algebra. Let us consider a representation with maximal vector v_Λ , of (highest) weight Λ , i.e., such that for any element h in the CSA, $h(v_\Lambda) = \Lambda(h)v_\Lambda$, where the Lie algebra acts on the representation vector space (or ‘module’), and where Λ is a linear form on the CSA. The Dynkin labels of this representation are defined as $p_i(\Lambda) \equiv 2\langle\alpha_i|\Lambda\rangle/\langle\alpha_i|\alpha_i\rangle$, where α_i are simple roots of the Lie algebra under consideration. Note that in the case of simply laced algebras such as $sl(3, \mathbb{R})$ and AE_3 , the simple roots have all a squared length equal to 2, so that the definition of the labels reduces to $p_i(\Lambda) = \langle\alpha_i|\Lambda\rangle$ (while the Cartan matrix similarly simplifies to $A_{ij} = \langle\alpha_i|\alpha_j\rangle$). In the case of $sl(3, \mathbb{R})$ we have two simple roots, α_1 and α_2 , and therefore two Dynkin labels: $p_1(\Lambda) = \langle\alpha_1|\Lambda\rangle$, $p_2(\Lambda) = \langle\alpha_2|\Lambda\rangle$. In terms of the Young tableau description of $sl(3, \mathbb{R})$ representations, the first Dynkin label p_1 counts the number of columns having two boxes, while p_2 counts the number of columns having only one box. For instance, $(p_1, p_2) = (1, 0)$ labels an antisymmetric two-index tensor, while $(p_1, p_2) = (0, 2)$ denotes a symmetric two-index tensor. Note also that the dimension of the $sl(3, \mathbb{R})$ representation described by the labels (p_1, p_2) is $(p_1 + 1)(p_2 + 1)(p_1 + p_2 + 2)/2$.

Let us determine the representations of $sl(3, \mathbb{R})$ appearing at the first level ($\ell = 1$) of AE_3 .¹⁷ At this level, one sees that, under the adjoint action of $sl(3, \mathbb{R})$, i.e., of (e_1, f_1, h_1) and (e_2, f_2, h_2) , the extra Chevalley–Serre generator f_3 is a maximal vector. Indeed,

$$e_1(f_3) \equiv [e_1, f_3] = 0 \quad e_2(f_3) \equiv [e_2, f_3] = 0. \quad (8.22)$$

The weight of f_3 is $\Lambda = -\alpha_3$ because, by definition $h(f_3) \equiv [h, f_3] = -\alpha_3(h)f_3$. Hence, the Dynkin labels of the representation built on the maximal vector f_3 are $p_i = -\langle\alpha_i|\alpha_3\rangle = -A_{i3}$ with $i = 1, 2$. Explicitly, this gives $(p_1, p_2) = (0, 2)$. These labels also show up as the eigenvalues of the actions of the (specially normalized) Cartan generators (h_1, h_2) on the maximal vector f_3 , namely,

$$h_1(f_3) \equiv [h_1, f_3] = 0 \quad h_2(f_3) \equiv [h_2, f_3] = 2f_3 \quad (8.23)$$

As we mentioned above, the representation $(p_1, p_2) = (0, 2)$ corresponds to a symmetric (two-index) tensor.

Thus, at the levels ± 1 we have AE_3 generators which can be represented as symmetric tensors $E^{ij} = E^{ji}$ and $F_{ij} = F_{ji}$. One verifies that all algebra relations are satisfied

¹⁷ This analysis is analogous to that performed in [38] for the affine subalgebra $A_1^{(1)}$.

with $(a_{(ij)} \equiv (a_{ij} + a_{ji})/2)$

$$\begin{aligned} [K^i_j, E^{kl}] &= \delta_j^k E^{il} + \delta_j^l E^{ki} \\ [K^i_j, F_{kl}] &= -\delta_k^i F_{jl} - \delta_l^i F_{kj} \\ [E^{ij}, F_{kl}] &= 2\delta_{(k}^{(i} K^{j)l)} - \delta_k^{(i} \delta_l^{j)} (K^1_1 + K^2_2 + K^3_3) \\ \langle F_{ij} | E^{kl} \rangle &= \delta_i^{(k} \delta_j^{l)} \end{aligned} \tag{8.24}$$

and the identifications

$$e_3 = E^{33}, \quad f_3 = F_{33}. \tag{8.25}$$

As one proceeds to higher levels, the classification of $SL(3, \mathbb{R})$ representations becomes rapidly more complicated due to the exponential increase in the number of representations with level ℓ . Generally, the representations that can occur at level $\ell + 1$ must be contained in the product of the level- ℓ representations with the level-1 representation $(0, 2)$. Working out these products is elementary, but cumbersome. Moreover, many of the representations constructed in this way will drop out (i.e., not appear as elements of the AE_3 Lie algebra). The complications are not yet visible at low levels; for instance, the level-2 generator $E^{ab|jk} \equiv \varepsilon^{abi} E_i^{jk}$, with labels $(1, 2)$, is straightforwardly obtained by commuting two level-1 elements

$$[E^{ij}, E^{kl}] = \varepsilon^{mk(i} E_m^{j)l} + \varepsilon^{ml(i} E_m^{j)k}. \tag{8.26}$$

A more economical way to sift the relevant representations is to work out the relation between Dynkin labels and the associated highest weights, using the fact that the highest weights of the adjoint representation are the roots. More precisely, the maximal vectors being (as exemplified above at level 1) of the ‘lowering type’, the corresponding highest weights are *negative* roots, say $\Lambda = -\alpha$ with α of the form (8.16) with *non-negative* integers ℓ, m, n . Working out the Dynkin labels of $\Lambda = -\alpha$ by computing the scalar products with the two simple roots of $sl(3, \mathbb{R})$ then yields

$$p_1 \equiv p = n - 2m, \quad p_2 \equiv q = 2\ell + m - 2n. \tag{8.27}$$

As indicated, we shall henceforth use the notation $(p_1, p_2) \equiv (p, q)$ for the Dynkin labels. This formula is restrictive because all the integers entering it must be non-negative.

Inverting this relation we obtain

$$m = \frac{2}{3}\ell - \frac{2}{3}p - \frac{1}{3}q \quad n = \frac{4}{3}\ell - \frac{1}{3}p - \frac{2}{3}q \tag{8.28}$$

with $n \geq 2m \geq 0$. A further restriction derives from the fact that the highest weight must be a root of AE_3 , namely, its square must be smaller or equal to 2:

$$\Lambda^2 = \frac{2}{3}(p^2 + q^2 + pq - \ell^2) \leq 2. \tag{8.29}$$

Consequently, the representations occurring at level ℓ must belong to the list of all the solutions of (8.28) which are such that the labels m, n, p, q are non-negative integers and the highest weight Λ is a root, i.e., $\Lambda^2 \leq 2$. These simple diophantine equations/inequalities can be easily evaluated manually up to rather high levels.

However, the task is not finished. Although the above procedure substantially reduces the number of possibilities, it does not tell us how often (including zero times!) a given representation appears (i.e., its *outer multiplicity*). For this purpose, we have to make use of more detailed information about AE_3 , namely the root multiplicities computed in [38, 66]. Matching the combined weight diagrams with the root multiplicities listed in table H_3 on page 215 of [66], we obtain the following representations in the decomposition of AE_3 w.r.t. its

$sl(3, \mathbb{R})$ subalgebra up to level $\ell \leq 5$:¹⁸

$$\begin{aligned}
 \ell = 1 & \rightarrow (p, q) = (0, 2) \\
 \ell = 2 & \rightarrow (p, q) = (1, 2) \\
 \ell = 3 & \rightarrow (p, q) = (2, 2) \\
 & \qquad \qquad (1, 1) \\
 \ell = 4 & \rightarrow (p, q) = (3, 2) \\
 & \qquad \qquad (1, 3) \\
 & \qquad \qquad (2, 1) \quad (\text{occurs twice}) \\
 & \qquad \qquad (0, 2) \\
 & \qquad \qquad (1, 0) \qquad \qquad \qquad (8.30) \\
 \ell = 5 & \rightarrow (p, q) = (4, 2) \\
 & \qquad \qquad (2, 3) \quad (\text{occurs twice}) \\
 & \qquad \qquad (0, 4) \quad (\text{occurs twice}) \\
 & \qquad \qquad (3, 1) \quad (\text{occurs three times}) \\
 & \qquad \qquad (1, 2) \quad (\text{occurs four times}) \\
 & \qquad \qquad (2, 0) \quad (\text{occurs three times}) \\
 & \qquad \qquad (0, 1) \quad (\text{occurs twice})
 \end{aligned}$$

(we have not listed the representations compatible with (8.28) and (8.29), which drop out, i.e., have outer multiplicity zero). Going to yet higher levels will require knowledge of AE_3 root multiplicities beyond those listed in [38, 66, 68]. There is, however, an infinite set of admissible Dynkin labels that can be readily identified by searching for ‘affine’ highest weights for which $m = 0$ in (8.16). They are

$$(p, q) = (\ell - 1, 2) \iff (m, n) = (0, \ell - 1). \quad (8.31)$$

Because the associated highest weights all obey $\Lambda^2 = 2$ all these representations (and their ‘transposed’ representations) appear with outer multiplicity 1 independent of ℓ . A second series of affine representations is

$$(p, q) = (\ell, 0) \iff (m, n) = (0, \ell). \quad (8.32)$$

Now we have $\Lambda^2 = 0$, and the corresponding representations have outer multiplicity zero because the corresponding states are already contained as lower weight states in the representations (8.31).

The Lie-algebra elements corresponding to (8.31) are thus given by the two conjugate infinite towers of $sl(3, \mathbb{R})$ representations (for $\ell = 1, 2, 3, \dots$)

$$E_{i_1 \dots i_{\ell-1}}^{jk}, \quad F^{i_1 \dots i_{\ell-1}}{}_{jk}. \quad (8.33)$$

The affine representations are distinguished because they contain the affine subalgebra $A_1^{(1)} \subset AE_3$. This embedding was already studied in [38], but in view of future applications and because the identification is a little subtle we here spell out some more details. The affine subalgebra is identified by requiring its elements to commute with the central charge c (8.20). From this requirement we infer that we must truncate the full algebra to the subalgebra generated by the elements $K^i{}_j$ for $i, j \in \{2, 3\}$ and those generators in (8.33) with $i_1 = \dots i_{\ell-1} = 1$ and $j, k \in \{2, 3\}$. In physicists’ notation, the affine subalgebra is spanned by the current algebra generators $\{T_m^-, T_m^3, T_m^+\}$ and the central charge c . The affine level $m \in \mathbb{Z}$ is

¹⁸ This decomposition is now available for levels $\ell \leq 56$; see <http://www.arXiv.org/e-print/hep-th/0301017>.

the eigenvalue of the affine level counting operator d : $[d, T_m] = mT_m$. We therefore conclude that the affine level-0 sector¹⁹ consists of the generators of the Ehlers $SL(2, \mathbb{R})$ group, namely,

$$T_0^- = f_3 (\equiv F_{33}), \quad T_0^3 = h_3, \quad T_0^+ = e_3 (\equiv E^{33}). \quad (8.34)$$

At affine level $m = -1$ we have

$$\begin{aligned} T_{-1}^- &= e_2 (\equiv K^2_3) \\ T_{-1}^3 &= [e_2, e_3] = 2E^{23} \\ T_{-1}^+ &= [[e_2, e_3], e_3] = 2[E^{23}, E^{33}] = 2E_1^{23}. \end{aligned} \quad (8.35)$$

Similarly, at affine level $m = +1$ we obtain

$$\begin{aligned} T_1^- &= [[f_2, f_3], f_3] = 2[F_{23}, F_{33}] = 2F^1_{23} \\ T_1^3 &= [f_2, f_3] = F_{23} \\ T_1^+ &= f_2 (\equiv K^3_2) \end{aligned} \quad (8.36)$$

and so on for higher affine levels. The precise identification of the affine subalgebra is of crucial importance for understanding the embedding of the Geroch group into the full hyperbolic algebra.

Other cases of interest, where similar decompositions can be worked out include the indefinite KM algebras E_{10} and E_{11} , both of which have been conjectured to appear in $D = 11$ supergravity and M-Theory, see [62, 64, 28] and [101], respectively. The first six rungs in the $sl(10, \mathbb{R})$ level decomposition of E_{10} have been worked out in [28], and are extended to higher levels, as well as to E_{11} , in [39].

8.3. Nonlinear σ -models in one dimension

Notwithstanding the fact that we know even less about the ‘groups’ associated with indefinite KM algebras, we will formulate nonlinear σ -models in one time dimension and thereby provide an effective and unified description of the asymptotic BKL dynamics for several physically important models. The basic object of interest is a one-parameter dependent KM group element $\mathcal{V} = \mathcal{V}(t)$, which is assumed to be an element of the coset space G/K , where G is the group obtained by formal exponentiation of the KM algebra \mathfrak{g} , and K its maximal compact subgroup, which is again obtained by formal exponentiation of the associated maximal compact subalgebra \mathfrak{k} defined above. (In the ‘transpose’ notation defined above, the group K is the group of ‘orthogonal elements’: $k^T = k^{-1}$.) For finite-dimensional $\mathfrak{g}(A)$ our definitions reduce to the usual ones, whereas for indefinite KM algebras they are formal constructs to begin with. (Formal constructs similar to our objects $\mathcal{V}, \mathcal{V}\mathcal{V}^{-1}$, etc have been used in somewhat different settings in [19, 20, 101].) However, to ensure that our definitions are meaningful operationally, we will make sure at every step that any finite truncation of the model is well defined and can be worked out explicitly in a finite number of steps.

In physical terms, \mathcal{V} can be thought of as an extension of the vielbein of general relativity, with G and K as generalizations of the $GL(d, \mathbb{R})$ and local Lorentz symmetries of general relativity. For infinite-dimensional G , the object \mathcal{V} thus becomes an ‘ ∞ -bein’. It is then natural to associate with this vielbein a ‘metric’, namely,

$$\mathcal{M} := \mathcal{V}^T \mathcal{V} \quad (8.37)$$

which is invariant under the left action ($\mathcal{V} \rightarrow k\mathcal{V}$) of the ‘Lorentz group’ K (actually the truncation of \mathcal{M} to the relevant $GL(n, \mathbb{R})$ subgroup of the KM algebras entering the models

¹⁹ Recall that the affine level m must not be confused with the level ℓ used in our decomposition of AE_3 , cf definition (8.16).

we study turns out to correspond to the matrix defined by the *contravariant* components, g^{ij} , of the spatial metric used in section 4). Exploiting this invariance, we can formally bring \mathcal{V} into a ‘triangular gauge’,

$$\mathcal{V} = \mathcal{A} \cdot \mathcal{N} \implies \mathcal{M} = \mathcal{N}^T \mathcal{A}^2 \mathcal{N} \quad (8.38)$$

where the Abelian part \mathcal{A} belongs to the exponentiation of the CSA, and the nilpotent part \mathcal{N} to the exponentiation of \mathfrak{n}^+ thus recovering the formulae which we have already used in sections 4.1 and 4.2. This formal Iwasawa decomposition, which is an infinite-dimensional generalization of that we used before, can be made fully explicit by decomposing \mathcal{A} and \mathcal{N} in terms of bases of \mathfrak{h} and \mathfrak{n}^+ ,

$$\mathcal{A}(t) = \exp(\beta^\mu(t) H_\mu), \quad \mathcal{N}(t) = \exp\left(\sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} v_{\alpha,s}(t) E_{\alpha,s}\right) \quad (8.39)$$

where Δ_+ denotes the set of positive roots. The components β^μ , parametrizing a generic element in the CSA \mathfrak{h} , will turn out to be in direct correspondence with the logarithmic scale factors $\beta^\mu = (\beta^a, \phi)$ introduced in section 3. We anticipated this correspondence by using the same notation. (As explained above, the apparently ‘wrong sign’ of the exponents of \mathcal{A} is due to the fact that \mathcal{M} will correspond to the inverse of the spatial metric.) The main technical difference with the kind of Iwasawa decompositions used in the foregoing sections is that now the matrix $\mathcal{V}(t)$ is infinite dimensional for indefinite $\mathfrak{g}(A)$. Observe that for finite-dimensional matrices, there was no need to worry about root multiplicities, as these are always 1. By contrast, there are now infinitely many v , and consequently \mathcal{N} contains an infinite tower of new degrees of freedom. Next we define

$$\mathcal{N}\mathcal{N}^{-1} = \sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} j_{\alpha,s} E_{\alpha,s} \in \mathfrak{n}^+ \quad (8.40)$$

with

$$j_{\alpha,s} = \dot{v}_{\alpha,s} + 'v\dot{v} + v\dot{v}\dot{v} + \dots' \quad (8.41)$$

(we put quotation marks to avoid having to write out the indices). To define a Lagrangian we proceed in the usual way. First we consider the quantity

$$\dot{\mathcal{V}}\mathcal{V}^{-1} = \dot{\beta}^\mu H_\mu + \sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} e^{\alpha(\beta)} j_{\alpha,s} E_{\alpha,s} \quad (8.42)$$

which has values in the Lie algebra $\mathfrak{g}(A)$. Here, we have set

$$\alpha(\beta) \equiv \alpha_\mu \beta^\mu \quad (8.43)$$

for the value of the root α (\equiv linear form) on the CSA element $\beta = \beta^\mu H_\mu$. Next we define

$$P := \frac{1}{2}(\dot{\mathcal{V}}\mathcal{V}^{-1} + (\dot{\mathcal{V}}\mathcal{V}^{-1})^T) = \dot{\beta}^\mu H_\mu + \frac{1}{2} \sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} j_{\alpha,s} e^{\alpha(\beta)} (E_{\alpha,s} + E_{-\alpha,s}) \in \mathfrak{g} \oplus \mathfrak{k}$$

where we arranged the basis so that $E_{\alpha,s}^T = E_{-\alpha,s}$. Then we define a KM-invariant σ -model action as $\int dt \mathcal{L}$ where the Lagrangian is defined by using the KM-invariant bilinear form $\langle \cdot | \cdot \rangle$ (cf (8.12))

$$\mathcal{L} = \frac{1}{2} n^{-1} \langle P | P \rangle = n^{-1} \left(\frac{1}{2} G_{\mu\nu} \dot{\beta}^\mu \dot{\beta}^\nu + \frac{1}{4} \sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} \exp(2\alpha(\beta)) j_{\alpha,s} j_{\alpha,s} \right). \quad (8.44)$$

Here, the Lorentzian metric $G_{\mu\nu}$ is the restriction of the invariant bilinear form to the CSA. It can be identified with the metric in the space of the scale factors, which is why we adopt the same notation. We have introduced in this Lagrangian a ‘lapse function’ n (not to be confused with the lapse function N introduced before), which ensures that our formalism is invariant under reparametrizations of the time variable. Finally, we can simply describe our Lagrangian (8.44) as that of a *null geodesic* in the coset space G/K .

Taking the algebra AE_3 as an example, this Lagrangian contains the Kasner Lagrangian (3.10) (without dilaton) as a special truncation. More specifically, retaining only the level-0 fields (corresponding to a $GL(3, \mathbb{R})/O(3)$ σ -model)

$$\mathcal{V}(t)|_{\ell=0} = \exp(h^a{}_b(t)K^b{}_a) \tag{8.45}$$

and defining from $h^a{}_b$ a vielbein by matrix exponentiation $e^a{}_b \equiv (e^h)^a{}_b$, and a corresponding contravariant metric $g^{ab} = e^a{}_c e^b{}_c$, one checks that the bilinear form (8.18) reproduces (half) the Lagrangian (3.10). (This computation works more generally for any $GL(n, \mathbb{R})$, and was already used in [28] in the $GL(10, \mathbb{R})$ decomposition of E_{10} .) The level-0 sector is thus associated with the Einsteinian dynamics of a spatial dreibein depending only on time.

Let us then consider the fields ϕ_{ij} associated with the level-1 generators E^{ij} . This leads to a truncation of our KM-invariant σ -model to the levels $\ell = 0, 1$, i.e.,

$$\mathcal{V}(t)|_{\ell=0,1} = \exp(h^a{}_b(t)K^b{}_a) \exp(\phi_{ab}E^{ab}). \tag{8.46}$$

In the gauge $n = 1$, the Lagrangian now has the form $\mathcal{L} \sim (g^{-1}\dot{g})^2 + g^{-1}g^{-1}\dot{\phi}\dot{\phi}$, where g denotes the *covariant* metric g_{ij} . As the ϕ_{ij} enter only through their time derivatives, their conjugate momenta Π^{ij} are constants of the motion. Eliminating the ϕ in favour of the constant momenta Π by a partial Legendre transformation yields the reduced Lagrangian (for the dynamics of $g_{ij}(t)$) of the form $\mathcal{L} \sim (g^{-1}\dot{g})^2 - g^{+1}g^{+1}\Pi\Pi$. In other words, the elimination of the ϕ has generated a potential $V_\phi = V_\phi(g)$ (in the gauge $n = 1$)

$$V_\phi(g) \propto +g_{ij}g_{kl}\Pi^{ik}\Pi^{jl}. \tag{8.47}$$

It is then easy to check that this potential can be identified with the leading (weight -2) gravitational potential in (6.14) (corresponding to the gauge $\tilde{N} = 1$), namely $\mathcal{V}_G^{\text{leading}} \equiv (-gR)^{\text{leading}} = +\frac{1}{4}gg_{ai}g^{bj}g^{ck}C^a{}_{bc}C^i{}_{jk}$ under the identification of the structure constants $C^i{}_{jk}$ with the momenta conjugate to ϕ_{ij} :

$$\Pi^{ij} \propto \varepsilon^{kl(i}C^j)_{kl}. \tag{8.48}$$

Note that the trace $C^j{}_{jk}$ drops out of this relation, and that Π^{ij} is of weight 1, like ε^{ijk} .

Consequently, when neglecting the subleading gravitational walls $\propto F_a$ in (6.14) and in four spacetime dimensions, the BKL dynamics at each given spatial point x_0 is equivalent to the $\ell = 0, \pm 1$ truncation of the AE_3 -invariant dynamics defined by (8.44). The fields $\phi_{ij}(t)$ parametrizing the components of the AE_3 coset element along the $\ell = 1$ generators are canonically conjugate to the structure constants $C^i{}_{jk}(t, x_0)$.

Similarly, one would like to associate the generators (8.33) with higher-order spatial gradients. However, the proper physical interpretation of these fields as well as of the other higher-level components remains yet to be found. In the case of the relation between supergravity in $D = 11$ and the E_{10} coset model, one could pursue the correspondence between spacetime fields and coset coordinates up to the $gl(10, \mathbb{R})$ level $\ell = 3$ included (corresponding to height 29) [28]. The correspondence worked, thanks to several ‘miraculous’ agreements between the numerical coefficients appearing in both Lagrangians.

Varying (8.44) w.r.t. the lapse function n gives rise to the constraint that the coset Lagrangian vanish. Let us define the canonical momenta

$$\pi_\mu := \frac{\delta \mathcal{L}}{\delta \dot{\beta}^\mu} = n^{-1}G_{\mu\nu}\dot{\beta}^\nu \tag{8.49}$$

and the (non-canonical) momentum-like variables

$$\Pi_{\alpha,s} := \frac{\delta \mathcal{L}}{\delta j_{\alpha,s}} = \frac{1}{2} n^{-1} e^{2\alpha(\beta)} j_{\alpha,s}. \quad (8.50)$$

The latter variables are related to the momenta canonically conjugate to the $v_{\alpha,s}$, i.e., $p_{\alpha,s} := \delta \mathcal{L} / \delta \dot{v}_{\alpha,s}$, by expressions of the form $\Pi_{\alpha,s} = p_{\alpha,s} + \nu p + \nu \nu p + \dots$, see section 8.4.

In terms of these variables, the KM-invariant Hamiltonian corresponding to the above Lagrangian reads

$$\mathcal{H}(\beta, \pi, v, p) = n \left(\frac{1}{2} G^{\mu\nu} \pi_\mu \pi_\nu + \sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} e^{-2\alpha(\beta)} \Pi_{\alpha,s} \Pi_{\alpha,s} \right). \quad (8.51)$$

The constraint of a vanishing Lagrangian becomes that of a vanishing Hamiltonian:

$$\mathcal{H} \approx 0. \quad (8.52)$$

The momentum-like variables $\Pi_{\alpha,s}$ can be thought of as infinite-dimensional generalizations of the asymptotically frozen combinations $P^j_a \mathcal{N}^b_j$ used in section 5.1, cf equation (5.14). They do not Poisson-commute with one another in general. Instead, one has Poisson brackets of the form $\{\Pi_{\alpha,r}, \Pi_{\beta,s}\} = \Omega_{\alpha,r\beta,s}^{\gamma,t}(\nu) \Pi_{\gamma,t}$. Note that π_μ is timelike or null because the potential is manifestly non-negative; the motion in the CSA is timelike—in fact, in the limit to be studied below, a broken future-oriented lightlike line.

Because the coefficients of the exponentials in (8.51) are non-negative we can now apply exactly the same reasoning as in section 5. The crucial point is that the Cartan variables β do not enter the commutation relations of the Π as indicated above. Therefore, as above, the off-diagonal components $v_{\alpha,s}$ and the momentum-like variables $\Pi_{\alpha,s}$ get frozen asymptotically, provided all spacelike walls are ‘behind the motion’ so that they do not conflict with the BKL limit. As before, we can introduce hyperbolic polar coordinates for parametrizing the dynamics of the Cartan variables $\beta^\mu = \rho \gamma^\mu$. When taking the limit $\rho \rightarrow \infty$, the Hamiltonian (8.52) behaves as

$$\mathcal{H} \sim \mathcal{H}_{\Delta_+}(\beta, \pi) = \frac{1}{2} G^{\mu\nu} \pi_\mu \pi_\nu + \sum_{\alpha \in \Delta_+} K_\alpha e^{-2\alpha(\beta)} \quad (8.53)$$

with constants $K_\alpha \geq 0$. This looks like a Toda Hamiltonian, except that the sum extends over all roots, rather than only the simple ones, and that the underlying KM algebra is indefinite, and not only finite or affine as in standard Toda theory. However, the *dominant potential walls* in the limit $\rho \rightarrow \infty$ are given by those terms $K_\alpha e^{-2\alpha(\beta)}$ for which α is a *simple root*. Indeed, by definition, the non-simple roots $\alpha = n_1 \alpha_1 + n_2 \alpha_2 + \dots$ give rise to potential terms $\propto [e^{-2\alpha_1(\beta)}]^{n_1} [e^{-2\alpha_2(\beta)}]^{n_2} \dots$ which are subdominant w.r.t. the set of potentials corresponding to the simple roots.²⁰ Therefore, the σ -model Hamiltonian is asymptotically equivalent to the truncation of (8.53) to the simple roots:

$$\mathcal{H} \sim \mathcal{H}_{\text{simple}}(\beta, \pi) = \frac{1}{2} G^{\mu\nu} \pi_\mu \pi_\nu + \sum_{i=1}^r K_i e^{-2\alpha_i(\beta)}. \quad (8.54)$$

Such hyperbolic Toda models, restricted to the simple roots, were first introduced and studied in [42].

Furthermore, as explained in section 5.2, the potentials become sharp wall potentials in the BKL limit $\beta^\mu \rightarrow \infty$. Making again the ‘genericity assumption’ $K_i > 0$ for the dominant

²⁰ Perhaps a useful analogy is to think of a mountainscape (defined by Toda exponential potentials for *all* roots); when the mountaintops rise into the sky, only the nearest mountains (corresponding to the simple roots) remain visible to the observer in the valley (the Weyl chamber).

simple root contributions (which is subject to the same caveats as before), we finally obtain the asymptotic Hamiltonian

$$\mathcal{H}_\infty(\beta, \pi) := \lim_{\rho \rightarrow \infty} \mathcal{H}(\beta, \pi) = \frac{1}{2} \pi^\mu \pi_\mu + \sum_{i=1}^r \Theta(-2\alpha_i(\beta)) \tag{8.55}$$

where the sum is over the simple roots only, and the motion of the β^μ is confined to the fundamental Weyl chamber $\alpha_i(\beta) \geq 0$. Note that, in the present KM set-up, all the walls enter on the same footing. There is nothing left of the distinctions that entered our foregoing BKL-type studies between different types of walls (symmetry walls, gravitational walls, electric walls and so on). The only important characteristic of a wall is its height $\text{ht } \alpha \equiv n_1 + n_2 + \dots$ for a root decomposed along simple roots as $\alpha = n_1\alpha_1 + n_2\alpha_2 + \dots$.

The sum in (8.52) ranges not only over the real roots, which give rise to infinitely many timelike walls, but also over null and purely imaginary roots of the KM algebra giving rise to lightlike and spacelike walls, respectively. In view of our discussion in section 7, the significance of the latter for the KM billiard remains to be fully elucidated. Here, we only remark that imaginary roots are of no relevance for the Weyl group of the KM algebra, which by definition consists only of reflections against the timelike hyperplanes orthogonal to real roots.

We conclude that *in the limit where one goes to infinity in the Cartan directions, the dynamics of the Cartan degrees of freedom of the coset model become equivalent to a billiard motion within the Weyl chamber, subject to the zero-energy constraint $\mathcal{H}(\beta, \pi) = 0$* . Therefore, in the cases where the cosmological billiards that we discussed in the first part of this paper are of KM type, the gravitational models are asymptotically equivalent (modulo the imposition of the additional momentum and Gauss constraints) to the product over the spatial points of independent (1 + 0)-dimensional KM coset models G/K .

8.4. Integrability, chaos and consistent truncations

In this final subsection, we show that the one-dimensional KM σ -models are formally integrable. Then we address the issue of why this formal integrability is not incompatible with the occurrence of chaos in the billiard description. We also discuss various possible finite-dimensional truncations of our infinite-dimensional KM-invariant σ -model.

The equations of motion of the off-diagonal fields constitute an infinite system of nonlinear ordinary differential equations of second order. As usual, they are equivalent to the conservation of the $\mathfrak{g}(A)$ -valued Noether charge

$$\begin{aligned} J &= \mathcal{V}^{-1} P \mathcal{V} \\ &\equiv J^\mu H_\mu + \sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} (J_{\alpha,s} E_{\alpha,s} + J_{-\alpha,s} E_{-\alpha,s}) \end{aligned} \tag{8.56}$$

in the gauge $n = 1$. It is easily checked from the first line of this equation that

$$J = \mathcal{M}^{-1} \dot{\mathcal{M}}. \tag{8.57}$$

This equation can be formally solved as

$$\mathcal{M}(t) = \mathcal{M}(0) \cdot e^{tJ}. \tag{8.58}$$

For finite-dimensional matrices \mathcal{M} , the system is thus completely integrable. This fact was already used in section 4 where we wrote down the exact solution (4.1) of a $GL(n, \mathbb{R})/O(n)$ model. More precisely, the $GL(n, \mathbb{R})$ analogue of the (Lie-algebra valued) conserved charge J is $\pi^i_j = \frac{1}{2} g^{ik} \dot{g}_{kj}$, whose conservation is clear from equation (3.13). (Note again that the

KM gauge $n = 1$ corresponds to $\tilde{N} = 1$ so that the KM time variable t corresponds to the gravitational time scale τ .) In the finite-dimensional case, we could derive the general solution by *diagonalizing* the constant matrix π^i_j , which led to the Kasner solution, written in terms of the eigenvalues v^a of π^i_j and of the diagonalizing matrix L as in (4.1). Then, thanks to finite dimensionality of the diagonalizing matrix, we showed there that, after a finite transition time (subsequently interpreted as linked to the finite number of collisions on the symmetry walls needed to reorder the eigenvalues v^a), the resulting asymptotic motion assumed a very simple monotonic form.

By contrast, for infinite-dimensional matrices the existence of the formal solution (8.58) does not necessarily imply regular behaviour for the CSA degrees of freedom in the asymptotic limit where all other degrees of freedom get frozen. First of all, and in marked contrast to the finite-dimensional case, it is not known for indefinite KM algebras whether a generic Lie-algebra element J can always be ‘diagonalized’, i.e., conjugated into the CSA by an element of the KM group G , see however [67]. Indeed, if that were the case, it would mean that a generic solution of the KM coset dynamics could be conjugated to a simple monotonic and diagonal Kasner-like solution $\beta^\mu = v^\mu t + \beta_0^\mu$ in the CSA, a conclusion difficult to reconcile with the chaotic motion that obtains in the BKL limit if the KM algebra is hyperbolic. Secondly, even if J could be conjugated into the CSA (as might be the case for certain non-generic elements) there remains the possibility that the required diagonalizing matrix (an infinite-dimensional analogue of the matrix L in (4.1)) introduces an infinite number of effective collisions in the time development of the analogue of (4.1), because the number of effective collisions was found to increase with the rank of the matrix. In the latter case, the chaotic motion for the CSA degrees of freedom would be the result of the projection onto a finite-dimensional space of a regular geodesic motion taking place in an infinite-dimensional phase space. In the former case, it might be the geodesic motion on the infinite-dimensional coset space G/K which could be intrinsically chaotic. We will leave further investigation of these delicate mathematical questions to future work.

In order to better understand why the existence of an infinite number of conserved quantities $\{J^\mu, J_{\alpha,s}, J_{-\alpha,s}\}$ does not rule out chaos, let us consider possible *consistent truncations* of our σ -model. By a ‘consistent truncation’ we here mean a submodel whose solutions are solutions of the full model. In order to formalize this notion, let us introduce a *gradation* \mathcal{D} of root space: the \mathcal{D} -degree of a root $\alpha = n_1\alpha_1 + n_2\alpha_2 + \dots$ is defined as $\mathcal{D}(\alpha) := n_1\mathcal{D}_1 + n_2\mathcal{D}_2 + \dots$, where $(\mathcal{D}_1, \mathcal{D}_2, \dots)$ is a given set of non-negative integers. Examples are the $sl(3, \mathbb{R})$ -level ℓ , the affine level m or the height $\mathcal{D}(\alpha) \equiv \text{ht}(\alpha) = n_1 + n_2 + \dots$. In terms of an expansion according to increasing gradation \mathcal{D} , the full σ -model Lagrangian (8.44) has the structure (with all coefficients suppressed and in the gauge $n = 1$)

$$\begin{aligned} \mathcal{L} \sim & \dot{\beta}^2 + e^{2\alpha_1(\beta)}[\dot{v}^1]^2 + \exp(2\alpha_2(\beta))[\dot{v}^2 + v^1\dot{v}^1]^2 \\ & + e^{2\alpha_3(\beta)}[\dot{v}^3 + v^1\dot{v}^2 + v^2\dot{v}^1 + v^1v^1\dot{v}^1]^2 + \dots \end{aligned} \quad (8.59)$$

The notation here is somewhat schematic: the lower indices on the roots and the upper indices on the off-diagonal fields refer to the gradation \mathcal{D} (so $\alpha_1, \alpha_2, \dots$ are *not* simple roots here). Furthermore, the degree zero term coincides with the free CSA kinetic term $\dot{\beta}^2$ only if $\mathcal{D}_i > 0$ for all i ; for other gradations, the level-0 sector is described by the σ -model Lagrangian for the respective level-0 subalgebra (which may be non-Abelian as was the case for the level ℓ used in section 4.2). The various terms within parentheses correspond to the gradation of the $j_{\alpha,s}$ above: $j_1 \sim \dot{v}^1$, $j_2 \sim \dot{v}^2 + v^1\dot{v}^1$, $j_3 \sim \dot{v}^3 + v^1\dot{v}^2 + v^2\dot{v}^1 + v^1v^1\dot{v}^1, \dots$. The momenta canonically conjugate to the off-diagonal variables v^1, v^2, \dots have a similar graded structure

$$\begin{aligned}
 p_1 &\sim e^{2\alpha_1(\beta)} j_1 + e^{2\alpha_2(\beta)} j_2 v^1 + e^{2\alpha_3(\beta)} j_3 [v^2 + v^1 v^1] + \dots \\
 &= \Pi_1 + \Pi_2 v^1 + \Pi_3 (v^2 + v^1 v^1) + \dots, \\
 p_2 &\sim e^{2\alpha_2(\beta)} j_2 + e^{2\alpha_3(\beta)} j_3 v^1 + \dots = \Pi_2 + \Pi_3 v^1 + \dots, \\
 p_3 &\sim e^{2\alpha_3(\beta)} j_3 + \dots = \Pi_3 + \dots,
 \end{aligned}
 \tag{8.60}$$

where, consistently with the above definition, we have introduced the short-hand notation $\Pi_n \equiv \exp 2\alpha_n(\beta) j_n$. Formally, the triangular relations above can be inverted iteratively to obtain infinite series of the form

$$\begin{aligned}
 \Pi_1 &\sim p_1 + p_2 v^1 + p_3 [v^2 + v^1 v^1] + \dots, \\
 \Pi_2 &\sim p_2 + p_3 v^1 + p_4 [v^2 + v^1 v^1] + \dots, \\
 \Pi_3 &\sim p_3 + p_4 v^1 + \dots.
 \end{aligned}
 \tag{8.61}$$

This yields a formal expression for the Hamiltonian in terms of the canonical variables $(\beta, \pi; v^n, p_n)$, where n again refers to the gradation,

$$\mathcal{H}(\beta, \pi; v^n, p_n) \sim \pi^2 + e^{-2\alpha_1(\beta)} \Pi_1^2 + e^{-2\alpha_2(\beta)} \Pi_2^2 + e^{-2\alpha_3(\beta)} \Pi_3^2 + \dots.
 \tag{8.62}$$

It is now easy to see that one can obtain a consistent *finite-dimensional* truncation of the dynamics by requiring that *all canonical momenta above a certain \mathcal{D} -degree n_0 vanish*. For instance, if we require $p_n = 0$ for $n \geq 3$ implying $\Pi_n = 0$ for all $n \geq 3$ we obtain a consistent dynamics following from the finite Hamiltonian

$$\mathcal{H}^{(2)}(\beta, \pi; v^1, p_1, v^2, p_2) \sim \pi^2 + e^{-2\alpha_1(\beta)} [p_1 + p_2 v^1]^2 + e^{-2\alpha_2(\beta)} p_2^2.
 \tag{8.63}$$

The finite-dimensional Lagrangian corresponding to this Hamiltonian reads

$$\mathcal{L}^{(2)} \sim \dot{\beta}^2 + e^{2\alpha_1(\beta)} [\dot{v}^1]^2 + e^{2\alpha_2(\beta)} [\dot{v}^2 + v^1 \dot{v}^1]^2.
 \tag{8.64}$$

and is obtained from the original Lagrangian (8.59) by setting equal to zero all j_n for $n \geq 3$, where, for instance, $j_3 = \dot{v}^3 + v^1 \dot{v}^2 + v^2 \dot{v}^1 + v^1 v^1 \dot{v}^1$, and so on. Note that the highest-degree variable v^2 does not explicitly appear in the truncated Hamiltonian $\mathcal{H}^{(2)}$. This implies that the corresponding highest-degree canonical momentum p_2 is a constant of the motion. However, non-trivial dynamics is obtained for the remaining degrees of freedom $\beta, \pi; v^1, p_1$. Note also that a *particular* set of solutions of the \mathcal{D} -degree-2 dynamics above is obtained when the constant of motion p_2 happens to vanish. This particular case of the \mathcal{D} -degree-2 dynamics is simply the \mathcal{D} -degree-1 truncation of the original dynamics, obtained by setting equal to zero all momenta above level 2, i.e., $p_2 = p_3 = \dots = 0$. The corresponding Hamiltonian is simply

$$\mathcal{H}^{(1)}(\beta, \pi; v^1, p_1) \sim \pi^2 + e^{-2\alpha_1(\beta)} p_1^2.
 \tag{8.65}$$

Again the highest-degree momentum, which is now p_1 , is a constant of the motion, so that $\mathcal{H}^{(1)}$ directly defines the reduced dynamics of the Cartan variables β, π .

For $\mathcal{D} \equiv \text{ht}$, the truncated dynamics (8.65) at degree $\mathcal{D} = 1$ defines the finite hyperbolic Toda model (8.54) involving only the simple roots. We have seen in the previous subsection that this height-1 Hamiltonian universally describes the asymptotic dynamics of the full σ -model. (Its BKL limit directly yields the universal Weyl-chamber billiard (8.55).) On the other hand, when the chosen gradation is the $sl(3)$ level of AE_3 , we obtain the low-level truncations of section 8.2 above: the level-0 truncation ($p_1 = \dots = 0$) yields the Kasner dynamics, while the level-1 dynamics describes the leading effect of the gravitational walls. A similar hierarchy of truncations appears in the E_{10} model of [28], where the analysis included levels $\ell = 0, 1, 2, 3$ w.r.t. the subalgebra $sl(10, \mathbb{R}) \equiv A_9$ (see also [25]).

Let us now clarify the meaning of the *infinite* set of conserved charges $J_{\pm\alpha,s}$ within the context of the *finite-dimensional* truncations which we just discussed. To this aim, one expresses J in terms of the (non-canonical) Hamiltonian variables (β, π, ν, Π) :

$$J = \pi^\mu \mathcal{N}^{-1} H_\mu \mathcal{N} + \sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} e^{-2\alpha(\beta)} \Pi_{\alpha,s} \mathcal{N}^{-1} E_{\alpha,s} \mathcal{N} + \sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} \Pi_{\alpha,s} \mathcal{N}^{-1} E_{-\alpha,s} \mathcal{N}. \quad (8.66)$$

Here only the third term spreads over the full algebra $\mathfrak{g} = \mathfrak{n}^- \oplus \mathfrak{h} \oplus \mathfrak{n}^+$, while the first and second are entirely contained in the parabolic subalgebras $\mathfrak{h} \oplus \mathfrak{n}^+$ and \mathfrak{n}^+ , respectively. From this fact, it is easy to see that any truncation where the $\Pi_{\alpha,s}$ vanish for all roots strictly above some degree, i.e., for $\mathcal{D}(\alpha) > n_0$, implies the vanishing of the correspondingly low ‘negative-root charges’ $J_{-\alpha,s}$. As for the non-zero highest-degree momenta $\Pi_{\alpha_0,s}$ with $\mathcal{D}(\alpha_0) = n_0$, equation (8.66) simply yields $\Pi_{\alpha_0,s} = J_{-\alpha_0,s}$. We thus recover the result above concerning the constancy of the highest-degree momenta. We can then move up (always in \mathcal{D} -degree) by adding simple roots α_i to α_0 . Equation (8.66) then implies $\Pi_{\alpha_0-\alpha_i,s} = J_{-\alpha_0+\alpha_i,s} + c\nu_{\alpha_i} J_{-\alpha_0,s}$, where the constant c is determined by the commutation relations. Continuing in this manner until one reaches the degree zero, one obtains for all the negative-degree (non-vanishing) momenta expressions which are linear in the charges and polynomial of ascending order in the off-diagonal fields $\nu_{\alpha,s}$. Note, however, that these relations do not suffice to eliminate non-trivial degrees of freedom from the truncated Hamiltonian, i.e., to replace it by an (on-shell) equivalent Hamiltonian depending on fewer variables. (The only degrees of freedom that one can straightforwardly eliminate are the trivial highest-degree ones, whose associated ‘position variables’ do not enter the Hamiltonian.) One might think that one would get further relations, and eventually enough relations to eliminate the off-diagonal degrees of freedom, by considering the higher-degree components of equation (8.66). However, it does not work this way. The degree-zero projection, namely

$$\pi^\mu H_\mu + \sum_{\alpha \in \Delta_+} \sum_{s=1}^{\text{mult}(\alpha)} \Pi_{\alpha,s} (\mathcal{N}^{-1} E_{-\alpha,s} \mathcal{N})|_h = c^\mu H_\mu \quad (8.67)$$

(where c^μ is a constant vector) gives an expression for the degree-zero momenta π^μ which is linear in the charges and polynomial in the $\nu_{\alpha,s}$. On the other hand, the positive-degree projection of equation (8.66) yields for the positive-degree components of the charges $J_{\alpha,s}$ expressions which (for any positive degree) will involve ‘hidden’ off-diagonal fields $\nu_{\alpha,s}$ of degree $\mathcal{D}(\alpha) > n_0$. These new variables did not enter the truncated Hamiltonians (8.63), which shows explicitly that the infinite number of conserved charges $J_{\alpha,s}$ of the original σ -model do not provide sufficiently many *autonomous* constants of the motion for the truncated Hamiltonian to guarantee its ‘integrability’ in the trivial sense of allowing one to reduce the number of degrees of freedom to zero.

Let us also dispose of another paradox concerning the formal integrability of KM-related models. Even if we consider the hyperbolic Toda model defined by the truncation of our general σ -model to height 1, i.e., $\mathcal{H}_{\text{simple}}(\beta, \pi)$ given by equation (8.54) with (strictly) positive constants K_i , one finds that it has formal integrability features. Indeed, the model (8.54) admits a Lax pair. Setting $s_i \equiv \sqrt{\langle \alpha_i | \alpha_i \rangle} / 2$, rescaling the Chevalley–Serre generators as $e'_i = s_i e_i$, $f'_i = s_i f_i$, $h'_i = s_i^2 f_i$, and defining $W_{ij} \equiv \langle \alpha_i | \alpha_j \rangle$, we replace the variables (β^μ, π_μ) by (l^i, p^i) by $l^i \equiv \sqrt{K_i} e^{-\alpha_i(\beta)}$ and $\sum_j W_{ij} p^j \equiv G^{\mu\nu}(\alpha_i)_\mu \pi_\nu$. One can then check that the Lie-algebra valued expressions

$$M = \sum_i [p^i h'_i + l^i (e'_i + f'_i)], \quad K = \sum_i l^i (e'_i - f'_i), \quad (8.68)$$

satisfy the Lax-pair evolution equation $\dot{M} = [K, M]$. One would then expect to be able to write down (infinitely) many conserved quantities associated with the isospectrality of the evolution of $M(t)$. (In the finite-dimensional case all the traces of the matrix $M(t)$ are constants of the motion, see, e.g., [90] and references therein.) However, in the KM case this formal Lax-pair integrability does not yield any concrete constants of the motion because there is no useful analogue of the matrix trace in the KM algebra. The integrability properties of hyperbolic Toda theories have also been discussed in [42].

Let us note also in passing that there is full compatibility between our general result of the asymptotic constancy of all the off-diagonal degrees of freedom $\nu_{\alpha,s}, \Pi_{\alpha,s}$ and the existence of an infinite number of conserved quantities $J_{\alpha,s}$. Morally speaking, the two infinite sets of asymptotic constants of integration $\{\lim \nu_{\alpha,s}\}, \{\lim \Pi_{\alpha,s}\}$ correspond to the doubly infinite set of charges $\{J_{\pm\alpha,s}\}$ (in non-zero degree), and one can check the compatibility of this correspondence by using the results above concerning truncated models. As for the zero-level conservation law (8.67) it is also compatible with the asymptotic limit because, in the gauge we use here (analogous to the β -space picture of the billiard) all the components of the zero-degree momenta π^μ tend to zero asymptotically (because of the redshifts accumulated on the receding walls).

In summary, there is thus no contradiction between the formal integrability of our σ -model and the chaos that appears in its generic asymptotic dynamics. This is mainly due to the presence of infinitely many degrees of freedom, and the dynamics of the finite-dimensional truncations of the σ -model of strictly positive degree.

9. Conclusions

In this paper, we have shown that theories involving gravity admit a remarkable asymptotic description in the vicinity of a spacelike singularity in terms of billiards in hyperbolic space. Depending on whether the actual billiard has finite or infinite volume, the dynamical evolution of the local scale factors is chaotic (oscillatory) or monotonic (Kasner-like). The billiard, and in particular its volume, is a fundamental characteristic of the theory, in the sense that it is determined solely by the field content and the parameters in the Lagrangian, and not by the initial conditions (in the generic case; i.e., there may be initial conditions for which some walls are absent—and the billiard is changed—but these are exceptional). Although we have not investigated the physical implications of this property for cosmological scenarios (in particular, string-inspired cosmologies [14, 40, 41, 76, 99]), nor its quantum analogue, we believe that this result is interesting in its own right because it uncovers an intrinsic feature of gravitational theories. As discussed in section 8, the regularity properties of some billiards appear to give a powerful handle on possible hidden Kac–Moody symmetries, which remain to be exploited for their full worth.

Acknowledgments

The work of MH is supported in part by the ‘Actions de Recherche Concertées’ of the ‘Direction de la Recherche Scientifique, Communauté Française de Belgique’, by a ‘Pôle d’Attraction Interuniversitaire’ (Belgium), by IISN-Belgium (convention 4.4505.86), by Proyectos FONDECYT 1970151 and 7960001 (Chile) and by the European Commission RTN programme HPRN-CT-00131, in which he is associated with KU Leuven. HN is partially supported by the EU contract HPRN-CT-2000-00122. Both MH and HN would like to thank IHES for hospitality while this work was carried out.

Appendix. Asymptotic freezing: a simple model

We have seen in the text that the ‘non-diagonal’ phase space variables (Q, P) (i.e., all variables except the extended scale factors β^μ parametrizing the diagonal-Iwasawa-part of the metric and the dilaton), get frozen to constant values in the BKL limit. We provide here a more detailed understanding of this property by discussing a simpler model which captures the essential features of the Hamiltonian (5.25).

Consider a system with two canonically conjugate pairs $(q, p), (Q, P)$ and time-dependent Hamiltonian

$$H(q, p, Q, P; T) = \frac{1}{2}p^2 + \frac{1}{2}(f(Q, P))^2 \rho^k e^{-\rho q} \quad (k \text{ being any real number}) \quad (\text{A.1})$$

where $\rho \equiv e^T$, with T the time variable. We have simplified here the Hamiltonian (5.26) by eliminating the variable λ , i.e., by replacing it by an explicit function of the coordinate time T , in the approximation of a constant π_λ , so that λ is a linear function of T (we scaled things so that $\lambda = T$). One can think of (q, p) as mimicking the scale factors, while (Q, P) mimics the off-diagonal components or the p -form variables. In (A.1), there is only one potential wall for q (namely, the second term). We shall consider later the case with several walls.

Let us start by remarking the important fact that the variable $f(Q, P)$ has zero Poisson bracket with H , therefore it is a constant of the motion. To simplify the following argument, we assume that we perform a canonical transformation such that $P^{\text{new}} = f(Q, P)$. Dropping henceforth the label ‘new’, we obtain $P = P_0$ where P_0 is a constant which we assume to be different from zero. The basic aim of this appendix is then to show that the (new) conjugate variable Q will also tend to a constant as $T \rightarrow +\infty$.

In the limit of large times, the motion in q is a free motion interrupted by a collision against the potential wall,

$$q = p_0|T - T_0| + q_0 \quad (\text{A.2})$$

where T_0 is the time of the collision, q_0 the turning point, and p_0 the constant momentum of q far from the wall. The location of the turning point is determined by energy conservation (using the fact that $p(T_0)$ vanishes before changing sign):

$$P_0^2 \rho_0^k e^{-\rho_0 q_0} = p_0^2. \quad (\text{A.3})$$

The time scale ΔT of the collision (during which one feels the influence of the exponential potential) is roughly of the order $1/(\rho_0 p_0)$: the later the collision, the sharper the wall. Let us evaluate the change in Q in the collision. To that end, we need to integrate $\dot{Q} = P \rho^k e^{-\rho q}$ over the collision, which yields

$$\Delta Q = P_0 \int_{-\infty}^{\infty} dT \rho^k e^{-\rho(p_0|T - T_0| + q_0)}. \quad (\text{A.4})$$

The integrand is maximum at $T = T_0$. We can approximate the integral by the value at the maximum times the time scale of the collision. Using (A.3), one obtains

$$\Delta Q \approx \frac{P_0}{P_0 \rho_0} = \frac{P_0}{P_0} e^{-T_0}. \quad (\text{A.5})$$

Hence, the variable Q receives a kick during the collision (which can be of order 1 at early times), but the later the collision (i.e., the larger T_0), the smaller the kick.

Assume now that there is another wall with the same pre-factor and the same time dependence, say at $q = d$, so that q bounces between these two walls,

$$V_{\text{additional}} = \frac{1}{2} P^2 \rho^k e^{-\rho(d-q)}.$$

At each collision Q receives a kick of order $1/\rho_0 = e^{-T_0}$. Because the speed of q remains constant (in the large T limit), the collisions are equally spaced in T . Therefore, the time of

the n th collision grows linearly with n : roughly $T_0^{(n)} \sim nd/p_0$. The total change in Q is then obtained by summing all the individual changes, which yields

$$(\Delta Q)_{\text{Total}} \sim \sum_n e^{-nd/p_0}. \quad (\text{A.6})$$

This sum converges. Therefore, after a while, one can neglect the further change in Q , i.e., assume $\dot{Q} = 0$. The Hamiltonian describing the large time limit is obtained by taking the sharp wall limit in the above H , and reads therefore

$$H = \frac{1}{2}p^2 + \Theta(-q) + \Theta(q - d). \quad (\text{A.7})$$

The pair (Q, P) drops out because it is asymptotically frozen. Our analysis justifies taking the sharp wall limit directly in H for this system, which is the procedure we followed in the text to obtain the gravitational billiards.

Note that, had we studied a more complicated model with different pre-factors for the different walls, the pre-factors $f_A(Q, P)$ would no longer have been exactly conserved. However, their non-conservation would only have been driven by the ‘far-away’ walls, so that their time variation would have been exponentially small.

References

- [1] Anderson E 2002 Strong-coupled relativity without relativity *Preprint* gr-qc/0205118
- [2] Andersson L and Rendall A D 2001 Quiescent cosmological singularities *Commun. Math. Phys.* **218** 479–511 (*Preprint* gr-qc/0001047)
- [3] Banks T, Fischler W and Motl L 1999 Dualities versus singularities *J. High Energy Phys.* JHEP01(1999)019 (*Preprint* hep-th/9811194)
- [4] Belinskii V A 1992 *JETP Lett.* **56** 422
- [5] Belinskii V A and Khalatnikov I M 1973 Effect of scalar and vector fields on the nature of the cosmological singularity *Sov. Phys.–JETP* **36** 591–7
- [6] Belinskii V A, Khalatnikov I M and Lifshitz E M 1970 Oscillatory approach to a singular point in the relativistic cosmology *Adv. Phys.* **19** 525
- [7] Belinskii V A, Khalatnikov I M and Lifshitz E M 1972 Construction of a general cosmological solution of the Einstein equation with a time singularity *Sov. Phys.–JETP* **35** 838–41
- [8] Belinskii V A, Khalatnikov I M and Lifshitz E M 1982 A general solution of the Einstein equations with a time singularity *Adv. Phys.* **31** 639
- [9] Berger B K 2002 Numerical approaches to spacetime singularities *Liv. Rev. Rel.* **5** 1 (*Preprint* gr-qc/0201056)
- [10] Berger B K, Garfinkle D, Isenberg J, Moncrief V and Weaver M 1998 The singularity in generic gravitational collapse is spacelike, local, and oscillatory *Mod. Phys. Lett. A* **13** 1565 (*Preprint* gr-qc/9805063)
- [11] Bergshoeff E, de Roo M, de Wit B and van Nieuwenhuizen P 1982 Ten-dimensional Maxwell–Einstein supergravity, its currents, and the issue of its auxiliary fields *Nucl. Phys. B* **195** 97
- [12] Breitenlohner P and Maison D 1986 On the Geroch group *Ann. Inst. H Poincaré* **46** 215
- [13] Breitenlohner P, Maison D and Gibbons G W 1988 Four-dimensional black holes from Kaluza–Klein theories *Commun. Math. Phys.* **120** 295
- [14] Buonanno A, Damour T and Veneziano G 1999 Pre-big bang bubbles from the gravitational instability of generic string vacua *Nucl. Phys. B* **543** 275 (*Preprint* hep-th/9806230)
- [15] Chapline G F and Manton N S 1983 Unification of Yang–Mills theory and supergravity in ten-dimensions *Phys. Lett. B* **120** 105
- [16] Chernoff D F and Barrow J D 1983 *Phys. Rev. Lett.* **50** 134
- [17] Chitre D M 1972 *PhD Thesis* University of Maryland
- [18] Cornish N J and Levin J J 1997 The mixmaster universe is chaotic *Phys. Rev. Lett.* **78** 998 (*Preprint* gr-qc/9605029)
Cornish N J and Levin J J 1997 The mixmaster universe: a chaotic Farey tale *Phys. Rev. D* **55** 7489 (*Preprint* gr-qc/9612066)
- [19] Cremmer E, Julia B, Lu H and Pope C N 1998 Dualisation of dualities: I *Nucl. Phys. B* **523** 73 (*Preprint* hep-th/9710119)
- [20] Cremmer E, Julia B, Lu H and Pope C N 1998 Dualisation of dualities: II. Twisted self-duality of doubled fields and superdualities *Nucl. Phys. B* **535** 242 (*Preprint* hep-th/9806106)

- [21] Cremmer E, Julia B, Lu H and Pope C N 1999 Higher-dimensional origin of $D = 3$ coset symmetries *Preprint* hep-th/9909099
- [22] Cremmer E, Julia B and Scherk J 1978 Supergravity theory in 11 dimensions *Phys. Lett. B* **76** 409
- [23] Damour T, de Buyl S, Henneaux M and Schomblond C 2002 Einstein billiards and overextensions of finite-dimensional simple Lie algebras *J. High Energy Phys.* JHEP08(2002)030 (*Preprint* hep-th/0206125)
- [24] Damour T and Henneaux M 2000 Chaos in superstring cosmology *Phys. Rev. Lett.* **85** 920 (*Preprint* hep-th/0003139)
See also a short version in Damour T and Henneaux M 2000 *Gen. Rel. Grav.* **32** 2339
- [25] Damour T and Henneaux M 2000 Oscillatory behaviour in homogeneous string cosmology models *Phys. Lett. B* **488** 108 (*Preprint* hep-th/0006171)
- [26] Damour T and Henneaux M 2001 $E(10)$, $BE(10)$ and arithmetical chaos in superstring cosmology *Phys. Rev. Lett.* **86** 4749 (*Preprint* hep-th/0012172)
- [27] Damour T, Henneaux M, Julia B and Nicolai H 2001 Hyperbolic Kac–Moody algebras and chaos in Kaluza–Klein models *Phys. Lett. B* **509** 323 (*Preprint* hep-th/0103094)
- [28] Damour T, Henneaux M and Nicolai H 2002 E_{10} and a small tension expansion of M theory *Phys. Rev. Lett.* **89** 221601 (*Preprint* hep-th/0207267)
- [29] Damour T, Henneaux M, Rendall A D and Weaver M 2002 Kasner-like behaviour for subcritical Einstein–matter systems *Ann. Inst. H Poincaré* **3** 1049 (*Preprint* gr-qc/0202069)
- [30] Dautcourt G 1998 On the ultrarelativistic limit of general relativity *Acta Phys. Polon. B* **29** 1047 (*Preprint* gr-qc/9801093)
- [31] Demaret J, De Rop Y and Henneaux M 1988 Chaos in nondiagonal spatially homogeneous cosmological models in space-time dimensions > 10 *Phys. Lett. B* **211** 37
- [32] Demaret J, Hanquin J L, Henneaux M, Spindel P and Taormina A 1986 The fate of the mixmaster behavior in vacuum inhomogeneous Kaluza–Klein cosmological models *Phys. Lett. B* **175** 129
- [33] Demaret J, Henneaux M and Spindel P 1985 Nonoscillatory behavior in vacuum Kaluza–Klein cosmologies *Phys. Lett. B* **164** 27
- [34] DeWitt B S 1967 Quantum theory of gravity: 1. The canonical theory *Phys. Rev.* **160** 1113
- [35] Eardley D, Liang E and Sachs R 1972 Velocity-dominated singularities in irrotational dust cosmologies *J. Math. Phys.* **13** 99
- [36] Eskin A and McMullen C 1993 Mixing, counting and equidistribution in Lie groups *Duke Math. J.* **71** 181–209
- [37] Ehlers J 1957 *Dissertation* Hamburg University
Ehlers J 1962 Transformation of static exterior solutions of Einstein’s gravitational field equations into different solutions by means of conformal mappings *Les Theories relativistes de la gravitation, Colloques Internationaux du CNRS* **91** 275
- [38] Feingold A J and Frenkel I B 1983 A hyperbolic Kac–Moody algebra and the theory of Siegel modular form of genus 2 *Math. Ann.* **263** 87
- [39] Fischbacher T and Nicolai H 2003 Low level representations for E_{10} and E_{11} *Preprint* hep-th/0301017
- [40] Gasperini M and Veneziano G 1993 Pre-big bang in string cosmology *Astropart. Phys.* **1** 317 (*Preprint* hep-th/9211021)
- [41] Gasperini M and Veneziano G 2002 The pre-big bang scenario in string cosmology *Preprint* hep-th/0207130
- [42] Gebert R W, Inami T and Mizoguchi S 1996 Toda field theories associated with hyperbolic Kac–Moody algebras—Painlevé properties and W algebras *Int. J. Mod. Phys. A* **11** 5479
- [43] Gebert R W and Nicolai H 1994 E_{10} for beginners *Strings and Symmetries, Proc. Gürsey Memorial Conf. I (Lecture Notes in Physics)* (Berlin: Springer)
- [44] Geroch R 1971 A method for generating solutions of Einstein’s equations *J. Math. Phys.* **12** 918
Geroch R 1972 A method for generating solutions of Einstein’s equations 2 *J. Math. Phys.* **13** 394
- [45] Gibbons G W 1985 Aspects of supergravity theories, Three lectures given at *GIFT Seminar on Theoretical Physics (San Feliu de Guixols, Spain, June 4–11, 1984)* (Singapore: World Scientific)
- [46] Gibbons G, Hashimoto K and Yi P 2002 Tachyon condensates, Carrollian contraction of Lorentz group, and fundamental strings *J. High Energy Phys.* JHEP09(2002)061 (*Preprint* hep-th/0209034)
- [47] Goddard P and Olive D I 1986 Kac–Moody and Virasoro algebras in relation to quantum physics *Int. J. Mod. Phys. A* **1** 303
- [48] Hawking S W and Penrose R 1970 The singularities of gravitational collapse and cosmology *Proc. R. Soc. A* **314** 529
- [49] Helgason S 2001 *Differential Geometry, Lie Groups, and Symmetric Spaces (Graduate Studies in Mathematics vol 34)* (Providence, RI: American Mathematical Society)
- [50] Henneaux M 1979 Geometry of zero signature space-times *Bull. Soc. Math. Belg.* **31** 47 (note the misprints in the published version, absent in the preprint version)

- [51] Henneaux M, Pilati M and Teitelboim C 1982 Explicit solution for the zero signature (strong coupling) limit of the propagation amplitude in quantum gravity *Phys. Lett. B* **110** 123
- [52] Howe R E and Moore C C 1979 Asymptotic properties of unitary representations *J. Funct. Anal.* **32** 72–96
- [53] Humphreys J E 1980 *Introduction to Lie Algebras and Representation Theory (Graduate Texts in Mathematics vol 9)* (Berlin: Springer)
- [54] Imponente G and Montani G 2001 On the covariance of the mixmaster chaoticity *Phys. Rev. D* **63** 103501 (Preprint astro-ph/0102067)
- [55] Isham C J 1976 Some quantum field theory aspects of the superspace quantization of general relativity *Proc. R. Soc. A* **351** 209
- [56] Ivashchuk V D, Kirillov A A and Melnikov V N 1994 Stochastic properties of multidimensional cosmological models near a singular point *JETP Lett.* **60** 235
Ivashchuk V D, Kirillov A A and Melnikov V N 1994 *Pis. Zh. Eksp. Teor. Fiz.* **60** 225 (in Russian)
- [57] Ivashchuk V D and Melnikov V N 1995 Billiard representation for multidimensional cosmology with multicomponent perfect fluid near the singularity *Class. Quantum Grav.* **12** 809
- [58] Ivashchuk V D and Melnikov V N 2000 Billiard representation for multidimensional cosmology with intersecting p-branes near the singularity *J. Math. Phys.* **41** 6341 (Preprint hep-th/9904077)
- [59] Ivashchuk V D and Melnikov V N 2001 Exact solutions in multidimensional gravity with antisymmetric forms *Class. Quantum Grav.* **18** R87
- [60] Jantzen R T 1986 Finite-dimensional Einstein–Maxwell–scalar-field system *Phys. Rev. D* **33** 2121–35
- [61] Jantzen R T 2001 Spatially homogeneous dynamics: a unified picture *Preprint gr-qc/0102035*
- [62] Julia B 1980 Group disintegrations *Superspace and Supergravity* ed S W Hawking and M Rocek (Cambridge: Cambridge University Press) p 331
- [63] Julia B 1984 Infinite Lie algebras in physics *Proc. Johns Hopkins Workshop on Current Problems in Particle Physics; Unified Theories and Beyond* (Baltimore: Johns Hopkins University) p 23
- [64] Julia B 1985 Kac–Moody symmetry of gravitation and supergravity *Lectures in Applied Mathematics* vol 21 (Providence, RI: American Mathematical Society) p 355
- [65] Julia B and Nicolai H 1996 Conformal internal symmetry of 2d σ -models coupled to gravity and a dilaton *Nucl. Phys.* **482** 431
- [66] Kac V G 1990 *Infinite Dimensional Lie Algebras* 3rd edn (Cambridge: Cambridge University Press)
- [67] Peterson D and Kac V G 1983 *Proc. Natl. Acad. Sci. USA* **80** 1778
- [68] Kang S-J 1994 Root multiplicities of the hyperbolic Kac–Moody algebra $HA_1^{(1)}$ *J. Algebra* **160** 623
- [69] Khalatnikov I M, Lifshitz E M, Khanin K M, Shchur L N and Sinai Ya G 1985 On the stochasticity in relativistic cosmology *J. Stat. Phys.* **38** 97
- [70] Khvedelidze A M and Mladenov D M 2002 Bianchi I cosmology and Euler–Calogero–Sutherland model *Preprint gr-qc/0208037*
- [71] Kirillov A A 1993 *Sov. Phys.–JETP* **76** 355
- [72] Kirillov A A and Kochnev A A 1987 *JETP Lett.* **46** 436
- [73] Kirillov A A and Melnikov V N 1995 Dynamics of inhomogeneities of metric in the vicinity of a singularity in multidimensional cosmology *Phys. Rev. D* **52** 723 (Preprint gr-qc/9408004)
- [74] Kirillov A A and Serebryakov G V 2001 Origin of a classical space in quantum cosmologies *Grav. Cosmol.* **7** 211 (Preprint hep-th/0012245)
- [75] LeBlanc V G 1997 Asymptotic states of magnetic Bianchi I cosmologies *Class. Quantum Grav.* **14** 2281–301
LeBlanc V G 1998 Bianchi II magnetic cosmologies *Class. Quantum Grav.* **15** 1607–26
- [76] Lidsey J E, Wands D and Copeland E J 2000 Superstring cosmology *Phys. Rep.* **337** 343 (Preprint hep-th/9909061)
- [77] Lifshitz E M, Lifshitz I M and Khalatnikov I M 1971 *Sov. Phys.–JETP* **32** 173
- [78] Lindstrom U and Svendsen H G 2001 A pedestrian approach to high energy limits of branes and other gravitational systems *Int. J. Mod. Phys. A* **16** 1347 (Preprint hep-th/0007101)
- [79] Lu H, Pope C N and Stelle K S 1996 Weyl group invariance and p-brane multiplets *Nucl. Phys. B* **476** 89 (Preprint hep-th/9602140)
- [80] Maldacena J M and Nunez C 2001 Supergravity description of field theories on curved manifolds and a no go theorem *Int. J. Mod. Phys. A* **16** 822 (Preprint hep-th/0007018)
- [81] Margulis G A 1969 Applications of ergodic theory to the investigation of manifolds of negative curvature *Funct. Anal. Appl.* **4** 335
- [82] Matzner R A and Misner C W 1967 *Phys. Rev.* **154** 1229
- [83] Misner C W 1969 Mixmaster universe *Phys. Rev. Lett.* **22** 1071–4
- [84] Misner C W 1969 Quantum cosmology: 1 *Phys. Rev.* **186** 1319
Misner C W 1972 *Minisuperspace Magic Without Magic* ed J R Klauder (San Francisco: Freeman) pp 441–73

- [85] Misner C W 1994 *Deterministic Chaos in General Relativity* ed D Hobill *et al* (New York: Plenum) pp 317–28 (Preprint gr-qc/9405068)
- [86] Moody R V and Pianzola A 1995 *Lie Algebras with Triangular Decomposition* (New York: Wiley)
- [87] Nicolai H 1992 A hyperbolic Lie algebra from supergravity *Phys. Lett. B* **276** 333
- [88] Nicolai H 1991 *Recent Aspects of Quantum Fields, Proc. Schladming 1991 (Lecture Notes in Physics)* (Berlin: Springer)
- [89] Obers N A, Pioline B and Rabinovici E 1998 M-theory and U-duality on T**d with gauge backgrounds *Nucl. Phys. B* **525** 163 (Preprint hep-th/9712084)
- [90] Olive D I and Turok N 1986 The Toda lattice field theory hierarchy and zero curvature conditions in Kac–Moody algebras *Nucl. Phys. B* **265** 469
- [91] Pioline B and Waldron A 2002 Quantum cosmology and conformal invariance *Phys. Rev. Lett.* **90** 031302 (Preprint hep-th/0209044)
- [92] Ringström H 2001 The Bianchi IX attractor *Ann. H Poincare* **2** 405–500 (Preprint gr-qc/0006035)
- [93] Ryan M P 1972 The oscillatory regime near the singularity in Bianchi-type IX universes *Ann. Phys., NY* **70** 301
- [94] Ryan M P 1972 *Hamiltonian Cosmology* (Heidelberg: Springer)
- [95] Ryan M P and Shepley L C 1975 *Homogeneous Relativistic Cosmologies (Princeton Series in Physics)* (Princeton, NJ: Princeton University Press)
- [96] Schnakenburg I and West P C 2001 Kac–Moody symmetries of IIB supergravity *Phys. Lett. B* **517** 421 (Preprint hep-th/0107181)
- [97] Schnakenburg I and West P C 2002 Massive IIA supergravity as a non-linear realisation *Phys. Lett. B* **540** 137 (Preprint hep-th/0204207)
- [98] Teitelboim C 1980 The Hamiltonian structure of space-time *General Relativity and Gravitation* vol 1 ed A Held (New York: Plenum Press)
- [99] Wands D 2002 String-inspired cosmology *Class. Quantum Grav.* **19** 3403 (Preprint hep-th/0203107)
- [100] Weaver M 2000 Dynamics of magnetic Bianchi VI₀ cosmologies *Class. Quantum Grav.* **17** 421–34 (Preprint gr-qc/9909043)
- [101] West P C 2001 E₁₁ and M theory *Class. Quantum Grav.* **18** 4443–60 (Preprint hep-th/0104081)
- [102] Zimmer R 1984 *Ergodic Theory and Semisimple Groups* (Boston: Birkhauser)