Institute of Physics Publishing Classical and Quantum Gravity

Class. Quantum Grav. **23** (2006) 3463–3489 doi:10.1088/0264-9381/23/10/016

Dynamics of the spatially homogeneous Bianchi type I Einstein–Vlasov equations

J Mark Heinzle[1,3] **and Claes Uggla**[2]

[1] Max Planck Institute for Gravitational Physics, Albert Einstein Institute, Am Mühlenberg 1, D-14476 Golm, Germany
[2] Department of Physics, University of Karlstad, S-651 88 Karlstad, Sweden

E-mail: Mark.Heinzle@aei.mpg.de and claes.uggla@kau.se

Received 5 December 2005
Published 28 April 2006
Online at stacks.iop.org/CQG/23/3463

Abstract
We investigate the dynamics of spatially homogeneous solutions of the Einstein–Vlasov equations with Bianchi type I symmetry by introducing a new formulation that allows an efficient use of dynamical systems methods. We find that all models are forever expanding and that they isotropize towards the future; towards the past there exists a singularity—we identify and describe all possible past asymptotic states. In this context, we establish the existence of a heteroclinic network, which is a new type of feature in general relativity. The past asymptotic structure illustrates that the dynamics of Vlasov matter models differs significantly from that of perfect fluid models.

PACS numbers: 04.20.Dw, 98.80.Bp

1. Introduction

In general relativity and cosmology, our knowledge about spatially homogeneous cosmological models has increased substantially over the years, and we are able to say that, for a large number of models, the qualitative behaviour of solutions is now well understood, see [1] for an overview. The majority of results, however, concerns solutions of the Einstein equations coupled to a perfect fluid, usually with a linear equation of state. It is thus important to note that these results are in general not robust, i.e., not structurally stable, under a change of the matter model; significant changes of the qualitative behaviour of solutions occur, for instance, for collisionless matter.

Several fundamental results on spatially homogeneous diagonal models of Bianchi type I with collisionless matter have been obtained in [2]. Diagonal locally rotationally symmetric (LRS) models have been investigated successfully by using dynamical systems methods,

[3] Current address: Institute for Theoretical Physics, Boltzmanngasse 5, A-1090 Vienna, Austria.

see [3] for the case of massless particles and [4, 5] for the massive case. In particular, solutions have been found whose qualitative behaviour is different from that of any perfect fluid model of the same Bianchi type.

The purpose of this paper is to re-investigate and give a detailed description of diagonal (non-LRS) Bianchi type I models with collisionless matter. Our analysis is based on a new formulation of Einstein's field equations that makes an efficient use of dynamical systems techniques possible. This formulation enables us to obtain a much more detailed picture of the global dynamics than the one previously given in [2]; in particular, we are able to determine the possible dynamical behaviour towards the past. Our analysis also reveals a completely new dynamical feature in general relativity—a heteroclinic network. Our results illustrate that there exist significant differences between collisionless matter models and perfect fluid models.

The outline of the paper is as follows. In section 2, we recast Einstein's field equations for the diagonal Bianchi type I case with collisionless matter to a reduced dimensionless dynamical system on a compact state space. In section 3, we give the fixed points of the reduced coupled system and list and discuss a hierarchy of invariant subsets of the state space, which is associated with a hierarchy of monotone functions. In section 4, we first present the results of the local dynamical systems analysis; subsequently we focus on the global dynamics and give two theorems and a conjecture. The first theorem states that all models isotropize asymptotically towards the future; this has been shown before, cf [2], but we give a new dynamical systems proof in appendix C. The past asymptotic dynamics is much more complicated since there exist several types of possible past asymptotic behaviour. In our second global theorem, we give a detailed description of all possible past asymptotic states. In connection with this theorem, we establish the existence of a so-called heteroclinic network, a completely new feature in general relativity, and hence the most interesting discovery in this paper. The conjecture regards details of how this structure affects the past dynamics. After we have stated the theorems and the conjecture about the past dynamics, we give a fairly non-technical description of their content and implications. This is followed by the proofs and our heuristic arguments for the conjecture; the arguments are based on rather technical methods from global dynamical systems analysis; in particular, we exploit the hierarchy of monotone functions in conjunction with the monotonicity principle. The reader who is mainly interested in our results may therefore skip this part and go directly to section 5, where we conclude with some further remarks and comments about our results and their implications. Appendix A provides a brief introduction to relevant background material from the theory of dynamical systems. In appendix B we investigate which conditions on the collisionless matter ensure compatibility with LRS and FRW symmetry. In appendix D, we discuss the physical interpretation of one of the most important boundaries of our state space formulation.

2. The reflection-symmetric Bianchi type I Einstein–Vlasov system

In a spacetime with Bianchi type I symmetry, the spacetime metric can be written as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + g_{ij}(t)\,\mathrm{d}x^i\,\mathrm{d}x^j, \qquad i, j = 1, 2, 3, \tag{1}$$

where g_{ij} is the induced Riemannian metric on the spatially homogeneous surfaces $t = \text{const}$. Since the metric is constant on $t = \text{const}$, it follows that the Ricci tensor of g_{ij} vanishes. Einstein's equations, in units $c = 1 = G$, decompose into the evolution equations,

$$\partial_t g_{ij} = -2k_{ij}, \qquad \partial_t k^i{}_j = \operatorname{tr} k\, k^i{}_j - 8\pi T^i{}_j + 4\pi\delta^i{}_j\left(T^k_{\ k} - \rho\right) - \Lambda\delta^i{}_j, \tag{2a}$$

and the Hamiltonian and momentum constraints

$$(\operatorname{tr} k)^2 - k^i{}_j k^j{}_i - 16\pi\rho - 2\Lambda = 0, \qquad j_k = 0, \tag{2b}$$

where k_{ij} denotes the second fundamental form of the surfaces $t = \text{const}$; see, e.g., [6]. The matter variables are defined as components of the energy–momentum tensor $T_{\mu\nu}$ ($\mu = 0, 1, 2, 3$), according to $\rho = T_{00}$, $j_k = T_{0k}$; T_{ij} denotes the spatial components. The cosmological constant Λ is set to zero in the following; the treatment of the case $\Lambda > 0$ is straightforward once the case $\Lambda = 0$ has been solved, cf the remarks in the conclusions.

In this paper, we consider collisionless matter (Vlasov matter), i.e., an ensemble of freely moving particles described by a non-negative distribution function f defined on the mass shell $PM \subseteq TM$ of the spacetime; for simplicity, we consider particles with equal mass m. The spacetime coordinates (t, x^i) and the spatial components v^i of the 4-momentum v^μ (measured w.r.t. $\partial/\partial x^\mu$) provide local coordinates on PM, since $v^\mu v_\mu = -m^2$. We thus find that f is a function $f = f(t, x^i, v^j)$. Compatibility with Bianchi type I symmetry forces the distribution function f to be homogeneous, i.e., $f = f(t, v^j)$. The evolution equation for f is the Vlasov equation (the Liouville equation)

$$\partial_t f + \frac{v^j}{v^0}\partial_{x^j} f - \frac{1}{v^0}\Gamma^j_{\mu\nu} v^\mu v^\nu \partial_{v^j} f = \partial_t f + 2k^j{}_l v^l \partial_{v^j} f = 0, \tag{2c}$$

see, e.g., [9] or [10]. The energy–momentum tensor associated with the distribution function f is given by

$$T^{\mu\nu} = \int f v^\mu v^\nu \mathrm{vol}_{PM},$$

where $\mathrm{vol}_{PM} = (\det g)^{1/2} v_0^{-1}\, \mathrm{d}v^1\, \mathrm{d}v^2\, \mathrm{d}v^3$ is the induced volume form on the mass shell; v_0 is understood as a function of the spatial components, i.e., $v_0^2 = m^2 + g_{ij} v^i v^j$. The components ρ, j_k and T_{ij}, which enter in (2a) and (2b), can thus be written as

$$\rho = \int f (m^2 + g^{ij} v_i v_j)^{1/2} (\det g)^{-1/2}\, \mathrm{d}v_1\, \mathrm{d}v_2\, \mathrm{d}v_3, \tag{2d}$$

$$j_k = \int f v_k (\det g)^{-1/2}\, \mathrm{d}v_1\, \mathrm{d}v_2\, \mathrm{d}v_3, \tag{2e}$$

$$T_{ij} = \int f v_i v_j (m^2 + g^{kl} v_k v_l)^{-1/2} (\det g)^{-1/2}\, \mathrm{d}v_1\, \mathrm{d}v_2\, \mathrm{d}v_3. \tag{2f}$$

The Einstein–Vlasov system (2) is usually considered for particles of mass $m > 0$, however, the system also describes massless particles if we set $m = 0$. (For a detailed introduction to the Einstein–Vlasov system we refer to [9] and [10].)

The general spatially homogeneous solution of the Vlasov equation (2c) in Bianchi type I is

$$f(t, v^i) = f_0(v_i), \tag{3}$$

where v_i are the covariant components of the momenta and f_0 is an arbitrary non-negative function, see [11]. (By inserting (3) into (2c) and using that $v_i = g_{ij}(t) v^j$ it is easy to check that $f_0(v_i)$ is a solution.) The momentum constraint in (2b) then reads

$$\int f_0(v_i) v_k\, \mathrm{d}v_1\, \mathrm{d}v_2\, \mathrm{d}v_3 = 0. \tag{4}$$

Henceforth, for simplicity, f_0 is assumed to be compactly supported, which ensures finiteness in (2d)–(2f).

There exists a subclass of Bianchi type I Einstein–Vlasov models that is naturally associated with the constraint (4): the class of 'reflection-symmetric' (or 'diagonal') models.

The following symmetry conditions are imposed on the initial data:

$$f_0(v_1, v_2, v_3) = f_0(-v_1, -v_2, v_3) = f_0(-v_1, v_2, -v_3) = f_0(v_1, -v_2, -v_3), \tag{5a}$$

$$g_{ij}(t_0), k_{ij}(t_0) \quad \text{diagonal.} \tag{5b}$$

These conditions automatically ensure that $j_k = 0$; furthermore, $T_{ij}(t_0)$ is diagonal, cf (2*f*); hence g_{ij}, k_{ij} and T_{ij} are diagonal for all times by the evolution equations. In the present paper, we will be concerned with this class of reflection-symmetric models.

The Einstein–Vlasov system (2) thus reduces to a system for six unknowns, the diagonal components of the metric $g_{ii}(t)$ and the second fundamental form $k^i{}_i(t)$ (no summation). The equations are (2*a*) and the Hamiltonian constraint in (2*b*). The initial data consist of $g_{ii}(t_0)$, $k^i{}_i(t_0)$; in addition we prescribe a distribution function $f_0(v_i)$ that provides the source terms in the equations via (2*d*) and (2*f*).

In the following, we reformulate the Einstein–Vlasov system as a dimensionless system on a compact state space. To that end we introduce new variables and modified matter quantities. Let

$$H := -\frac{\operatorname{tr} k}{3}, \qquad x := g^{11} + g^{22} + g^{33}, \tag{6}$$

and define the dimensionless variables

$$s_i := \frac{g^{ii}}{x}, \qquad \Sigma_i := -\frac{k^i{}_i}{H} - 1, \qquad z := \frac{m^2}{m^2 + x}, \tag{7a}$$

where

$$s_1 + s_2 + s_3 = 1, \qquad \Sigma_1 + \Sigma_2 + \Sigma_3 = 0. \tag{7b}$$

The transformation from the variables $\left(g_{ii}, k^i{}_i\right)$ to (s_i, Σ_i, x, H), where (s_i, Σ_i) are subject to the above constraints, is one-to-one. (Note that x can be obtained from z when $m > 0$.) By distinguishing one direction (1, 2 or 3), one can decompose s_i and simultaneously introduce a trace-free adaption of the shear to new $\Sigma_\pm$ variables as is done in, e.g., [1]; however, since Bianchi type I does not have a preferred direction we will refrain from doing so here.

Next, we replace the matter quantities ρ, T^i_i (no summation) by the dimensionless quantities

$$\Omega := \frac{8\pi\rho}{3H^2}, \qquad w_i := \frac{T^i_i}{\rho}, \qquad w := \frac{1}{3}\sum_i w_i = \frac{1}{3}\frac{\sum_i T^i_i}{\rho}. \tag{8}$$

Expressed in the new variables, w_i can be written as

$$w_i = \frac{(1-z)s_i \int f_0 v_i^2 \left[z + (1-z)\sum_k s_k v_k^2\right]^{-1/2} \mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3}{\int f_0 \left[z + (1-z)\sum_k s_k v_k^2\right]^{1/2} \mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3}. \tag{9}$$

Finally, let us introduce a new dimensionless time variable τ defined by

$$\partial_\tau = H^{-1}\partial_t, \tag{10}$$

henceforth a prime denotes differentiation w.r.t. τ.

We now rewrite the Einstein–Vlasov equations as a set of dimensional equations that decouple for dimensional reasons and a reduced system of dimensionless coupled equations on a compact state space. The decoupled dimensional equations are

$$H' = -3H\left[1 - \frac{\Omega}{2}(1 - w)\right] \tag{11a}$$

$$x' = -2x\left(1 + \sum_k s_k \Sigma_k\right). \tag{11b}$$

The reduced dimensionless system consists of the Hamiltonian constraint, cf (2*b*),

$$1 - \Sigma^2 - \Omega = 0, \qquad \text{where} \quad \Sigma^2 := \tfrac{1}{6}\left(\Sigma_1^2 + \Sigma_2^2 + \Sigma_3^2\right), \tag{12}$$

and a coupled system of evolution equations

$$\Sigma_i' = -3\Omega\left[\frac{1}{2}(1-w)\Sigma_i - (w_i - w)\right] \tag{13a}$$

$$s_i' = -2s_i\left[\Sigma_i - \sum_k s_k \Sigma_k\right] \tag{13b}$$

$$z' = 2z(1-z)\left(1 + \sum_k s_k \Sigma_k\right). \tag{13c}$$

In the system (13), Ω is regarded as $\Omega = 1 - \Sigma$ because of (12), and $w = w(z, s_j)$, $w_i = w_i(z, s_j)$ from (8) and (9). Note that the functional dependence of w and w_i on z and (s_1, s_2, s_3) involves the (arbitrary) distribution function f_0 through an integration; this makes the system (13) non-standard and numerical investigations non-trivial.

In the massive case $m > 0$, the decoupled equation for x is redundant since the equation for z is equivalent. In the massless case $m = 0$ we have $z = 0$; hence, x is needed in order to reconstruct the spatial metric from the new variables, although the equation for x does not contribute to the dynamics.

The dimensionless dynamical system (13) together with the constraint (12) describes the full dynamics of the Einstein–Vlasov system of Bianchi type I. In the massive case, the state space associated with this system is the space of the variables $\{(\Sigma_i, s_i, z)\}$, i.e.,

$$\mathcal{X} := \{(\Sigma_i, s_i, z) \mid (\Sigma^2 < 1) \wedge (s_i > 0) \wedge (0 < z < 1)\}, \tag{14}$$

where s_i and Σ_i are subject to the constraints $s_1 + s_2 + s_3 = 1$, $\Sigma_1 + \Sigma_2 + \Sigma_3 = 0$, cf (7*b*). (The inequalities for s_i and Σ_i follow from the definition (7*a*) and the constraint (12), respectively.) The state space $\mathcal{X}$ is thus five dimensional.

It will turn out eventually that all solutions asymptotically approach the boundaries of $\mathcal{X}$: $z = 0, z = 1, s_i = 0, \Omega = 0$ ($\Leftrightarrow \Sigma^2 = 1$). This suggests the inclusion of these sets in the analysis, whereby we obtain a compact state space $\bar{\mathcal{X}}$.

The equations on the invariant subset $z = 0$ of $\bar{\mathcal{X}}$ are identical to the coupled dimensionless system in the case of massless particles $m = 0$. We will therefore refer to the subset $z = 0$ as the massless subset; it represents the four-dimensional state space for the massless case.

We conclude this section by looking at some variables in more detail. The inequality $\Sigma^2 \leqslant 1$ together with the constraint $\Sigma_1 + \Sigma_2 + \Sigma_3 = 0$ results in $|\Sigma_i| \leqslant 2$ for all i. Note that equality is achieved when $(\Sigma_1, \Sigma_2, \Sigma_3) = (\pm 2, \mp 1, \mp 1)$ and permutations thereof, cf figure 1. The matter quantities satisfy

$$0 \leqslant w \leqslant \tfrac{1}{3}, \qquad 0 \leqslant w_i \leqslant 3w \leqslant 1. \tag{15}$$

The equalities hold at the boundaries of the state space: $w_i = 0 = w_i$ iff $z = 1$; $w = \frac{1}{3}$ iff $z = 0$; $w_i = 0$ iff $s_i = 0$; $w_i = 3w$ when $z < 1$ iff $s_i = 1$.

There exist a number of useful auxiliary equations that complement the system (13):

$$\Omega' = \Omega\left[3(1-w)\Sigma^2 - \sum_k w_k \Sigma_k\right], \tag{16}$$

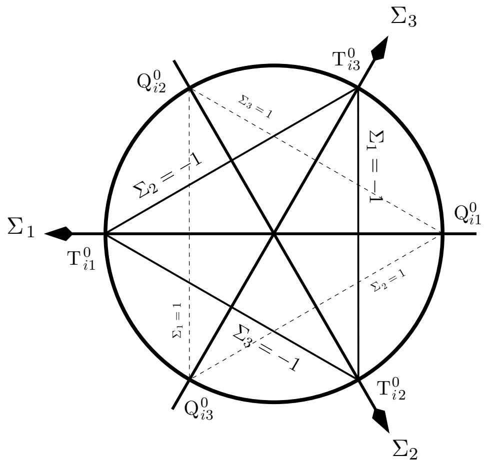

Figure 1. The disc $\Sigma^2 \leqslant 1$ and the Kasner circle KC^0_i.

Table 1. The fixed point sets. The range of the index i is always $i = 1, \ldots, 3$. The superscript denotes the value of z; the first kernel letter describes the type of fixed point set; if there is no second kernel letter, the fixed point set is just a point; if there is a second kernel letter, this letter denotes the dimensionality and character of the set—S refers to surface, L stands for line and C for circle.

Fixed point set	Defined by	Interpretation
FS^1	$z = 1, \Sigma_j = 0\, \forall j$	FRW dust
KC^1_i	$z = 1, \Sigma^2 = 1, s_i = 1, s_j = 0\, \forall j \neq i$	Kasner
TS_i	$0 \leqslant z \leqslant 1, \Sigma_i = 2, \Sigma_j = -1\, \forall j \neq i, s_i = 0$	Taub
KC^0_i	$z = 0, \Sigma^2 = 1, s_i = 1, s_j = 0\, \forall j \neq i$	Kasner
QL^0_i	$z = 0, \Sigma_i = -2, \Sigma_j = 1\, \forall j \neq i, s_i = 0$	Non-flat LRS Kasner
F^0	$z = 0, \Sigma_j = 0\, \forall j, w_j = 1/3\, \forall j$	FRW radiation
D^0_i	$z = 0, s_i = 0, \Sigma_i = -1, \Sigma_j = 1/2 = w_j\, \forall j \neq i$	Distributional LRS

$$\rho' = -\rho \left[3(1+w) + \sum_k w_k \Sigma_k \right] \leqslant -2\rho. \tag{17}$$

The inequality in (17) follows by using $\Sigma_i \geqslant -2\, \forall i$ and (15). This shows that ρ increases monotonically towards the past which yields a matter singularity, i.e., $\rho \to \infty$ for $\tau \to -\infty$. It is often beneficial to consider the equations of the original variables as auxiliary equations, e.g., $(g^{ii})' = -2g^{ii}(1 + \Sigma_i)$.

3. Fixed points, invariant subsets and monotone functions

3.1. Fixed points

The dynamical system (13) possesses a number of fixed points, all residing on the boundaries of the state space; see table 1.

- FS^1 is a surface of fixed points that correspond to the flat isotropic Friedmann–Robertson–Walker (FRW) dust solution.

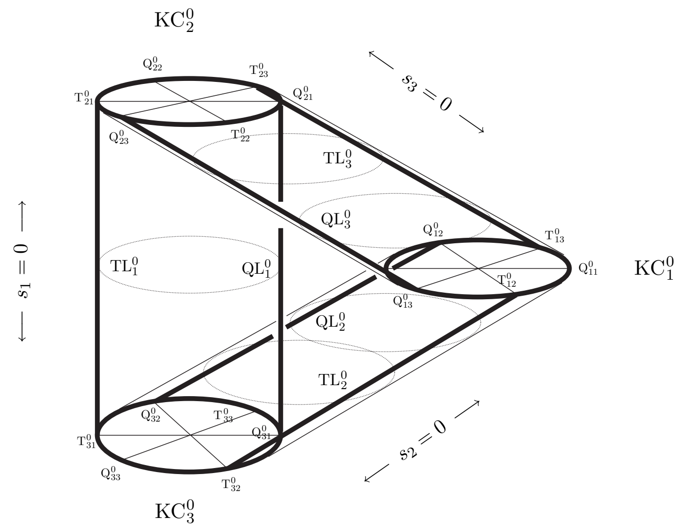

Figure 2. A schematic depiction of the fixed points on $z = 0$. The underlying structure is the three sides of the s_i-triangle $s_1 + s_2 + s_3 = 1$: each point represents a disc $\Sigma^2 \leqslant 1$; the vertices contain the Kasner circles KC_i^0. Bold lines denote the lines of fixed points TL_i^0, QL_i^0 and KC_i^0.

- The circles $\mathrm{KC}_i^{1,0}$ consist of fixed points that correspond to Kasner solutions, see figure 1.
- The fixed points on TS_i are associated with the Taub representation of Minkowski spacetime, see, e.g., [1]. The intersection of TS_i with $(z = 0)$ yields a line of fixed points which we denote by TL_i^0.
- The fixed points on QL_i^0 correspond to the non-flat LRS Kasner solution; note that each fixed point on one line QL_i^0 represents the same LRS Kasner solution.
- F^0 is a fixed point that corresponds to the flat isotropic FRW radiation solution. The location of F^0 depends on the chosen distribution function since the equations $w_1 = w_2 = w_3 = 1/3$, which are to be solved for (s_1, s_2, s_3), involve f_0; see appendix B for details.
- The fixed points D_i^0 are associated with a self-similar LRS solution connected with a distributional f_0; see appendix D for details. Their location depends on f_0, and is determined by $w_j = 1/2$ $(\forall\, j \neq i)$.

The LRS points on KC_i^0 play a particularly important role in the following, which motivates that they are given individual names:

- The three Taub points on KC_i^0 defined by $\Sigma_j = 2$ (and thus $\Sigma_l = -1\, \forall l \neq j$) are denoted by T_{ij}^0.
- The three non-flat LRS points on KC_i^0 given by $\Sigma_j = -2$ (and thus $\Sigma_l = 1\, \forall l \neq j$) are denoted by Q_{ij}^0.

The Kasner circles KC_j^0 and KC_k^0 are connected by the lines TL_i^0 and QL_i^0; the end points of the line TL_i^0 are the Taub points T_{ji}^0 and T_{ki}^0; analogously, the end points of QL_i^0 are the points Q_{ji}^0 and Q_{ki}^0, where (i, j, k) is an arbitrary permutation of $(1, 2, 3)$. The remaining points T_{ll}^0 and Q_{ll}^0 $(l = 1, \ldots, 3)$ do not lie on any of the fixed point sets TL_i^0 or QL_i^0. This fixed point structure is depicted in figure 2.

3.2. Invariant subsets and monotone functions

The dynamical system (13) possesses a hierarchy of invariant subsets and monotone functions. Since this feature of the dynamical system will turn out to be of crucial importance in the analysis of the global dynamics, we give a detailed discussion.

$\mathcal{X}$: on the full (interior) state space $\mathcal{X}$, we define

$$M_{(1)} = (s_1 s_2 s_3)^{-1/3} \frac{z}{1-z}. \tag{18a}$$

A straightforward computation shows

$$M'_{(1)} = 2M_{(1)}, \tag{18b}$$

i.e., $M_{(1)}$ is strictly monotonically increasing along orbits in $\mathcal{X}$. Note that $M_{(1)}$ is intimately related to the spatial volume density since $M_{(1)} = m^2 \det(g_{ij})^{1/3}$.

$\mathcal{Z}^1$: this subset is characterized by $z = 1$. Since $w_i = w = 0$, the equations for s_i decouple, and the essential dynamics is described by the equations $\Sigma'_i = -(3/2)(1-\Sigma^2)\Sigma_i$. (Note that these equations are identical to the Bianchi type I equations for dust—it is therefore natural to refer to $\mathcal{Z}^1$ as the dust subset.) Explicit solutions for these equations can be obtained by noting that $\Sigma_1 \propto \Sigma_2 \propto \Sigma_3$ for all solutions or by using that $\Omega' = 3\Sigma^2\Omega$.

$\mathcal{Z}^0$: this subset is the massless boundary set $z = 0$. Since $w = 1/3$, the dynamical system (13) reduces to

$$\Sigma'_i = -\Omega[1 + \Sigma_i - 3w_i], \qquad s'_i = -2s_i\left[\Sigma_i - \sum_k s_k \Sigma_k\right]. \tag{19}$$

Consider the function

$$M_{(2)} = (1-\Sigma^2)^{-1}(s_1 s_2 s_3)^{-1/6} \int f_0 \left[\sum_k s_k v_k^2\right]^{1/2} \mathrm{d}v_1\, \mathrm{d}v_2\, \mathrm{d}v_3. \tag{20a}$$

The derivative is

$$M'_{(2)} = -2\Sigma^2 M_{(2)}, \tag{20b}$$

which yields monotonicity when $\Sigma^2 \neq 0$. If $\Sigma^2 = 0$, then

$$M'_{(2)} = 0, \qquad M''_{(2)} = 0, \qquad M'''_{(2)} = -6M_{(2)} \sum_i \left(w_i - \frac{1}{3}\right)^2. \tag{20c}$$

Hence, $M_{(2)}$ is strictly monotonically decreasing everywhere on $z = 0$, except at the fixed point F^0 (for which $\Sigma^2 = 0$ and $w_1 = w_2 = w_3 = 1/3$), where $M_{(2)}$ attains a positive minimum. The latter follows from the fact that $(1-\Sigma^2)^{-1}$ is minimal at the point $\Sigma_i = 0\ \forall i$ and that $\partial M_{(2)}/\partial s_i = (2s_i)^{-1}[w_i - 1/3]M_{(2)}$.

$\mathcal{S}_i$ ($i = 1, 2, 3$): these invariant boundary subsets are defined by $s_i = 0$ (which yields $w_i = 0$). There exists a monotone function on $\mathcal{S}_1$,

$$M_{(3)} = (s_2 s_3)^{-1/2} \frac{z}{1-z}, \qquad M'_{(3)} = (2 - \Sigma_1)M_{(3)}, \tag{21}$$

analogous functions can be obtained on $\mathcal{S}_2$ and $\mathcal{S}_3$ through permutations. In appendix D, we show that the sets $\mathcal{S}_i$ are associated with the Einstein–Vlasov equations stemming from distributional distribution functions; hence, we will refer to these subsets as distributional subsets.

$\mathcal{K}$: this boundary subset is the vacuum subset defined by $\Omega = 0$ (or equivalently $\Sigma^2 = 1$). Σ_i are constant on this subset, which completely determines the dynamics of the s_i variables (via (13*b*) or via the auxiliary equation for g^{ii}). The Bianchi type I vacuum solution is the familiar Kasner solution and we thus refer to $\mathcal{K}$ as the Kasner subset.

Intersections of the above boundary subsets yield boundary subsets of lower dimensions; those that are relevant for the global dynamics are discussed in the following.

$\mathcal{S}_i^0$ and $\mathcal{S}_i^1$: the intersection between the subset $\mathcal{S}_i$ and $\mathcal{Z}^0$ and $\mathcal{Z}^1$ yields three-dimensional invariant subsets $(s_i = 0) \cap (z = 0)$ and $(s_i = 0) \cap (z = 1)$, respectively. On $\mathcal{S}_i^0$ there exists a monotonically decreasing function:

$$M_{(4)} = (1 + \Sigma_i)^2, \qquad M'_{(4)} = -2\Omega M_{(4)}. \tag{22}$$

$\mathcal{S}_{ij}$: these subsets are defined by setting $s_i = 0$ and $s_j = 0$ $(j \neq i)$, i.e., $\mathcal{S}_{ij} = \mathcal{S}_i \cap \mathcal{S}_j$. On $\mathcal{S}_{ij}$ we have $s_k = 1$ $(k \neq i, j)$ and $w_k = 3w$, since $w_i = w_j = 0$.

$\mathcal{D}_i^0$: the subsets $\mathcal{S}_i^0$ admit two-dimensional invariant subsets $\mathcal{D}_i^0$ characterized by $(z = 0) \cap (s_i = 0) \cap (\Sigma_i = -1)$. On $\mathcal{D}_1^0$ consider the function

$$M_{(5)} = (2 + \Sigma_2\Sigma_3)^{-1}\,(s_2 s_3)^{-1/4} \int f_0 \left[s_2 v_2^2 + s_3 v_3^2\right]^{1/2} \mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3, \tag{23a}$$

analogous functions can be defined on $\mathcal{D}_2^0$ and $\mathcal{D}_3^0$. Equations (19) imply

$$M'_{(5)} = -\tfrac{1}{12} M_{(5)}[(1 - 2\Sigma_2)^2 + (1 - 2\Sigma_3)^2], \tag{23b}$$

i.e., $M_{(5)}$ is strictly monotonically decreasing unless $\Sigma_2 = 1/2 = \Sigma_3$. In the special case $\Sigma_2 = 1/2 = \Sigma_3$, we obtain

$$M'_{(5)} = 0, \qquad M''_{(5)} = 0, \qquad M'''_{(5)} = -\tfrac{27}{8} M_{(5)}\left[\left(w_2 - \tfrac{1}{2}\right)^2 + \left(w_3 - \tfrac{1}{2}\right)^2\right]. \tag{23c}$$

Hence, $M_{(5)}$ is strictly monotonically decreasing everywhere on $\mathcal{D}_1^0$ except for at the fixed point D_1, for which $\Sigma_2 = \Sigma_3 = w_2 = w_3 = \frac{1}{2}$, cf table 1. The function $M_{(5)}$ possesses a positive minimum at D_1. This is because $(2+\Sigma_2\Sigma_3)^{-1}$ is minimal at the point $\Sigma_2 = \Sigma_3 = 1/2$ and $\partial M_{(5)}/\partial s_i = (2s_i)^{-1}[w_i - 1/2]M_{(5)}$ for $i = 2, 3$.

$\mathcal{K}^0$: the intersection of the Kasner subset $\mathcal{K} = (\Sigma^2 = 1)$ with the $z = 0$ subset yields a three-dimensional subset, $\mathcal{K}^0$. This subset will play a prominent role in the analysis of the past asymptotic behaviour of solution.

The remaining subsets are located in the interior state space and are associated with additional spacetime symmetries. They only exist if f_0 satisfies certain conditions, which are less restrictive than requiring that f_0 shares the spacetime symmetries; thus, e.g., LRS symmetry does not necessarily imply that f_0 is LRS!

LRS_i: a solution of the Einstein–Vlasov equations is locally rotationally symmetric (LRS) if $\Sigma_j = \Sigma_k$ and $w_j = w_k$ for some $j \neq k$; without loss of generality we set $(j, k) = (2, 3)$. Accordingly, we define the subset LRS_1 of $\mathcal{X}$ through the equations $\Sigma_2 = \Sigma_3$, $w_2 = w_3$; $\mathrm{LRS}_{2,3}$ are defined analogously. If the set LRS_1 is invariant under the flow of the dynamical system, LRS initial data remain LRS under the evolution, i.e., the general LRS solution exists. However, invariance of LRS_1 requires that the distribution function f_0 satisfies conditions that ensure compatibility with $\Sigma_2 = \Sigma_3$, $w_2 = w_3$; this yields a class of functions, interestingly enough, that is larger than the proper LRS distribution functions, see appendix B for details. Let us consider a compatible f_0. For an orbit lying on LRS_1, equation (13*b*) entails that $s_2(\tau) \propto s_3(\tau)$ (where the proportionality constant exhibits a dependence on f_0, which enters through the equation $w_2 = w_3$), and hence $g_{22} \propto g_{33}$; by rescaling the coordinates one

Table 2. The key invariant subsets; additional invariant subsets can be formed by further intersections. The range of the indices i, j, k is 1, 2, 3.

Subset	Defined by	Comment
$\mathcal{Z}^1$	$z = 1, w_i = w = 0$	Dust subset
$\mathcal{Z}^0$	$z = 0, w = 1/3$	Massless subset
$\mathcal{S}_i$	$s_i = w_i = 0$	Distributional subsets
$\mathcal{K}$	$\Omega = 0$	Vacuum Kasner subset
$\mathcal{S}_i^1$	$z = 1, s_i = w_1 = w_2 = w_3 = 0$	
$\mathcal{S}_i^0$	$z = s_i = w_i = 0, w_j + w_k = 1\ (i \neq j \neq k)$	
$\mathcal{S}_{ij}$	$s_i = s_j = w_i = w_j = 0, s_k = 1\ (i \neq j \neq k)$	
$\mathcal{D}_i^0$	$z = s_i = w_i = \Sigma_i + 1 = 0, w_j + w_k = 1\ (i \neq j \neq k)$	
$\mathcal{K}^0$	$z = \Omega = 0$	
LRS$_i$	$\Sigma_j = \Sigma_k, w_j = w_k, s_j \propto s_k\ (i \neq j \neq k)$	Requires special f_0
FRW	$\Sigma_1 = \Sigma_2 = \Sigma_3 = 0, s_i \propto s_j \propto s_k\ (i \neq j \neq k)$	Requires special f_0 if $m \neq 0$

can achieve $g_{22} = g_{33}$, i.e., a line element in an explicit LRS form. Hence, the LRS$_i$ subsets comprise the solutions with LRS geometry. For distribution functions f_0 that are not LRS-compatible, the LRS$_i$ subsets are not invariant under the flow of the dynamical system; therefore, in general, solutions with LRS geometry do not exist (except for special solutions).

FRW: FRW models with collisionless matter are described by isotropic solutions of the Einstein–Vlasov equations. A solution is isotropic if $\Sigma_1 = \Sigma_2 = \Sigma_3 = 0$ and $w_1 = w_2 = w_3 = w$ for all times τ. The first condition implies isotropy of the geometry: namely, the equations $\Sigma_1 = \Sigma_2 = \Sigma_3 = 0$ yield $s_i(\tau) = \text{const}$ via (13*b*) (where the constants need not equal 1/3 in general), whereby we obtain a FRW geometry, since the spatial coordinates can be rescaled so that $g_{ij}(t) \propto \delta_{ij}$. The second condition $w_1 = w_2 = w_3 = w$ represents isotropy of the pressures. If this condition is violated, $\Sigma_1 = \Sigma_2 = \Sigma_3 = 0$ is impossible, cf (13*a*), i.e., the assumption of isotropic pressures is necessary for an isotropic geometry. One might think that, consistently, the distribution function f_0 must be isotropic. Remarkably, this is not the case: the massless Einstein–Vlasov equations admit a FRW solution *independently of the prescribed distribution function* f_0—the FRW solution uniquely corresponds to the fixed point F^0. (Note that the position $(s_1, s_2, s_3) = \left(s_1^{\mathrm{F}}, s_2^{\mathrm{F}}, s_3^{\mathrm{F}}\right)$ of F^0 depends on f_0, but the fixed point exists for arbitrary f_0.) This solution can be interpreted as the flat isotropic radiation solution. In contrast, the massive Einstein–Vlasov equations do not admit a FRW solution for an arbitrary f_0: the straight line in $\mathcal{X}$ given by $\Sigma_1 = \Sigma_2 = \Sigma_3 = 0$ and $s_i = s_i^{\mathrm{F}}\ \forall i$ is in general not a solution of equations (13). This is because $s_i = s_i^{\mathrm{F}}\ \forall i$ implies $w_1 = w_2 = w_3$ for $z = 0$, but this is not the case for $z > 0$ in general. A distribution function f_0 such that $s_i = s_i^{\mathrm{F}}$ entails $w_1 = w_2 = w_3$ for all z is called compatible with an isotropic geometry; a simple example is $f_0 = \tilde{f}_0\left(v_1^2 + v_2^2 + v_3^2\right)$, see appendix B for details. In the case of compatibility, the line $\Sigma_1 = \Sigma_2 = \Sigma_3 = 0, s_i = s_i^{\mathrm{F}}$ ($\leftrightarrow w_1 = w_2 = w_3$) is a solution of the dynamical system, the FRW orbit; it can be regarded as the intersection of the three LRS subsets. In general, f_0 is incompatible with a FRW geometry and there is no FRW solution of the massive Einstein–Vlasov equations. However, in section 4.2 we will see that there exists one unique solution that isotropizes towards the past (with an isotropic singularity) and towards the future; it thus possesses FRW asymptotic states.

A brief summary of the invariant subsets discussed in this section is given in table 2.

4. Local and global dynamics

4.1. Local dynamics

Let us consider smooth reflection-symmetric Bianchi type I Vlasov solutions that approach fixed point sets when $\tau \to -\infty$.

Theorem 4.1. *In the massive (massless) case, there exists*

(a) a single orbit that approaches (corresponds to) F^0*,*
(b) three equivalent one-parameter sets of orbits (three single orbits) that approach D_i^0*,* $i = 1, \ldots, 3$*,*
(c) one three-parameter (two-parameter) set of orbits that approaches QL_1^0*;* QL_2^0 *and* QL_3^0 *yield equivalent sets,*
(d) one four-parameter (three-parameter) set of orbits that approaches the part of KC_1^0 *defined by* $1 < \Sigma_1 < 2$*; similarly,* KC_2^0 *and* KC_3^0 *yield equivalent sets.*

Proof. The statements of the theorem follow from the local stability analysis of the fixed point sets F^0, D_i^0, QL_i^0, KC_i^0, when combined with the Hartman–Großman and the reduction theorem, since the fixed points F^0, D_i^0 are hyperbolic and QL_i^0, KC_i^0 are transversally hyperbolic. This requires the dynamical system to be $\mathcal{C}^1$ and this leads to some restrictions on f_0. However, it is possible to obtain an alternative proof that does not require such restrictions. Such a proof can be obtained from the hierarchical structure of invariant sets and monotone functions; we will refrain from making the details explicit here, since our analysis of the global dynamics below contains all essential ingredients implicitly. □

Interpretation of theorem 4.1 (massive case). A three-parameter set of solutions converges to each individual non-LRS Kasner solution as $t \to 0$. (In the state space description three equivalent sets of orbits approach three equivalent transversally stable Kasner arcs that cover all non-LRS Kasner solutions; the equivalence reflects the freedom of permuting the spatial coordinates.) Furthermore, a three-parameter set of solutions approaches the non-flat LRS Kasner solution. Hence, in total, a four-parameter set of solutions asymptotically approaches non-flat Kasner states. There exist special solutions with non-Kasner behaviour towards the singularity: one solution isotropizes towards the singularity, i.e., only one solution has an isotropic singularity; a one-parameter set of solutions approaches a non-Kasner LRS solution of the type (D.6) (three equivalent one-parameter sets of orbits approach three equivalent non-Kasner LRS fixed points D_i^0 associated with this solution). For the latter solutions $\Omega = 3/4$; these solutions cannot be interpreted as perfect fluid solutions since they possess anisotropic pressures.

In the following, we show that the list of theorem 4.1 is almost complete: there exist no other attracting sets towards the singularity with one exception, a heteroclinic network.

4.2. Global dynamics

Theorem 4.2. *All orbits in the interior of the state space* $\mathcal{X}$ *of massive particles (state space* $\mathcal{Z}^0$ *of massless particles) converge to* FS^1 (F^0) *when* $\tau \to +\infty$*.*

A proof of theorem 4.2 has been given in [2]; in appendix C, we present an alternative proof based on dynamical systems techniques.

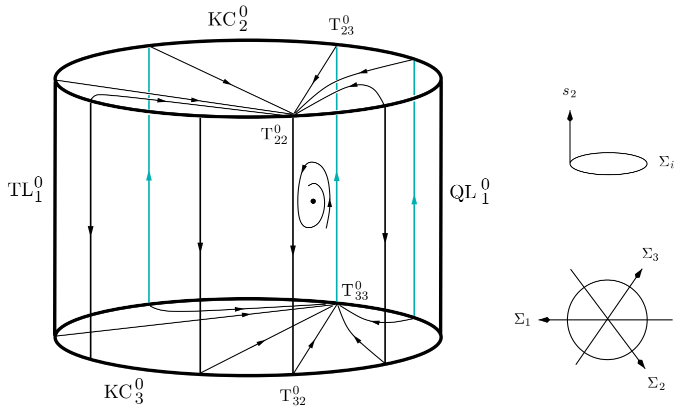

Figure 3. Flow on the boundaries and on the invariant subset $\Sigma_1 = -1$ on $\mathcal{S}^0_1$. The fixed point on $\Sigma_1 = -1$ is the point D^0_1; the heteroclinic cycle $\mathcal{H}^0_1$ consists of the fixed points T^0_{22}, T^0_{32}, T^0_{33}, T^0_{23} and the connecting orbits.

Interpretation of theorem 4.2 (massive case). Since all fixed points on FS^1 correspond to isotropic dust solutions, the theorem states that all smooth reflection-symmetric Bianchi type I Vlasov solutions behave like infinitely diluted isotropized dust solutions towards the future. Combining this result with the first result of theorem 4.1, we find that there exists one unique solution that becomes isotropic towards the past (i.e., possesses an isotropic singularity) and towards the future. Consequently, although for general f_0 there does not exist any isotropic (FRW) solution of the equations, there exists at least one unique solution with isotropic asymptotic states (but the intermediate behaviour is anisotropic). If f_0 is compatible with FRW symmetry, the solution is isotropic for all times. In this case, one can interpret the matter content as a perfect fluid that behaves like a radiation perfect fluid asymptotically towards the past and like dust towards the future. The FRW solution is special—all other solutions of the Einstein–Vlasov equation have anisotropic matter that is impossible to interpret as a perfect fluid. This feature is the reason for the quite different behaviour of Vlasov matter as compared to perfect fluids, both as regards intermediate and asymptotic past behaviour.

The past asymptotic behaviour of solutions is much more complicated. One structure that appears in this context is particularly interesting: a heteroclinic network, $\mathcal{H}^0$. For the definition of a heteroclinic network, we refer the reader to [12] and references therein; here, we restrict ourselves to a discussion of the network $\mathcal{H}^0$. The heteroclinic network $\mathcal{H}^0$ can be regarded as a collection of entangled heteroclinic cycles: $\mathcal{H}^0 = \mathcal{H}^0_1 \cup \mathcal{H}^0_2 \cup \mathcal{H}^0_3$. The heteroclinic cycle $\mathcal{H}^0_1$ consists of four Taub points and the heteroclinic orbits that connect them,

$$\mathcal{H}^0_1 : \mathrm{T}^0_{22} \to \mathrm{T}^0_{32} \to \mathrm{T}^0_{33} \to \mathrm{T}^0_{23} \to \mathrm{T}^0_{22}, \tag{24}$$

see figure 3; $\mathcal{H}^0_2$ and $\mathcal{H}^0_3$ are defined analogously. The network $\mathcal{H}^0$ hence consists of the nine Taub points which are joined by heteroclinic orbits to form a connected path, see figure 5 and the schematic representation of figure 6; details follow in the proof of theorem 4.3.

Our dynamical systems formulation presented in section 2 makes it possible to describe the global dynamics towards the past:

Theorem 4.3. *The α-limit set of an orbit in the interior of the state space is one of the fixed points of the fixed point sets* F^0, D^0_i, QL^0_i, KC^0_i, *see theorem 4.1, or, possibly, the heteroclinic network* $\mathcal{H}^0$.

Conjecture. *In analogy with each individual fixed point on* KC_i^0 *(with* $1 < \Sigma_i < 2$*) and with the lines* QL_i^0*, the heteroclinic network* $\mathcal{H}^0$ *attracts a three-parameter (two-parameter) set of solutions in the massive (massless) case.*

The proof of theorem 4.3 and evidence for the conjecture will be presented below, split into five subsections. Since the arguments of the proof are lengthy and rather technical, we will first give a fairly non-technical interpretation of the theorem and the conjecture.

Interpretation of theorem 4.3 and the conjecture. Theorem 4.3 essentially states that the list given in theorem 4.1 is complete with the possible exception of the heteroclinic network $\mathcal{H}^0$. One individual transversally stable fixed point on KC_i^0 with $1 < \Sigma_i < 2$ attracts a three-parameter (two-parameter) set of orbits in the massive (massless) case; each such point represents a single non-LRS Kasner solution. Analogously, the line QL_i^0, taken as a whole, attracts a three-parameter (two-parameter) set of orbits. Each point on QL_i^0 represents one and the same non-flat LRS Kasner solution, i.e., the line QL_i^0, taken as a whole, is associated with a non-flat LRS Kasner solution. We see that points on KC_i^0 and QL_i^0 are not on an equal footing; this is an artefact due to our choice of variables. Instead, one should treat the individual fixed points on KC_i^0 (with $1 < \Sigma_i < 2$) and the line QL_i^0 (taken as a whole) 'democratically'—each represents a particular non-flat Kasner solution—each attracts a three-parameter (two-parameter) set of solutions in the massive (massless) case.

The Taub points T_{ij}^0 represent flat LRS Kasner solutions; they are associated with the Taub representation of Minkowski spacetime. None of the nine Taub points attracts orbits of the dynamical system, therefore there do not exist solutions of the Einstein–Vlasov system that possess a flat LRS Kasner solution as a past asymptotic state. Although the individual Taub points are not attractors, the heteroclinic network $\mathcal{H}^0$, which is built upon the Taub points, is: we conjecture (and prove the analogous statement for the special case of distributional distribution functions) that the heteroclinic network $\mathcal{H}^0$ attracts a three-parameter (two-parameter) set of solutions in the massive (massless) case. Hence, $\mathcal{H}^0$ should be treated on an equal footing with one individual fixed point of KC_i^0 (with $1 < \Sigma_i < 2$) or the lines QL_i^0. Solutions attracted by $\mathcal{H}^0$ oscillate between different representations of the Taub solution. The evolution is dominated by episodes when the solution is close to a particular Taub solution on standard form, but there will be transitions between such episodes that transport the solution from the neighbourhood of one Taub solution to the neighbourhood of another. It is important to note that the transitions lead to perpetual oscillations in Ω so that Ω does not converge to zero in the limit. For the particular case of solutions of the Einstein–Vlasov system associated with a distributional distribution function, the solutions that have $\mathcal{H}_1^0$ as the past asymptotic state are analysed in detail, see appendix D.

By definition, the past attractor of a dynamical system is the smallest closed invariant set that contains the α-limits of generic orbits, see appendix A. If we assume that the conjecture is correct (or if $\mathcal{H}^0$ attracts a lower dimensional set of solutions than in the conjecture), then the application of this definition yields that the past attractor is the union of the sets KC_i^0 (with $1 \leqslant \Sigma_i \leqslant 2$). From the point of view of physics this is somewhat misleading. The attractor excludes all asymptotic states corresponding to LRS Kasner solutions. Both the fixed points Q_{ij} ($i \neq j$) and the Taub points are part of the past attractor (since the attractor is defined to be a closed set), but there do not exist any solutions that converge to these points. From a physical perspective, the closure of the set KC_i^0 (with $1 < \Sigma_i < 2$) rather involves the whole lines QL_i^0 as well as the entire network $\mathcal{H}^0$: each element of this set (a point on KC_i^0 with $1 < \Sigma_i < 2$, a line QL_i^0, the network $\mathcal{H}^0$) attracts a three-parameter (two-parameter) set of solutions. Finally, note that if the conjecture were wrong and $\mathcal{H}^0$ attracted a four-parameter (three-parameter) set of solutions, then, of course, $\mathcal{H}^0$ would have to be added to the above mathematically defined attractor.

It is of interest to compare the dynamics of the Einstein–Vlasov equations with the perfect fluid situation (with a linear equation of state, $p = (\gamma - 1)\rho$ where $1 \leqslant \gamma < 2$). In this case each individual Kasner solution, including the flat Taub solutions, attracts one solution towards the past (see, e.g., [1, p 135]); this means that all solutions asymptotically approach Kasner states except for the isotropic and flat FRW solution. In the Vlasov case, both the non-generic and generic situations are more complicated: both F^0 and D_i^0 attract non-generic solutions, see theorem 4.1; the flat Taub solutions are not among the past asymptotic states, but the heteroclinic network $\mathcal{H}^0$ takes the role of the Taub solutions. Hence, instead of $\Omega \to 0$, Ω oscillates for solutions that approach the heteroclinic network; the consequences of this will be discussed in the concluding remarks.

In the remainder of this section, we give the proof of theorem 4.3 and evidence for the conjecture. (As the arguments of the proof are rather technical, the reader who is not interested in the details of the global dynamical systems analysis may skip ahead to the concluding remarks.) The first step in the proof is to gain a detailed understanding of the dynamics on the relevant invariant subspaces of the dynamical system.

4.2.1. Dynamics on $\mathcal{S}_i^0$

Lemma 4.4. *Consider an orbit in the interior of $\mathcal{S}_i^0$. Its α-limit set is a fixed point on* KC_j^0 *or* KC_k^0 $(i \neq j \neq k)$, QL_i^0 *or* TL_i^0, *or it is the heteroclinic cycle $\mathcal{H}_i^0$, defined in (24). The ω-limit set is the fixed point* D_i^0.

Proof. Without loss of generality we consider $\mathcal{S}_1^0$, which can be described by the variables

$$0 < s_2 < 1 \quad (s_3 = 1 - s_2) \qquad \text{and} \qquad \Sigma_1, \Sigma_2, \Sigma_3 \quad (\Sigma_1 + \Sigma_2 + \Sigma_3 = 0, \Sigma^2 < 1), \quad (25)$$

hence $\mathcal{S}_1^0$ is represented by the interior of a cylinder, cf figure 3. The boundary of $\mathcal{S}_1^0$ consists of the lateral boundary $\mathcal{S}_1^0 \cap \mathcal{K}^0$, the base $\mathcal{S}_{12}^0$ and the top surface $\mathcal{S}_{13}^0$.

Since $\mathcal{S}_1^0 \cap \mathcal{K}^0$ is part of $\mathcal{K}$, it follows that $\Sigma_i \equiv$ const for all orbits on $\mathcal{S}_1^0 \cap \mathcal{K}^0$. We observe that s_2 is monotonically increasing (decreasing) when $\Sigma_2 < \Sigma_3$ ($\Sigma_2 > \Sigma_3$), since $s_2' = -2s_2(1 - s_2)(\Sigma_2 - \Sigma_3)$; the two domains are separated by the lines of fixed points TL_1^0 and QL_1^0, see figure 3.

The key equations to understand the flow on $\mathcal{S}_{12}^0$ are

$$\Omega' = \Omega(2\Sigma^2 - \Sigma_3) \qquad \text{and} \qquad \Sigma_3' = \Omega(2 - \Sigma_3). \quad (26)$$

From the first equation it follows that all points on KC_3^0 are transversally hyperbolic repelling fixed points except for T_{33}^0; from the second equation we infer that T_{33}^0 is the attractor of the entire interior of $\mathcal{S}_{12}^0$. Similarly, T_{22}^0 is the attractor on $\mathcal{S}_{13}^0$, see figure 3.

The plane $\mathcal{D}_1^0$, defined by $\Sigma_1 = -1$, is an invariant subset in $\mathcal{S}_1^0$. In the interior of the plane we find the fixed point D_1^0; the boundary consists of a heteroclinic cycle $\mathcal{H}_1^0$, see (24). (Note that analogous cycles $\mathcal{H}_2^0$ and $\mathcal{H}_3^0$ exist on $\mathcal{S}_2^0$ and $\mathcal{S}_3^0$, respectively.) The function $M_{(5)}$ is monotonically decreasing on $\mathcal{D}_1^0$, cf (23). Application of the monotonicity principle, see appendix A, yields that D_1^0 is the ω-limit and that $\mathcal{H}_1^0$ is the α-limit for all orbits on $\mathcal{D}_1^0$, cf figure 3.

Consider now an orbit in $\mathcal{S}_1^0$ with $\Sigma_1 \neq -1$. The function $M_{(4)} = (1 + \Sigma_1)^2$ is monotonically decreasing on $\mathcal{S}_1^0$, cf (22). The monotonicity principle implies that the ω-limit lies on $\Sigma_1 = -1$ or $\Sigma^2 = 1$ (but $\Sigma_1 \neq \pm 2$). Since the logarithmic derivative of Ω is positive everywhere on $\mathcal{S}_1^0 \cap \mathcal{K}^0$ (except at T_{22}^0 and T_{33}^0), i.e., $\Omega^{-1}\Omega'|_{\Omega=0} = 2 - \sum_k w_k \Sigma_k > 0$, it follows that the 'wall' $\mathcal{S}_1^0 \cap \mathcal{K}^0$ of the cylinder is repelling everywhere away from $\Sigma_1 = -1$. Consequently, the ω-limit of the orbit cannot lie on $\Sigma^2 = 1$ but is contained in $\Sigma_1 = -1$. The fixed point D_1^0 is a hyperbolic sink, as we conclude from the dynamics on $\Sigma_1 = -1$ and from $(1+\Sigma_1)^{-1}(1+\Sigma_1)'|_{\mathrm{D}_1^0} = -3/4$. Therefore, the *a priori* possible ω-limit sets on $\Sigma_1 = -1$ are D_1^0 and the heteroclinic cycle $\mathcal{H}_1^0$.

To prove that D_1^0 is the actual ω-limit, we again consider the function $M_{(5)}$. However, we no longer restrict its domain of definition to $\mathcal{D}_1^0$, but view it as a function on $\mathcal{S}_1^0$; we obtain

$$12M'_{(5)} = -M_{(5)}[(\Sigma_1 + 2\Sigma_2)^2 + (\Sigma_1 + 2\Sigma_3)^2 + 6(\Sigma_1 + 1)^2 - 6(\Sigma_1 + 1)]. \tag{27}$$

The bracket is positive when $\Sigma_1 < -1$; hence $M_{(5)}$ is decreasing when $\Sigma_1 < -1$. This prevents orbits with $\Sigma_1 < -1$ from approaching $\mathcal{H}_1^0$, since the cycle is characterized by $M_{(5)} = \infty$. Now suppose that there exists an orbit in $\Sigma_1 > -1$, whose ω-limit is $\mathcal{H}_1^0$. At late times the trajectory shadows the cycle; hence, for late times, the bracket in (27) is almost always positive along the trajectory—only when the trajectory passes through a small neighbourhood of $(\Sigma_1, \Sigma_2, \Sigma_3) = (-1, 1/2, 1/2)$ is the bracket marginally negative. Since the trajectory spends large amounts of time near the fixed points and the passages from one fixed point to another become shorter and shorter in proportion, it follows that at late times $M_{(5)}$ is decreasing along the orbit (with ever shorter periods of limited increase). This is a contradiction to the assumption that the orbit is attracted by the heteroclinic cycle. We therefore draw the conclusion that D_1^0 is the global sink on $\mathcal{S}_1^0$.

Consider again an orbit in $\mathcal{S}_1^0$ with $\Sigma_1 \neq -1$. Invoking the monotonicity principle with the function $M_{(4)}$ we find that the α-limit of the orbit must be located on $\Sigma^2 = 1$, $\Sigma_1 \neq -1$. From the analysis of the flow on the boundaries of the cylinder, we obtain that all fixed points on $\Sigma^2 = 1$ except for T_{22}^0 and T_{33}^0 are transversally hyperbolic. The fixed points on KC_2^0 with $\Sigma_2 < \Sigma_3$ and the points on KC_3^0 with $\Sigma_2 > \Sigma_3$ are saddles; the fixed points on KC_2^0 with $\Sigma_2 > \Sigma_3$ and those on KC_3^0 with $\Sigma_2 < \Sigma_3$ are transversally hyperbolic sources (except for T_{22}^0, T_{33}^0): every point attracts a one-parameter set of orbits from $\mathcal{S}_1^0$ as $\tau \to -\infty$. In contrast, each fixed point on TL_1^0 and QL_1^0 is a source for exactly one orbit. The structure of the flow on $\Sigma^2 = 1$ implies that the α-limit of the orbit in $\mathcal{S}_1^0$ with $\Sigma_1 \neq -1$ must be one of the transversally hyperbolic sources. This establishes lemma 4.4. □

Mathematically, the past attractor is the union of the sets KC_2^0 (with $\Sigma_2 \geqslant \Sigma_3$) and KC_3^0 (with $\Sigma_2 \leqslant \Sigma_3$), see appendix A. However, the above establishes that each transversally stable fixed point on KC_2^0 and KC_3^0 attracts a one-parameter set of orbits towards the past and that the same is true for the lines QL_1^0 and TL_1^0 considered as a whole. As discussed in the concluding remarks, the present choice of variables yields multiple representations of the same solutions, thus, e.g., each fixed point on QL_1^0 represents the same non-flat LRS Kasner solution. Hence, the above statement can be interpreted physically by saying that the individual non-flat Kasner states and the Taub solution corresponding to TL_1^0 attract a one-parameter set of solutions each. The Taub states (flat Kasner solutions) corresponding to the points T_{ij} $(i, j \in \{2, 3\})$ are not attractive, but the heteroclinic cycle $\mathcal{H}_1^0$, which is built upon these points, is: the heteroclinic cycle $\mathcal{H}_1^0$ also attracts a one-parameter set of solutions. Hence, individual non-flat Kasner states, the Taub state TL_1^0 and $\mathcal{H}_1^0$ should be treated 'democratically' as regards the dynamics towards the past; physically it makes sense to define the past attractor on $\mathcal{S}_i^0$ as consisting of the union of the transversally stable fixed points on KC_2^0, KC_3^0, the lines QL_1^0, TL_1^0 and the network $\mathcal{H}_1^0$; i.e., the physical attractor consists of the non-flat Kasner states, the Taub state TL_1^0 and $\mathcal{H}_1^0$.

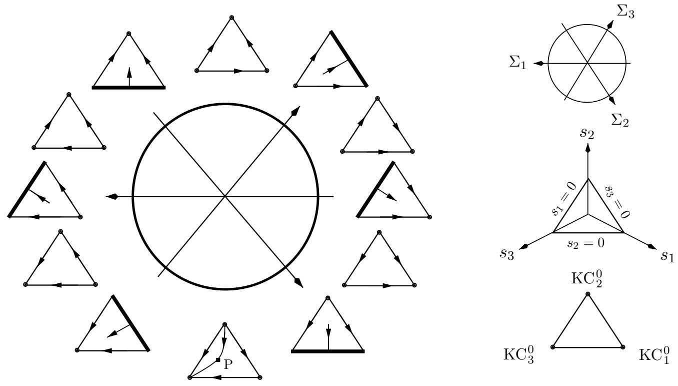

Figure 4. Schematic representation of the flow of the dynamical system on the invariant set $\mathcal{K}^0 = (z = 0) \cap (\Sigma^2 = 1)$, which is the Cartesian product of the $(\Sigma^2 = 1)$-circle and the s_i-triangle. The depicted fixed points are the Kasner fixed points and the lines of fixed points TL_i^0 $(\leftrightarrow \Sigma_i = 2)$ and QL_i^0 $(\leftrightarrow \Sigma_i = -2)$. All orbits are heteroclinic, like the orbit through the arbitrary point P that connects KC_2^0 with KC_3^0.

4.2.2. Dynamics on $\mathcal{K}^0$. The invariant subset $\mathcal{K}^0$ is defined by setting $z = 0$ and $\Sigma^2 = 1$; it can be represented by the Cartesian product of the circle $(\Sigma^2 = 1)$ in the Σ_i-space times the s_i-triangle given by $\left\{0 < s_1, s_2, s_3 < 1, \sum_k s_k = 1\right\}$. The flow on this space possesses a simple structure: since $\Sigma_i' \equiv \text{const}$ for all orbits, the dynamical freedom resides in the s_i-spaces.

A schematic depiction of the flow on $\mathcal{K}^0$ is given in figure 4. All fixed points are located on the boundaries of $\mathcal{K}^0$, i.e., on $s_1 = 0, s_2 = 0$ or $s_3 = 0$. The vertices of the s_i-triangle are the Kasner circles KC_i^0. If $(\Sigma_1, \Sigma_2, \Sigma_3) \in (\Sigma^2 = 1)$ is such that $\Sigma_k = 2$ (respectively $\Sigma_k = -2$) for some k, then the side $s_k = 0$ of the triangle is a line of fixed points, TL_k^0 (respectively QL_k^0). Note that all fixed points are transversally hyperbolic on $\mathcal{K}^0$ and that they constitute the α- and ω-limit sets for all orbits on $\mathcal{K}^0$. The character of the fixed points, i.e., whether they are (transversal) attractors or repellors, depends on the sector of the circle $(\Sigma^2 = 1)$, see figure 4.

The results about the global dynamics on $\mathcal{S}_i^0$ and $\mathcal{K}^0$ will turn out to be an integral part in the proof of theorem 4.3, which we will address next. First, we will prove the massless case of the theorem.

4.2.3. Dynamics on $\mathcal{Z}^0$. Let γ be an arbitrary orbit in the interior of $\mathcal{Z}^0, \gamma \neq \{\mathrm{F}^0\}$. The function $M_{(2)}$ is strictly monotonically decreasing on $\mathcal{Z}^0$ (except at F^0, where it has a minimum), cf (20) ff; hence we can use the monotonicity principle: the α-limit set $\alpha(\gamma)$ of γ must be located on the boundaries of $\mathcal{Z}^0$, i.e., on $\mathcal{S}_i^0$ or $\mathcal{K}^0$. The first step in our analysis is to prove that the interior of the subsets $\mathcal{S}_i^0$ and $\mathcal{K}^0$ cannot belong to $\alpha(\gamma)$, unless γ is one of three special orbits.

Recall from our analysis of $\mathcal{S}_i^0$ that the fixed point $\mathrm{D}_i^0 \in \mathcal{S}_i^0$ is a hyperbolic sink on $\mathcal{S}_i^0$. In the orthogonal direction, however, we obtain $s_i^{-1} s_i'|_{\mathrm{D}_i^0} = 3$. It follows that D_i^0 is a hyperbolic saddle in the state space $\mathcal{Z}^0$ and that there exists exactly one orbit δ_i^0 that emanates from D_i^0 into the interior of $\mathcal{Z}^0$. (Theorem 4.2 implies that δ_i^0 converges to the global sink F^0 as $\tau \to \infty$.)

Henceforth, let γ be different from δ_i^0. In order to show that $\alpha(\gamma)$ does not contain any point in the interior of $\mathcal{S}_i^0$, we perform a proof by contradiction: assume that $\alpha(\gamma)$ contains a point P in the interior of $\mathcal{S}_i^0$; then the whole orbit through P and the ω-limit $\omega(\mathrm{P})$ (as well as the α-limit) of that orbit must be contained in $\alpha(\gamma)$. As already shown, D_i^0 is the global attractor on $\mathcal{S}_i^0$, hence $\alpha(\gamma) \ni \omega(P) = D_i^0$. Since the saddle D_i^0 is in $\alpha(\gamma)$, the unique orbit δ_i^0 emanating from it is contained in $\alpha(\gamma)$ as well. Thus, ultimately, $\omega(\delta_i^0)$, i.e., the point F^0, must be contained in $\alpha(\gamma)$; this is a contradiction, since F^0 is a sink. Therefore, γ cannot contain any α-limit point in the interior of $\mathcal{S}_i^0$. We will now use similar reasoning repeatedly.

Assume next that $\alpha(\gamma)$ contains a point P in the interior of $\mathcal{K}^0$, i.e., a point with $\Sigma^2 = 1$, $0 < s_i < 1\ \forall i$. Suppose first that $\Sigma_k \neq \pm 2$ for all k. Since P is an element of $\alpha(\gamma)$, the whole orbit through P and the α-limit $\alpha(P)$ of that orbit must be contained in $\alpha(\gamma)$. From the dynamics on $\mathcal{K}^0$, cf figure 4, it follows that $\alpha(P)$ is one of the Kasner fixed points on KC_i^0, where i corresponds to the direction determined by $\Sigma_i = \max_k \Sigma_k$; we hence denote $\alpha(P)$ as K_P. Since $\Sigma_k \neq \pm 2$, it follows from the previous analysis that K_P is a transversally hyperbolic source on the subspace $\mathcal{K}^0$; $\Omega^{-1}\Omega'|_{\Sigma^2=1} = 2 - \sum_k w_k \Sigma_k > 0$ yields that K_P is a transversally hyperbolic source on the whole space $\mathcal{Z}^0$. Since $\alpha(\gamma)$ contains the transversally hyperbolic source K_P, that fixed point necessarily constitutes the entire α-limit set, i.e., $\alpha(\gamma) = K_P$. This is in contradiction to our assumption $\alpha(\gamma) \ni P$. The omitted cases $\Sigma_i = \pm 2$ for some i will be dealt with next.

Suppose that $\Sigma_i = -2$ for one index i. Assume that P lies in $\alpha(\gamma)$, therefore $\alpha(P)$ is contained in $\alpha(\gamma)$ as well. The dynamics on $\mathcal{K}^0$ implies that $\alpha(P)$ is a fixed point Q_P on QL_i^0, cf figure 4. This point is a transversally hyperbolic source; $\Omega^{-1}\Omega'|_{QL_i^0} = 1$ in this case. By the same argument as above we obtain a contradiction to the assumption $\alpha(\gamma) \ni P$.

Finally suppose that $\Sigma_i = 2$ for one index i. When we assume that P is in $\alpha(\gamma)$, then the ω-limit $\omega(P)$ is contained in $\alpha(\gamma)$. From figure 4, we see that $\omega(P)$ is a fixed point T_P on TL_i^0. The point T_P is a transversally hyperbolic saddle, since $\Omega^{-1}\Omega'|_{TL_i^0} = 3$, and there exists exactly one orbit that emanates from it, namely the orbit that connects T_P with D_i in $\mathcal{S}_i^0$. Since $T_P \in \alpha(\gamma)$, that orbit must also be contained in $\alpha(\gamma)$. This is in contradiction to the previous result: $\alpha(\gamma)$ cannot contain interior points of $\mathcal{S}_i^0$. Hence, our assumption $\alpha(\gamma) \ni P$ was false: the α-limit of γ cannot contain any interior point of $\mathcal{K}^0$.

Our analysis results in the following statement: there exist four special orbits, one trivial orbit corresponding to the fixed point F^0, and three orbits, the orbits δ_i^0, that converge to the fixed points $D_i^0 \in \mathcal{D}_i^0$. The α-limit set of every orbit γ in $\mathcal{Z}^0$ different from F^0 and δ_i^0 must be located on the boundaries of the spaces $\mathcal{S}_i^0$ and $\mathcal{K}^0$, i.e., on the union of the boundaries of the cylinders $\mathcal{S}_i^0$, which we denote by $\partial\mathcal{S}^0 = \partial\mathcal{S}_1^0 \cup \partial\mathcal{S}_2^0 \cup \partial\mathcal{S}_3^0$. The set $\partial\mathcal{S}^0$ is depicted in figure 2: it comprises the lateral surfaces of the cylinders and the base/top surfaces.

All fixed points on $\partial\mathcal{S}^0$ are transversally hyperbolic except for the points T_{ii}^0: TL_i^0 consists of transversally hyperbolic saddles; in contrast, the fixed points on QL_i^0 are transversally hyperbolic sources; points on KC_i^0 with $\Sigma_i > 1$, $\Sigma_i \neq 2$ are sources while those with $\Sigma_i < 1$ are saddles. Combining the analysis of the preceding sections, see figures 3 and 4, we obtain, more specifically: each point on QL_i^0 is a source for a one-parameter family of orbits that emanate into the interior of $\mathcal{Z}^0$ and each point on KC_i^0 with $\Sigma_i > 1$ ($\Sigma_i \neq 2$) is the source for a two-parameter family. (The points with $\Sigma_i = 1$ on KC_i^0 are the two points $Q_{ij}^0 \in QL_j^0$ and $Q_{ik}^0 \in QL_k^0$. Each of these two points is a transversally hyperbolic source for a one-parameter family of orbits; however, those orbits are not interior orbits, but remain on the boundary of $\mathcal{Z}^0$.)

The non-transversally hyperbolic fixed points T_{ii}^0 are part of a special structure that is present on $\partial\mathcal{S}^0$: the set $\partial\mathcal{S}^0$ exhibits a robust heteroclinic network $\mathcal{H}^0$ (of depth 1), see, e.g., [12] for a discussion of heteroclinic networks; the network $\mathcal{H}^0$ is depicted in figure 5; a schematic depiction is given in figure 6. In particular, we observe that the heteroclinic cycles $\mathcal{H}_i^0$ of the spaces $\mathcal{S}_i^0$ appear as heteroclinic subcycles of the network.

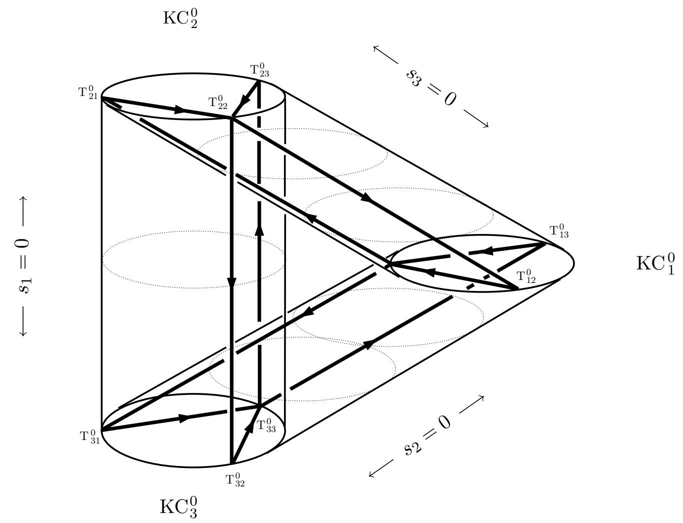

Figure 5. The heteroclinic network $\mathcal{H}^0$ that exists on the set $\partial\mathcal{S}^0$. Its building blocks are the heteroclinic cycles $\mathcal{H}^0_1$, $\mathcal{H}^0_2$, $\mathcal{H}^0_3$.

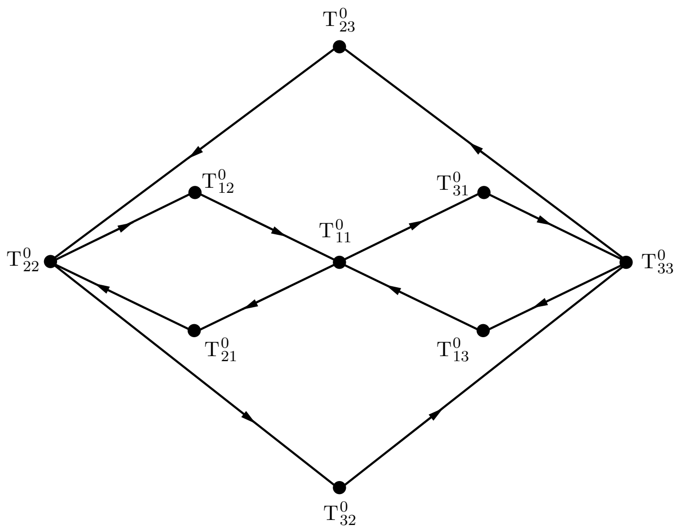

Figure 6. Schematic representation of $\mathcal{H}^0$.

A straightforward analysis of the flow on $\partial\mathcal{S}^0$ using the same type of reasoning as above leads to the result that there exist no other structures on $\partial\mathcal{S}^0$ that could serve as α-limits for an interior $\mathcal{Z}^0$-orbit γ. We have thus proved the following statement: the α-limit of γ is one of the transversally hyperbolic sources listed above or it is the heteroclinic network (or a heteroclinic subcycle thereof). This concludes the proof of the massless case of theorem 4.3.

4.2.4. Global dynamics in the massive case. Let γ be an arbitrary orbit in the interior of the state space $\mathcal{X}$. The function $M_{(1)}$ is strictly monotonically increasing on $\mathcal{X}$ (and on $\mathcal{K}$), cf (18) ff; moreover, $M_{(1)}$ vanishes for $z \to 1$ and $s_i \to 0$ (unless $z \to 0$ simultaneously).

Hence, by applying the monotonicity principle we obtain that the α-limit set $\alpha(\gamma)$ of γ must be located on $\mathcal{Z}^0$ including its boundaries.

Consider the fixed point $\mathrm{F}^0 \in \mathcal{Z}^0$. By theorem 4.2 this fixed point is a global sink on $\mathcal{Z}^0$. In the orthogonal direction, however, we have $z^{-1}z'|_{\mathrm{F}^0} = 2$. It follows that F^0 is a hyperbolic saddle in the space $\mathcal{X}$ and that there exists exactly one orbit ϕ that emanates from F^0 into the interior of $\mathcal{X}$. (Theorem 4.2 implies that ϕ converges to FS^1 as $\tau \to \infty$; thus, ϕ represents the unique solution of the Einstein–Vlasov equations that isotropizes towards the past and the future.)

Let γ be different from ϕ. Assume that $\alpha(\gamma)$ contains a point P in the interior of $\mathcal{Z}^0$; then the whole orbit through P and the ω-limit $\omega(\mathrm{P})$ must be contained in $\alpha(\gamma)$. Theorem 4.2 implies $\omega(\mathrm{P}) = \mathrm{F}^0$, hence $\mathrm{F}^0 \in \alpha(\gamma)$. Since the saddle F^0 is in $\alpha(\gamma)$, the unique orbit ϕ emanating from it is contained in $\alpha(\gamma)$ as well. Thus, ultimately, $\omega(\phi)$, i.e., a point on FS^1, must be contained in $\alpha(\gamma)$; this is a contradiction, since FS^1 consists of transversally hyperbolic sinks. We conclude that γ cannot contain any α-limit point in the interior of $\mathcal{Z}^0$.

Since $\alpha(\gamma)$ must be located on the boundary on $\mathcal{Z}^0$, i.e., on $\mathcal{S}_i^0$ or $\mathcal{K}^0$, the proof can be completed in close analogy to the proof in the massless case. We thus restrict ourselves here to giving some relations that establish that the sources on $\mathcal{Z}^0$ generalize to sources on $\mathcal{X}$: on KC_i^0 we have $z^{-1}z'|_{\mathrm{KC}_i^0} = 2(1+\Sigma_i)$, which is positive for all $\Sigma_i > -1$ and thus for $\Sigma_i > 1$ in particular; for QL_i^0 we obtain $z^{-1}z'|_{\mathrm{QL}_i^0} = 4$. We further note that $z^{-1}z'|_{\mathrm{D}_i^0} = 3$; thus D_i^0 possesses a two-dimensional unstable manifold. (Orbits in that manifold converge to FS^1.) Finally, note that along the heteroclinic cycle $\mathcal{H}_1^0 : \mathrm{T}_{22}^0 \to \mathrm{T}_{32}^0 \to \mathrm{T}_{33}^0 \to \mathrm{T}_{23}^0 \to \mathrm{T}_{22}^0$, we obtain that $z^{-1}z'$ equals $6s_2, 2(1+\Sigma_3), 2(1+\Sigma_2), 6(1-s_2)$, respectively (and similarly for $\mathcal{H}_2^0$ and $\mathcal{H}_3^0$); from this it follows that $z^{-1}z'$ is positive on $\mathcal{H}^0$, except at T_{ij}^0 $(i \neq j)$ where $z^{-1}z' = 0$.

This concludes the proof of theorem 4.3.

4.2.5. Heuristic motivation for the conjecture. Let us first consider the space $\mathcal{S}_1^0$ (or, analogously, $\mathcal{S}_2^0, \mathcal{S}_3^0$). In the proof of lemma 4.4, we have classified the α-limits of orbits in $\mathcal{S}_1^0$. Each individual point on KC_2^0 with $\Sigma_2 > \Sigma_3$ (but $\Sigma_2 \neq 2$) and each point on KC_3^0 with $\Sigma_2 < \Sigma_3$ (but $\Sigma_3 \neq 2$) attracts a one-parameter family of orbits. Each point on QL_1^0 (where $\Sigma_2 = \Sigma_3 = 1$) attracts one orbit, hence QL_1^0 considered as a whole attracts a one-parameter family; analogously, TL_1^0 (where $\Sigma_2 = \Sigma_3 = -1$) attracts a one-parameter family, and so does the heteroclinic cycle $\mathcal{H}_1^0$; $\mathcal{H}_1^0$ is on an equal footing with an individual fixed point on KC_2^0 with $\Sigma_2 > \Sigma_3$ (but $\Sigma_2 \neq 2$), a point on KC_3^0 with $\Sigma_2 < \Sigma_3$ (but $\Sigma_3 \neq 2$) and with the lines QL_1^0 and TL_1^0. This constitutes a proof for the conjecture in the special case of the Einstein–Vlasov equations with a distributional distribution function.

We now note that $s_1^{-1}s_1' = z^{-1}z'$ along the heteroclinic cycle $\mathcal{H}_1^0$ and that $s_1^{-1}s_1' = z^{-1}z'$ is *positive* on $\mathcal{H}_1^0$, except at $\mathrm{T}_{23}^0, \mathrm{T}_{32}^0$, where the quantity vanishes. This strongly suggests that $\mathcal{H}_1^0$ (or its generalization $\mathcal{H}^0$) attracts a three-parameter (two-parameter) set of solutions in the massive (massless) case and that $\mathcal{H}^0$ can be viewed as being on an equal footing with the individual fixed points on KC_i^0 (with $1 < \Sigma_i < 2$) and with the lines QL_i^0. A proof of the conjecture would require a theorem generalizing the Hartman–Großman theorem and the centre manifold reduction theorem to heteroclinic networks.

5. Concluding remarks

In this paper, we have analysed the asymptotic behaviour of solutions of the Einstein–Vlasov equations with Bianchi type I symmetry. To that end we have reformulated the equations as a system of autonomous differential equations on a compact state space, which enabled us to employ powerful techniques from dynamical systems theory. However, our formulation yielded multiple representations of some structures, e.g., the Kasner solutions. This could have been avoided to a considerable extent by using other variables. Replacing s_i with $E_i = \sqrt{g^{ii}}/H$, i.e., the Hubble-normalized spatial frame variables of [7, 13], and using $y = m^2 H^{-2}$ instead of z, yields a single Kasner circle on the massless boundary instead of three. The latter variables, however, are not bounded; indeed, they blow up towards the future in the present case. It is possible to replace the variables by bounded variables; however, variables of this type lead to differentiability difficulties towards the singularity. Issues like these made the variables we employed in this paper more suitable for the kind of analysis we have performed. However, E_i-variables, or 'E_i-based' variables, would have yielded a more direct physical interpretation, and would have been more suitable to relate the present results to a larger context; but it is not difficult to translate our results to the E_i-variables, used in e.g. [7, 13], where the relationship between the dynamics of inhomogeneous and spatially homogeneous models was investigated and exploited.

In the present work we have not considered a cosmological constant, Λ. The effects of a positive cosmological constant can be outlined as follows: since $\rho \to \infty$ towards the singularity, it follows that Λ can be asymptotically neglected and hence that the singularity structure is qualitatively the same as for $\Lambda = 0$. However, towards the future Λ destabilizes FS^1, which becomes a saddle, and instead solutions isotropize and asymptotically reach a de Sitter state.

Based on the global dynamical systems analysis, we have identified all possible attracting sets of orbits—both in the massless and massive cases. We have proved that for the invariant subset $\mathcal{S}_1^0$, the heteroclinic cycle $\mathcal{H}_1^0$ attracts a one-parameter set of orbits just like the non-flat Kasner states; in a sense, $\mathcal{H}_1^0$ adopts the role of (two of the three equivalent) flat Taub Kasner states in the perfect fluid case. However, there is a difference: Ω oscillates for solutions that approach $\mathcal{H}_1^0$ while $\Omega \to 0$ for solutions that approach Kasner states. We have presented a heuristic argument that provides support for the conjecture that the heteroclinic network $\mathcal{H}^0$ will play an analogous role in the full state space, but unfortunately we have not been able to prove this. If true, $\mathcal{H}^0$ yields a new example of self-similar breaking at the initial singularity, see [14, 15].

If correct, the conjecture states that the heteroclinic network $\mathcal{H}^0$ takes on the role played by the flat Taub solutions in Bianchi type I. In the perfect fluid case, a solution that is associated with a Taub asymptote has a so-called weak null singularity, and therefore the solution can be $\mathcal{C}^0$ extended to the Minkowski spacetime (see [16, p 176]). This is not going to be the case for Vlasov matter: the Taub points themselves do not attract any solutions; a solution converging to $\mathcal{H}^0$ oscillates forever between different Taub states, which leads to an oscillation in Ω—Vlasov matter seems to prevent weak null singularities from forming.

Heuristically, it is reasonable to assume that the Bianchi type I Einstein–Vlasov case will play a role that is similar to that of the Bianchi type I perfect fluid case as regards singularities in a more general spatially homogeneous, or even spatially inhomogeneous, context; the mechanisms that make Bianchi type I so prominent—symmetry and source 'contractions', and asymptotic silence (see, e.g., [17, 7])—should work in both cases. This is of relevance for the above statement about weak null singularities, but even more so for generic spacelike singularities. For such singularities, the Taub states play a key role in the vacuum and perfect fluid cases—indeed they are one of the main obstacles for producing theorems in this context. Considering the above and that $\mathcal{H}^0$ takes on the role of the Taub points for Vlasov matter leads to an issue: does Vlasov matter generically lead to a 'simpler' singularity structure that makes it easier to establish new singularity theorems?

Under all circumstances, $\mathcal{H}^0$ will be important for the intermediate dynamical behaviour of many models, and thus there are significant differences between perfect fluid models and models with Vlasov matter. The existence of $\mathcal{H}^0$ is directly related to the anisotropy of the energy–momentum tensor; this leads to another question: which sources yield heteroclinic networks and the associated mathematically and physically interesting phenomena?

Acknowledgments

We gratefully acknowledge the hospitality and the support of the Isaac Newton Institute for Mathematical Sciences in Cambridge, UK, where part of this work was done. We also thank Alan Rendall and an anonymous referee for useful questions and comments. CU is supported by the Swedish Research Council.

Appendix A. Dynamical systems

In this appendix, we briefly recall some concepts from the theory of dynamical systems which we use in the paper.

Consider a dynamical system defined on an invariant set $X \subseteq \mathbb{R}^m$. The ω-limit set $\omega(x)$ (α-limit set $\alpha(x)$) of a point $x \in X$ is defined as the set of all accumulation points of the future (past) orbit of x. The simplest examples are fixed points and periodic orbits.

The monotonicity principle [1] gives information about the global asymptotic behaviour of the dynamical system. If $M : X \to \mathbb{R}$ is a $\mathcal{C}^1$ function which is strictly decreasing along orbits in X, then

$$\omega(x) \subseteq \left\{\xi \in \bar{X}\backslash X \;\Big|\; \lim_{\zeta\to\xi} M(\zeta) \neq \sup_X M\right\} \tag{A.1a}$$

$$\alpha(x) \subseteq \left\{\xi \in \bar{X}\backslash X \;\Big|\; \lim_{\zeta\to\xi} M(\zeta) \neq \inf_X M\right\} \tag{A.1b}$$

for all $x \in X$.

Locally in the neighbourhood of a fixed point, the flow of the dynamical system is determined by the stability features of the fixed point. If the fixed point is hyperbolic, i.e., if the linearization of the system at the fixed point is a matrix possessing eigenvalues with non-vanishing real parts, then the Hartman–Großman theorem applies: in a neighbourhood of a hyperbolic fixed point, the full nonlinear dynamical system and the linearized system are topologically equivalent. Non-hyperbolic fixed points are treated in centre manifold theory: the reduction theorem generalizes the Hartman–Großman theorem; for further details see, e.g., [18]. If a fixed point is an element of a connected fixed point set (line, surface, . . .) and the number of eigenvalues with zero real parts is equal to the dimension of the fixed point set, then the fixed point is called transversally hyperbolic. Application of the centre manifold reduction theorem is particularly simple in this case. (The situation is analogous in the more general case when the fixed point is an element of an *a priori* known invariant set that coincides with the centre manifold of the fixed point.)

Given a flow on a state space X, the future (past) attractor $A^+(A^-)$ is defined as the smallest closed invariant set such that $\omega(x) \subset A^+$ ($\alpha(x) \subset A^-$) for all points $x \in X$ apart from a set of measure zero.

Appendix B. FRW and LRS$_i$ symmetry

In this section, we discuss the sets FRW and LRS$_i$ in detail.

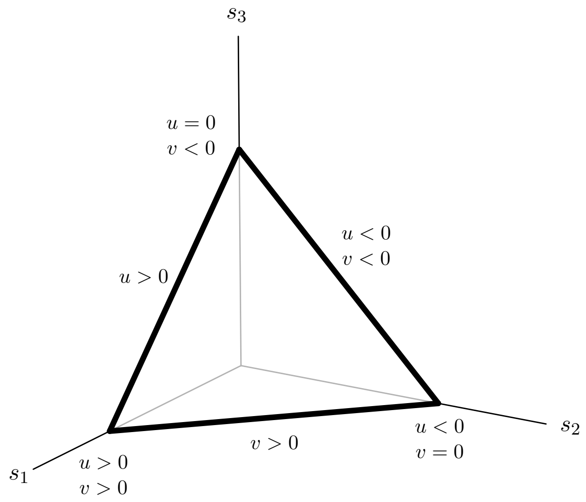

Figure 7. The functions u and v are monotonic along the boundaries of the space $\{(s_1, s_2, s_3) \,|\, s_i \geqslant 0, \sum_k s_k = 1\}$.

To begin with, we prove that the fixed point F^0 on $z = 0$ is well-defined and unique. Since the defining equations for F^0 are $w_1 = w_2 = w_3 = 1/3$, we must show that these equations indeed possess a unique solution $(s_1, s_2, s_3) = \left(s_1^{\mathrm{F}}, s_2^{\mathrm{F}}, s_3^{\mathrm{F}}\right)$ for all distribution functions f_0. Setting $z = 0$ in (9) implies that equations $w_1 = w_2 = w_3 = 1/3$ are equivalent to the system

$$u := \int f_0 \left[s_1 v_1^2 - s_2 v_2^2\right] \left(\sum_k s_k v_k^2\right)^{-1/2} \mathrm{d}^3 v = 0 \qquad \text{(B.1)}$$

and $v = 0$, where v is defined by replacing $\left[s_1 v_1^2 - s_2 v_2^2\right]$ by $\left[s_1 v_1^2 - s_3 v_3^2\right]$ in (B.1). On the three boundaries of the space $\left\{(s_1, s_2, s_3) \,\middle|\, s_i \geqslant 0, \sum_k s_k = 1\right\}$, the functions u and v are monotonic; their signs are given in figure 7. The derivative $\partial u/\partial s_1$ is manifestly positive, $\partial u/\partial s_2$ is negative, hence $\operatorname{grad} u$ is linearly independent of the surface normal $(1, 1, 1)$, and it follows that $u = \text{const}$ describes a curve for all const $\in \mathbb{R}$. The same argument applies to v, since $\partial v/\partial s_1 > 0$ and $\partial v/\partial s_3 < 0$. Figure 7 reveals that $u = 0$ ($v = 0$) connects the upper (right) vertex of the (s_1, s_2, s_3)-space with the opposite side. Investigating $(\operatorname{grad} u - \lambda \operatorname{grad} v)$ we find that the first component is manifestly positive when $\lambda \leqslant 2/3$ and negative when $\lambda \geqslant 3/2$, the second component is negative when $\lambda \leqslant 3$ and the third component is positive when $\lambda \geqslant 1/3$, which implies that $(\operatorname{grad} u - \lambda \operatorname{grad} v)$ is linearly independent of the surface normal $(1, 1, 1)$ for all λ. It follows that all equipotential curves of the functions u and v intersect transversally; hence $u = 0$ and $v = 0$ possess a unique point of intersection, which proves the claim.

The established existence and uniqueness result for the fixed point F^0 is independent of the prescribed distribution function f_0. Therefore, we have shown that for each distribution function there exists a unique FRW solution of the massless Einstein–Vlasov equations (represented by F^0).

The situation is different in the massive case. A FRW solution must satisfy the equations $\Sigma_i = 0$, $w_1 = w_2 = w_3 = w$ and $s_i = \text{const}\ \forall i$ (then one can use a rescaling of the spatial coordinates $\{x^i\}$ to bring the metric to a form that is explicitly isotropic, $g_{11}(t) = g_{22}(t) = g_{33}(t)$.) However, for a general distribution function f_0, these equations are incompatible with the Einstein–Vlasov equations; in other words, the straight line $\Sigma_i = 0$, $s_i = s_i^{\mathrm{F}}\ \forall i$ is not an orbit of the dynamical system. Hence, in the massive case, the Einstein–Vlasov equations do not admit a FRW solution for arbitrary f_0; the distribution function f_0 is required to satisfy FRW compatibility conditions, see below, in order for a FRW solution to exist.

Note, however, that for each f_0, there exists exactly one orbit that originates from F^0 and ends on FS^1, see section 4, i.e., there exists a unique solution of the Einstein–Vlasov equations that isotropizes towards the past and towards the future. This anisotropic solution can be regarded as a generalized FRW solution; if f_0 is compatible with the FRW geometry, then the generalized FRW solution reduces to an ordinary FRW solution.

The treatment of the LRS case is analogous: the subset LRS_1 (and, analogously, $\mathrm{LRS}_{2,3}$), defined through the equations $\Sigma_2 = \Sigma_3, w_2 = w_3$, describes solutions exhibiting LRS geometry. (For a solution on LRS_1, equation (13*b*) entails $s_2(\tau) \propto s_3(\tau)$; by rescaling the coordinates one can achieve $g_{22} = g_{33}$, i.e., a line element in an explicit LRS form.) However, for general f_0, the set LRS_1 is not invariant under the flow of the dynamical system, so that orbits will not remain on LRS_1. Consequently, for general f_0, the Einstein–Vlasov equations do not admit solutions with LRS geometry.

More specifically, consider

$$(\Sigma_2 - \Sigma_3)' = -3\Omega\left[\tfrac{1}{2}(1-w)(\Sigma_2-\Sigma_3) - (w_2 - w_3)\right]. \tag{B.2}$$

Hence, $(\Sigma_2 - \Sigma_3)'$ vanishes when $\Sigma_2 = \Sigma_3$ and $w_2 = w_3$. From (13*b*) and (13*c*), we obtain an equation for w_i',

$$w_i' = -2w_i\left[\Sigma_i - \sum_k \Sigma_k\left(\frac{1}{2}w_k + \frac{1}{2}w_i^{-1}\beta_{ik}^{(0)}\right) + \frac{z}{2}\left(w_i^{-1}\beta_i^{(1)} + \beta^{(1)}\right)\right], \tag{B.3}$$

where we have defined

$$\beta_{i_1\ldots i_k}^{(m)} = \frac{(1-z)^k \int f_0\left(\Pi_{n=1}^k s_{i_n} v_{i_n}^2\right)\left[z + (1-z)\sum_k s_k v_k^2\right]^{1/2-k-m}\,\mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3}{\int f_0\left[z+(1-z)\sum_k s_k v_k^2\right]^{1/2}\,\mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3}; \tag{B.4}$$

note that $w_i = \beta_i^{(0)}$. Equation (B.3) implies

$$(w_2 - w_3)' = -(\Sigma_1 - \Sigma_2)\left(\beta_{22}^{(0)} - \beta_{33}^{(0)}\right) - z(\Sigma_1 + 1)\left(\beta_2^{(1)} - \beta_3^{(1)}\right), \tag{B.5}$$

when $\Sigma_2 = \Sigma_3$ and $w_2 = w_3$. We conclude that the set $\Sigma_2 \equiv \Sigma_3, w_2 \equiv w_3$ is an invariant set of the dynamical system, iff $w_2 = w_3$ implies $\beta_{22}^{(0)} = \beta_{33}^{(0)}$ and $\beta_2^{(1)} = \beta_3^{(1)}$. (In the massless case, only the first condition is required.) These conditions are violated for general distribution functions; if the conditions hold for f_0, then this distribution function is said to be compatible with the LRS symmetry. This is the case, for instance, when there exist constants $a_2 > 0, a_3 > 0$, such that f_0 is invariant under the transformation $v_2 \to (a_3/a_2)v_3, v_3 \to (a_2/a_3)v_2$; e.g., $f_0 = \tilde{f}_0\big(v_1, v_2^2 v_3^2\big)$ or $f_0 = \tilde{f}_0\big(v_1, a_2^2 v_2^2 + a_3^2 v_3^2\big)$; in the latter case $w_2(\tau) \equiv w_3(\tau)$ implies $a_3^2 s_2(\tau) \equiv a_2^2 s_3(\tau)$.

Finally, note that a distribution function f_0 is compatible with a FRW geometry, if it is compatible with all LRS symmetries. This means that, for instance, $f_0 = \tilde{f}_0\big(a_1^2 v_1^2 + a_2^2 v_2^2 + a_3^2 v_3^2\big)$ is compatible with the FRW symmetry and thus admits a unique FRW solution of the Einstein–Vlasov equations.

Appendix C. Future asymptotics

In this section, we give the proof of theorem 4.2:

Theorem 4.2. *The ω-limit of every orbit in the interior of the massive state space $\mathcal{X}$ (massless state space $\mathcal{Z}^0$) is one of the fixed points* FS^1 *(the fixed point* F^0*).*

Proof. Consider first the state space $\mathcal{Z}^0$ of massless particles and the associated system (19). The function $M_{(2)}$, cf (20) ff, is well-defined and monotonically decreasing everywhere except for at the fixed point F^0, where it has a global minimum. On the boundaries $\mathcal{S}_i^0$ (given by $s_i = 0$) and $\mathcal{K}^0$ ($\Sigma^2 = 1$) of the state space $\mathcal{Z}^0$, the function $M_{(2)}$ is infinite. Therefore, the application of the monotonicity principle yields that the ω-limit of every orbit must be the fixed point F^0.

In the massive case consider (13a) in the form

$$\Sigma_i' = -3\Omega\left[\tfrac{1}{2}(1-w)(1+\Sigma_i) - \tfrac{1}{2}(1-3w) - w_i\right]. \tag{C.1}$$

The rhs is positive when $\Sigma_i \leqslant -1$ and $z > 0$ ($w < 1/3$). This implies that the hyperplanes $\Sigma_i = -1$ constitute semipermeable membranes in the state space $\mathcal{X}$, whereby the 'triangle' $(\Sigma_1 > -1) \cap (\Sigma_2 > -1) \cap (\Sigma_3 > -1)$ becomes a future invariant subset of the flow (13).

The first part of the proof is to show that every orbit enters the triangle at some time τ_e (and consequently remains inside for all later times).

Assume that there exists an orbit with $\Sigma_i(\tau) \leqslant -1$ for all τ (for some i). From (13b) we infer that

$$s_i' = -2s_i[s_j(\Sigma_i - \Sigma_j) + s_k(\Sigma_i - \Sigma_k)] > 0 \tag{C.2}$$

if $\Sigma_i < -1$ and that $s_i' \geqslant 0$ if $\Sigma_i = -1$; hence $s_i(\tau) \geqslant s_i(\tau_0) = \mathrm{const} > 0$ for all $\tau \in [\tau_0, \infty)$. From (16) we obtain

$$\begin{aligned} \tfrac{1}{3}\Omega^{-1}\Omega'\big|_{\Omega=0} &= 1 - \tfrac{1}{3}w_i(1+\Sigma_i) - \tfrac{1}{3}w_j(1+\Sigma_j) - \tfrac{1}{3}w_k(1+\Sigma_k) \\ &\geqslant 1 - w_j - w_k = (1-3w) + w_i \geqslant \mathrm{const} > 0, \end{aligned} \tag{C.3}$$

since $s_i \geqslant \mathrm{const} > 0$. Consequently, $\Omega(\tau) \geqslant \mathrm{const} > 0$ for all $\tau \in [\tau_0, \infty)$. It follows from (C.1) that

$$\Sigma_i' \geqslant \mathrm{const} > 0 \tag{C.4}$$

for all $\tau \in [\tau_0, \infty)$ by the same argument. This is in contradiction to the assumption $\Sigma_i \leqslant -1$ for all τ.

Thus, in the second part of the proof, we can consider an arbitrary orbit γ and assume, without loss of generality, that $\gamma(\tau)$ lies in the Σ-triangle for all $\tau \in [\tau_e, \infty)$. Equation (13$c$) leads to

$$z' = 2z(1-z)\sum_n s_n(1+\Sigma_n) \geqslant 0 \tag{C.5}$$

for all $\tau \in [\tau_e, \infty)$, hence $z(\tau) \geqslant z(\tau_e) > 0$ for all $\tau \in [\tau_e, \infty)$.

We define the function N by

$$N = (1+\Sigma_1)(1+\Sigma_2)(1+\Sigma_3). \tag{C.6}$$

The derivative can be estimated by

$$N' \geqslant 3\Omega N\left[-\frac{3}{2}(1-w) + \frac{1}{2}\sum_n \frac{1-3w}{1+\Sigma_n}\right]. \tag{C.7}$$

Since $w(\tau) \leqslant \mathrm{const} < 1/3$ (because $z(\tau) \geqslant \mathrm{const} > 0$), N' is positive when at least one of the Σ_i is sufficiently small, i.e., when N itself is small (a detailed analysis shows that $N' \geqslant 3\Omega N[-(3/2)(1-w) + \sqrt{3}(1-3w)N^{-1/2}]$). We conclude that there exists a positive constant N_0 such that $N(\tau) \geqslant N_0$ for all $\tau \in [\tau_e, \infty)$. This in turn implies that there exists

$\nu > 0$ such that $\Sigma_i(\tau) \geqslant -1 + \nu$ for all i, for all $\tau \in [\tau_e, \infty)$, whereby $z' \geqslant 2z(1-z)\nu$ for all $\tau \in [\tau_e, \infty)$.

It follows that the ω-limit of γ must lie on $z = 1$, i.e., on $\mathcal{Z}^1$. Taking into account the simple structure of the flow on $\mathcal{Z}^1$, characterized by $\Omega' = 3(1-\Omega)\Omega$, we conclude that the fixed points FS^1 given by $\Sigma_1 = \Sigma_2 = \Sigma_3 = 0$ are the only possible ω-limits. □

Remark. In order to demonstrate the versatility of the dynamical systems methods, we have chosen here to prove theorem 4.2 by using techniques that are slightly different from those employed in section 4 (which exploit the monotonicity principle). However, it is straightforward (in fact, even simpler) to give a proof by making use of the hierarchy of monotone functions. Indeed, the function $M_{(1)}$ ensures that the ω-limit of every orbit lies on $\mathcal{Z}^1$ or $\mathcal{S}_i$; modulo some subtleties, we can exclude that $\mathcal{S}_i$ is attractive by using the monotone function $M_{(3)}$ and the local properties of the fixed points.

Appendix D. The spaces $\mathcal{S}_i^0$—interpretation of solutions

The flow on the boundary subsets $\mathcal{S}_i^0$ is of fundamental importance in the analysis of the global dynamics of the state space, see section 4.2. Note that except for F^0 all attractors (D_i^0, QL_i^0, KC_i^0 and the heteroclinic network) lie on $\mathcal{S}_i^0$. For a depiction of the flow on $\mathcal{S}_1^0$, see figure 3. In the following we show that orbits on $\mathcal{S}_1^0$ represent solutions of the Einstein–Vlasov system that are associated with a special class of distribution functions. Furthermore, we investigate in detail solutions that converge to the subcycle $\mathcal{H}_1^0$ of the heteroclinic network.

Consider a distribution function f_0 of the form

$$f_0(v_1, v_2, v_3) = \delta(v_1) f_0^{\mathrm{red}}(v_2, v_3), \tag{D.1}$$

where $f_0^{\mathrm{red}}(v_2, v_3)$ is even in v_2 and v_3. In the case of massless particles, $m = 0$ (and $z = 0$ respectively), we obtain

$$w_1 = 0, \qquad w_j = \frac{g^{jj}\int f_0^{\mathrm{red}} v_j^2 \left[g^{22}v_2^2 + g^{33}v_3^2\right]^{-1/2} \mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3}{\int f_0^{\mathrm{red}} \left[g^{22}v_2^2 + g^{33}v_3^2\right]^{1/2} \mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3} \quad (j = 2, 3), \tag{D.2}$$

where g^{22} and g^{33} can be replaced by s_2 and s_3, if desired. In the unbounded variables g^{ii}, the equations read

$$\Sigma_1' = -\Omega[1 + \Sigma_1], \qquad (g^{11})' = -2g^{11}(1 + \Sigma_1) \tag{D.3a}$$

$$\Sigma_j' = -\Omega[1 + \Sigma_j - 3w_j], \qquad (g^{jj})' = -2g^{jj}(1 + \Sigma_j) \quad (j = 2, 3), \tag{D.3b}$$

cf the remark at the end of section 2. In particular, we note that the equation for g^{11} decouples; hence the full dynamics is represented by a reduced system in the variables $(\Sigma_1, \Sigma_2, \Sigma_3, g^{22}, g^{33})$, which coincides with the system (D.3) on the invariant subset $g^{11} = 0$. In analogy to the definitions (7a), we set

$$s_1 = 0, \qquad s_2 = \frac{g^{22}}{g^{22} + g^{33}}, \qquad s_3 = \frac{g^{33}}{g^{22} + g^{33}}, \tag{D.4}$$

so that $s_2 + s_3 = 1$. This results in the dynamical system

$$\Sigma_1' = -\Omega[1 + \Sigma_1], \qquad s_1 \equiv 0 \tag{D.5a}$$

$$\Sigma_j' = -\Omega[1 + \Sigma_j - 3w_j], \qquad s_j' = -2s_j[\Sigma_j - (s_2\Sigma_2 + s_3\Sigma_3)] \quad (j = 2, 3). \tag{D.5b}$$

This system (D.5a) coincides with the dynamical system (13) induced on $\mathcal{S}_1^0$ (which is obtained by setting $z = 0$, thus $w = 1/3$ and $s_1 = 0$ in (13).

Our considerations show that the flow on $\mathcal{S}_1^0$ possesses a direct physical interpretation: orbits on $\mathcal{S}_1^0$ represent solutions of the massless Einstein–Vlasov system of Bianchi type I with a 'distributional' distribution function of the type (D.1). Note that the system (D.5) on $\mathcal{S}_1^0$ must be supplemented by the decoupled equations (11b) and $(g^{11})' = -2g^{11}(1+\Sigma_1)$ in order to construct the actual solution from an orbit in $\mathcal{S}_1^0$.

Two structures in $\mathcal{S}_1^0$ are of particular interest: the fixed point D_1^0 and the heteroclinic cycle $\mathcal{H}_1^0$, see figure 3. The fixed point D_1^0 represents an LRS solution (associated with a distributional f_0); it is straightforward to show that the metric is of the form

$$g_{11} = \text{const}, \qquad g_{22} \propto t^{4/3}, \qquad g_{33} \propto t^{4/3}, \tag{D.6}$$

and $H = (4/9)t^{-1}$.

The orbit $\mathrm{T}_{22}^0 \to \mathrm{T}_{32}^0$, which is part of $\mathcal{H}_1^0$, corresponds to a solution

$$g_{11} = g_{11}^0, \qquad g_{22} = g_{22}^0(3H_0t)^2, \qquad g_{33} = g_{33}^0, \tag{D.7a}$$

here, $H = (3t)^{-1}$; H_0 is a characteristic value of H. For the orbit $\mathrm{T}_{33}^0 \to \mathrm{T}_{23}^0$, the result is analogous with g_{22} and g_{33} interchanged. A more extensive computation shows that the orbit $\mathrm{T}_{32}^0 \to \mathrm{T}_{33}^0$ leads to

$$g_{11} = g_{11}^0, \qquad g_{22} = g_{22}^0[\log(1+3H_0t)]^2, \qquad g_{33} = g_{33}^0(1+3H_0t)^2, \tag{D.7b}$$

together with $H = H_0(1+3H_0t)^{-1}(1+[\log(1+3H_0t)]^{-1})$. (Note that $3Ht$ is always close to 1 and approaches 1 for $t \to 0$ and $t \to \infty$.) The result for the orbit $\mathrm{T}_{23}^0 \to \mathrm{T}_{22}^0$ is analogous with g_{22} and g_{33} interchanged.

An orbit close to the fixed point T_{32}^0 corresponds to a solution

$$g_{11} = g_{11}^0, \qquad g_{22} = g_{22}^0(3H_0t)^2, \qquad g_{33} = g_{33}^0. \tag{D.7c}$$

This is reflected in (D.7a) and in (D.7b) (since (D.7b) approximates this behaviour in the limit of small t). The result for the fixed point T_{23}^0 is analogous with g_{22} and g_{33} interchanged. Orbits in the neighbourhood of T_{22}^0 lead to more complicated behaviour:

$$g_{11} = g_{11}^0, \qquad g_{22} = g_{22}^0(3H_0t)^2, \qquad g_{33} = g_{33}^0 h(3H_0t), \tag{D.7d}$$

where h is a function that grows logarithmically initially to become approximately constant (set to one). The analogous result holds for T_{33}^0.

Now consider an orbit converging to the heteroclinic cycle as $\tau \to -\infty$, i.e., $t \searrow 0$. Epochs associated with the orbit being close to one of the fixed points lead to (D.7c) and (D.7d); transition periods, i.e., episodes in which the orbit is close to one of the four heteroclinic orbits (and far from the fixed points) yield characteristic behaviour of the types (D.7a) and (D.7b). The joining of epochs and transition episodes is reflected in a matching of the constants in (D.7).

Example. Let $t^{(n)}$ denote a monotone sequence of times such that the solution is in transition episode (n) at time $t^{(n)}$ (i.e., the orbit is close to one of the four heteroclinic orbits and far from the fixed points); $t^{(n)} \searrow 0$ as $n \to \infty$. Since $3Ht \approx 1$ as $t \searrow 0$, the sequence $t^{(n)}$ gives rise to a sequence $H^{(n)}$ defined by $3H^{(n)}t^{(n)} = 1$. During episode (n) the solution exhibits characteristic behaviour of the type (D.7a) or (D.7b) with $H_0 = H^{(n)}$ (and $g_{22}^0 = g_{22}^{(n)}$, $g_{33}^0 = g_{33}^{(n)}$). Suppose that the orbit is close to the heteroclinic orbit $\mathrm{T}_{32}^0 \to \mathrm{T}_{33}^0$ in episode (n). We obtain behaviour of the type (D.7b) with $H_0 = H^{(n)}$. As $H^{(n)}t$ gets small, we see that $g_{22} \approx g_{22}^{(n)}(3H^{(n)}t)^2$, $g_{33} \approx g_{33}^{(n)}$, i.e., the orbit enters an epoch described by (D.7c). The next (as $t \searrow 0$) episode corresponds to the orbit running close to $\mathrm{T}_{22}^0 \to \mathrm{T}_{32}^0$;

the behaviour of the solution is (D.7*a*) with $g_{22}^{(n+1)}$, $g_{33}^{(n+1)}$ and $H_0 = H^{(n+1)}$. The matching between the episodes (n) and $(n+1)$ is thus straightforward: $g_{22}^{(n+1)}(H^{(n+1)})^2 = g_{22}^{(n)}(H^{(n)})^2$ and $g_{33}^{(n+1)} = g_{33}^{(n)}$. Matching episodes $(n+1)$ and $(n+2)$ is slightly more involved. As $t \searrow 0$, episode $(n+1)$ is followed by an epoch characterized by (D.7*d*), i.e., $g_{22} = g_{22}^{(n+1)}(3H_0 t)^2$ and $g_{33} = g_{33}^{(n+1)} h(3H_0 t)$. In episode $(n+2)$, the orbit is close to the heteroclinic orbit $\mathrm{T}_{23}^0 \to \mathrm{T}_{22}^0$, where

$$g_{11} = g_{11}^0, \qquad g_{22} = g_{22}^{(n+2)}(1 + 3H^{(n+2)}t)^2, \qquad g_{33} = g_{33}^{(n+2)}[\log(1 + 3H^{(n+2)}t)]^2. \tag{D.8}$$

When $H^{(n+2)}t$ is large, as it is as the beginning (as $t \searrow 0$) of episode $(n+2)$, we get $g_{22} = g_{22}^{(n+2)}(3H^{(n+2)}t)^2$ and $g_{33} = g_{33}^{(n+2)}(\log 3H^{(n+2)}t)^2$. The matching between episode $(n+1)$ and $(n+2)$ thus yields $g_{22}^{(n+2)}(H^{(n+2)})^2 = g_{22}^{(n+1)}(H^{(n+1)})^2$; the matching of g_{33} involves information on the function h.

References

[1] Wainwright J and Ellis G F R 1997 *Dynamical Systems in Cosmology* (Cambridge: Cambridge University Press)

[2] Rendall A D 1996 The initial singularity in solutions of the Einstein–Vlasov system of Bianchi type I *J. Math. Phys.* **37** 438

[3] Rendall A D and Tod K P 1999 Dynamics of spatially homogeneous solutions of the Einstein–Vlasov equations which are locally rotationally symmetric *Class. Quantum Grav.* **16** 1705

[4] Rendall A D and Uggla C 2000 Dynamics of spatially homogeneous locally rotationally symmetric solutions of the Einstein–Vlasov equations *Class. Quantum Grav.* **17** 4697

[5] Rendall A D 2002 Cosmological models and centre manifold theory *Gen. Rel. Grav.* **34** 1277

[6] Wald R M 1984 *General Relativity* (Chicago, IL: University of Chicago Press)

[7] Uggla C, van Elst H, Wainwright J and Ellis G F R 2003 The past attractor in inhomogeneous cosmology *Phys. Rev.* D **68** 103502

[8] Misner C W, Thorne K S and Wheeler J A 1973 *Gravitation* (San Francisco, CA: Freeman)

[9] Andréasson H 2005 The Einstein–Vlasov system/kinetic theory *Living Rev. Rel.* **8** lrr-2005-2, online journal article (cited on 30 June 2005) (http://www.livingreviews.org/llr-2005-2)

[10] Rendall A D 2004 The Einstein–Vlasov system *The Einstein Equations and the Large Scale Behaviour of Gravitational Fields* ed P T Chruściel and H Friedrich (Basle: Birkhäuser) p 235

[11] Maartens R and Maharaj S D 1990 Collision-free gases in Bianchi space-times *Gen. Rel. Grav.* **22** 595

[12] Ashwin P and Field M 1999 Heteroclinic networks in coupled cell systems *Arch. Ration. Mech. Anal.* **148** 107

[13] Röhr N and Uggla C 2005 Conformal regularization of Einstein's field equations *Class. Quantum Grav.* **22** 3775

[14] Wainwright J, Hancock M J and Uggla C 1999 Asymptotic self-similarity breaking at late times in cosmology *Class. Quantum Grav.* **16** 2577

[15] Lim W C, Uggla C and Wainwright J 2006 Asymptotic self-similarity breaking at late times in cosmology *Class. Quantum Grav.* **23** 2607

[16] Hawking S W and Ellis G F R 1973 *The Large Scale Structure of Space-Time* (Cambridge: Cambridge University Press)

[17] Ashtekar A, Tate R S and Uggla C 1993 Minisuperspaces: observables and quantization *Int. J. Mod. Phys.* D **2** 15

[18] Crawford J D 1991 Introduction to bifurcation theory *Rev. Mod. Phys.* **63–4** 991