MATHEMATICS OF GRAVITATION
PART I, LORENTZIAN GEOMETRY AND EINSTEIN EQUATIONS
BANACH CENTER PUBLICATIONS, VOLUME 41
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 1997

AN INTRODUCTION TO THE EINSTEIN–VLASOV SYSTEM

ALAN D. RENDALL
Max-Planck-Institut für Gravitationsphysik
Schlaatzweg 1, 14473 Potsdam, Germany
E-mail: rendall@aei-potsdam.mpg.de

1. Introduction. The Vlasov equation arises in kinetic theory. It gives a statistical description of a collection of particles. It is distinguished from other equations of kinetic theory by the fact that there is no direct interaction between particles. In particular, no collisions are included in the model. Each particle is acted on only by fields which are generated collectively by all particles together. The fields which are taken into account depend on the physical situation being modelled. In plasma physics, where this equation is very important, the interaction is electromagnetic and the fields are described either by the Maxwell equations or, in a quasi-static approximation, by the Poisson equation [26]. In gravitational physics, which is the subject of the following, the fields are described by the Einstein equations or, in the Newtonian approximation, by the Poisson equation. (There is a sign difference in the Poisson equation in comparison with the electromagnetic case due to the replacement of a repulsive by an attractive force.) The best known applications of the Vlasov equation to self-gravitating systems are to stellar dynamics [3]. It can also be applied to cosmology. In the first case the systems considered are galaxies or parts of galaxies where there is not too much dust or gas which would require a hydrodynamical treatment. Possible applications are to globular clusters, elliptical galaxies and the central bulge of spiral galaxies. The 'particles' in all these cases are stars. In the cosmological case they are galaxies or even clusters of galaxies. The fact that they are modelled as particles reflects the fact that their internal structure is believed to be irrelevant for the dynamics of the system as a whole.

These lectures are concerned not with the above physical applications but with some basic mathematical aspects of the Einstein–Vlasov system. First the definition and general mathematical properties of this system of partial differential equations are discussed and then the Cauchy problem for this system is formulated. The central theme in what follows is the global Cauchy problem, where 'global' means global in time. Up to now

---

1991 *Mathematics Subject Classification*: Primary 83C05; Secondary 83C55.
The paper is in final form and no version of it will be published elsewhere.

global results have only been obtained in very special cases and one of these, the case of spherically symmetric asymptotically flat solutions, is discussed in detail. The global existence results presented depend on having a suitable local existence result. At first the necessary local existence theorem is stated without proof and used in obtaining global theorems. The proof of the local existence theorem is discussed in some detail afterwards. Further information on kinetic theory in general relativity may be found in [10].

Let $(M, g_{\alpha\beta})$ be a spacetime, i.e. M is a four-dimensional manifold and $g_{\alpha\beta}$ is a metric of Lorentz signature $(-,+,+,+)$. Note that $g_{\alpha\beta}$ denotes a geometric object here and not the components of the geometric object in a particular coordinate system. In other words the indices are abstract indices. (See [28], section 2.4 for a discussion of this notation.) It is always assumed that the metric is time-orientable, i.e. that the two halves of the light cone at each point of M can be labelled past and future in a way which varies continuously from point to point. With this global direction of time, it is possible to distinguish between future-pointing and past-pointing timelike vectors. The worldline of a particle of non-zero rest mass m is a timelike curve in spacetime. The unit future-pointing tangent vector to this curve is the 4-velocity v^α of the particle. Its 4-momentum p^α is given by mv^α. There are different variants of the Vlasov equation depending on the assumptions made. Here it is assumed that all particles have the same mass m but it would also be possible to allow a continuous range of masses. When all the masses are equal, units can be chosen so that $m = 1$ and no distinction need be made between 4-velocity and 4-momentum. There is also the possibility of considering massless particles, whose wordlines are null curves. In the case $m = 1$ the possible values of the four-momentum are precisely all future-pointing unit timelike vectors. These form a hypersurface P in the tangent bundle TM called the mass shell. The distribution function f, which represents the density of particles with given spacetime position and four-momentum, is a non-negative real-valued function on P. A basic postulate in general relativity is that a free particle travels along a geodesic. Consider a future-directed timelike geodesic parametrized by proper time. Then its tangent vector at any time is future-pointing unit timelike. Thus this geodesic has a natural lift to a curve on P, by taking its position and tangent vector together. This defines a flow on P. Denote the vector field which generates this flow by X. (This vector field is what is sometimes called the geodesic spray in the mathematics literature.) The condition that f represents the distribution of a collection of particles moving freely in the given spacetime is that it should be constant along the flow, i.e. that $Xf = 0$. This equation is the Vlasov equation, sometimes also known as the Liouville or collisionless Boltzmann equation.

To get an explicit expression for the Vlasov equation, it is necessary to introduce local coordinates on the mass shell. In the following local coordinates x^α on spacetime are always chosen such that the hypersurfaces x^0=const. are spacelike. (Greek and Roman indices take the values $0, 1, 2, 3$ and $1, 2, 3$ respectively.) Intuitively this means that x^0, which may also be denoted by t, is a time coordinate and that the x^a are spatial coordinates. A timelike vector is future-pointing if and only if its zero component in a coordinate system of this type is positive. It is not assumed that the vector $\partial/\partial x^0$ is timelike. One way of defining local coordinates on P is to take the spacetime coordinates x^α together with the spatial components p^a of the four-momentum in these coordinates.

Then the explicit form of the Vlasov equation is:

$$\partial f/\partial t + (p^a/p^0)\partial f/\partial x^a - (\Gamma^a_{\beta\gamma} p^\beta p^\gamma/p^0)\partial f/\partial p^a = 0 \tag{1.1}$$

where $\Gamma^\alpha_{\beta\gamma}$ are the Christoffel symbols associated to the metric $g_{\alpha\beta}$. Here it is understood that p^0 is to be expressed in terms of p^a and the metric using the relation $g_{\alpha\beta}p^\alpha p^\beta = -1$. An alternative way of coordinatizing the mass shell which is often useful is to use the components of the four-momentum in an orthormal frame, which has a priori nothing to do with the frame defined by the coordinates. It should be chosen so that the first vector is future-pointing timelike. Here only the case where the first vector is the unit normal to the hypersurfaces of constant time will be considered. The explicit form of the Vlasov equation in these coordinates is similar to (1.1), with the Christoffel symbols replaced by the Ricci rotation coefficients $\gamma^\mu_{\nu\rho}$ of the frame. Explicitly, it is given by:

$$e^0_0\partial f/\partial t + (v^\mu/v^0)e^a_\mu \partial f/\partial x^a - (\gamma^i_{\mu\nu} v^\mu v^\nu/v^0)\partial f/\partial v^i = 0 \tag{1.2}$$

The convention is used that frame indices are denoted by letters from the middle of the alphabet, while coordinate indices are taken from the beginning of the alphabet. The components of the frame vectors are denoted by e^α_μ and p^α and v^μ are related by the equation $p^\alpha = e^\alpha_\mu v^\mu$. Since the frame is orthonormal $v^0 = \sqrt{1 + \delta^{ij}v_i v_j}$, where δ^{ij} is the Kronecker delta.

The Vlasov equation can be coupled to the Einstein equations as follows, giving rise to the Einstein–Vlasov system. The unknowns are a 4-manifold M, a (time orientable) Lorentz metric $g_{\alpha\beta}$ on M and a non-negative real-valued function f on the mass shell defined by $g_{\alpha\beta}$. The field equations consist of the Vlasov equation defined by the metric $g_{\alpha\beta}$ for f and the Einstein equation $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$. (Units are chosen here so that the speed of light and the gravitational constant both have the numerical value unity.) To obtain a complete system of equations it remains to define $T_{\alpha\beta}$ in terms of f and $g_{\alpha\beta}$. It is defined as an integral over the part of the mass shell over a given spacetime point with respect to a measure which will now be defined. The metric at a given point of spacetime defines in a tautological way a metric on the tangent space at that point. The part of the mass shell over that point is a submanifold of the tangent space and as such has an induced metric, which is Riemannian. The associated measure is the one which we are seeking. It is evidently invariant under Lorentz transformations of the tangent space, a fact which may be used to simplify computations in concrete situations. In the coordinates (x^α, p^a) on P the explicit form of the energy-momentum tensor is:

$$T_{\alpha\beta} = -\int f p_\alpha p_\beta |g|^{1/2}/p_0 dp^1 dp^2 dp^3 \tag{1.3}$$

A simple computation in normal coordinates based at a given point shows that $T_{\alpha\beta}$ defined by (1.3) is divergence-free, independently of the Einstein equations being satisfied. This is of course a necessary compatibility condition in order for the Einstein–Vlasov system to be a reasonable set of equations. Another important quantity is the particle current density, defined by:

$$N^\alpha = -\int f p^\alpha |g|^{1/2}/p_0 dp^1 dp^2 dp^3 \tag{1.4}$$

A computation in normal coordinates shows that $\nabla_\alpha N^\alpha = 0$. This equation is an expression of the conservation of the number of particles. There are some inequalities which follow immediately from the definitions (1.3) and (1.4). Firstly $N_\alpha V^\alpha \leq 0$ for any future-pointing timelike or null vector V^α, with equality only if $f = 0$ at the given point. Hence unless there are no particles at some point, the vector N^α is future-pointing timelike. Next, if V^α and W^α are any two future-pointing timelike vectors then $T_{\alpha\beta}V^\alpha W^\beta \geq 0$. This is the dominant energy condition ([12], p. 91). Finally, if X^α is a spacelike vector then $T_{\alpha\beta}X^\alpha X^\beta \geq 0$. This is the non-negative pressures condition. This condition, the dominant energy condition and the Einstein equations together imply that the Ricci tensor satisfies the inequality $R_{\alpha\beta}V^\alpha V^\beta \geq 0$ for any timelike vector V^α. The last inequality is called the strong energy condition. These inequalities constitute one of the reasons which mean that the Vlasov equation defines a well-behaved matter model in general relativity. However this is not the only reason. A perfect fluid with a reasonable equation of state or matter described by the Boltzmann equation also have energy-momentum tensors which satisfy these inequalities.

The Vlasov equation in a fixed spacetime is a linear hyperbolic equation for a scalar function and hence solving it is equivalent to solving the equations for its characteristics. In coordinate components these are:

$$\begin{aligned} dX^a/ds &= P^a \\ dP^a/ds &= -\Gamma^a_{\beta\gamma}P^\beta P^\gamma/P^0 \end{aligned} \tag{1.5}$$

Let $X^a(s, x^\alpha, p^a)$, $P^a(s, x^\alpha, p^a)$ be the unique solution of (1.5) with initial conditions $X^a(t, x^\alpha, p^a) = x^a$ and $P^a(t, x^\alpha, p^a) = p^a$. Then the solution of the Vlasov equation can be written as:

$$f(x^\alpha, p^a) = f_0(X^a(0, x^\alpha, p^a), P^a(0, x^\alpha, p^a)) \tag{1.6}$$

where f_0 is the restriction of f to the hypersurface $t = 0$. This function f_0 serves as initial datum for the Vlasov equation. It follows immediately from this that if f_0 is bounded by some constant C, the same is true of f. This obvious but important property of the solutions of the Vlasov equation is used frequently without comment in the study of this equation.

The above calculations involving $T_{\alpha\beta}$ and N^α were only formal. In order that they have a precise meaning it is necessary to impose some fall-off in the momentum variables on f so that the integrals occurring exist. The simplest condition to impose is that f has compact support for each fixed t. This property holds if the initial datum f_0 has compact support and if each hypersurface $t = t_0$ is a Cauchy hypersurface. For by the definition of a Cauchy hypersurface, each timelike curve which starts at $t = 0$ hits the hypersurface $t = t_0$ at a unique point. Hence the geodesic flow defines a continuous mapping from the part of the mass shell over the initial hypersurface $t = 0$ to the part over the hypersurface $t = t_0$. The support of $f(t_0)$, the restriction of f to the hypersurface $t = t_0$ is the image of the support of f_0 under this continuous mapping and so is compact. Let $P(t)$ be the supremum of the values of $|p^a|$ attained on the support of $f(t)$. It turns out that in many cases controlling the solution of the Vlasov equation coupled to some field equation in the case of compactly supported initial data for the

distribution function can be reduced to obtaining a bound for $P(t)$. An example of this is given below.

The data in the Cauchy problem for the Einstein equations coupled to any matter source consist of the induced metric g_{ab} on the initial hypersurface, the second fundamental form k_{ab} of this hypersurface and some matter data. In fact these objects should be thought of as objects on an abstract 3-dimensional manifold S. Thus the data consist of a Riemannian metric g_{ab}, a symmetric tensor k_{ab} and appropriate matter data, all defined intrinsically on S. The nature of the initial data for the matter will now be examined in the case of the Einstein–Vlasov system. It is not quite obvious what to do, since the distribution function f is defined on the mass shell and so the obvious choice of initial data, namely the restriction of f to the initial hypersurface, is not appropriate. For it is defined on the part of the mass shell over the initial hypersurface and this is not intrinsic to S. This difficulty can be overcome as follows. Let ϕ be the mapping which sends a point of the mass shell over the initial hypersurface to its orthogonal projection onto the tangent space to the initial hypersurface. The map ϕ is a diffeomorphism. The abstract initial datum f_0 for f is taken to be a function on the tangent bundle of S. The initial condition imposed is that the restriction of f to the part of the mass shell over the initial hypersurface should be equal to f_0 composed with ϕ. An initial data set for the Einstein equations must satisfy the constraints and in order that the definition of an abstract initial data set for the Einstein equations be adequate it is necessary that the constraints be expressible purely in terms of the abstract initial data. The constraint equations are:

$$R - k_{ab}k^{ab} + (\mathrm{tr}k)^2 = 16\pi\rho \tag{1.7}$$

$$\nabla_a k^a_b - \nabla_b(\mathrm{tr}k) = 8\pi j_b \tag{1.8}$$

Here R denotes the scalar curvature of the metric g_{ab}. If n^α denotes the future-pointing unit normal vector to the initial hypersurface and $h^{\alpha\beta} = g^{\alpha\beta} + n^\alpha n^\beta$ is the orthogonal projection onto the tangent space to the initial hypersurface then $\rho = T_{\alpha\beta}n^\alpha n^\beta$ and $j^\alpha = -h^{\alpha\beta}T_{\beta\gamma}n^\gamma$. The vector j^α satisfies $j^\alpha n_\alpha = 0$ and so can be naturally identified with a vector intrinsic to the initial hypersurface, denoted here by j^a. What needs to be done is to express ρ and j_a in terms of the intrinsic initial data. They are given by the following expressions:

$$\rho = \int f_0(p^a)p^a p_a/(1 + p^a p_a)^{1/2}({}^{(3)}g)^{1/2}dp^1dp^2dp^3 \tag{1.9}$$

$$j_a = \int f_0(p^a)p_a({}^{(3)}g)^{1/2}dp^1dp^2dp^3 \tag{1.10}$$

If a three-dimensional manifold on which an initial data set for the Einstein–Vlasov system is defined is mapped into a spacetime by an embedding ψ then the embedding is said to induce the given initial data on S if the induced metric and second fundamental form of $\psi(S)$ coincide with the results of transporting g_{ab} and k_{ab} with ψ and the relation $f = f_0 \circ \phi$ holds, as above. A form of the local existence and uniqueness theorem can now be stated. This will only be done for the case of smooth (i.e. infinitely differentiable) initial data although versions of the theorem exist for data of finite differentiability.

THEOREM 1.1. *Let S be a 3-dimensional manifold, g_{ab} a smooth Riemannian metric on S, k_{ab} a smooth symmetric tensor on S and f_0 a smooth non-negative function of compact support on the tangent bundle TS of S. Suppose further that these objects satisfy the constraint equations (1.7)-(1.8). Then there exists a smooth spacetime $(M, g_{\alpha\beta})$, a smooth distribution function f on the mass shell of this spacetime and a smooth embedding ψ of S into M which induces the given initial data on S such that $g_{\alpha\beta}$ and f satisfy the Einstein–Vlasov system and $\psi(S)$ is a Cauchy hypersurface. Moreover, given any other spacetime $(M', g'_{\alpha\beta})$, distribution function f' and embedding ψ' satisfying these conditions, there exists a diffeomorphism χ from an open neighbourhood of $\psi(S)$ in M to an open neighbourhood of $\psi'(S)$ in M' which satisfies $\chi \circ \psi = \psi'$ and carries $g_{\alpha\beta}$ and f to $g'_{\alpha\beta}$ and f' respectively.*

The formal statement of this theorem is rather complicated, but its essential meaning is as follows. Given an initial data set (satisfying the constraints) there exists a corresponding solution of the Einstein–Vlasov system and this solution is locally unique up to diffeomorphism. There also exists a global uniqueness statement which uses the notion of the maximal Cauchy development of an initial data set, but this is not required in the following. The first proof of a theorem of this kind for the Einstein–Vlasov system is due to Choquet-Bruhat [7].

In the following we are mainly concerned with asymptotically flat spacetimes. These are the spacetimes which are appropriate for describing isolated systems in general relativity. It is assumed that these spacetimes admit a Cauchy hypersurface with topology $\mathbf{R}^3$, although more general cases could also be considered. The smooth data set (g_{ab}, k_{ab}, f_0) on $\mathbf{R}^3$ is said to be asymptotically flat if there exist global coordinates x^a such that as $|x|$ tends to infinity the components g_{ab} in these coordinates tend to δ_{ab}, the components k_{ab} tend to zero, f_0 has compact support and certain norms are finite. These are weighted Sobolev norms. If u is a smooth function on $\mathbf{R}^3$ define

$$\|u\|_{H^s_\delta} = \left[\sum_{i=0}^{s} \int (1 + |x|^2)^{(\delta+i)} |D^i u|^2 dx\right]^{1/2} \tag{1.11}$$

where $|D^i u|$ denotes the maximum modulus of any partial derivative of order i of the function u. If the quantity (1.11) is finite then u is said to belong to the weighted Sobolev space H^s_δ. Assume for asymptotic flatness that $g_{ab} - \delta_{ab} \in H^s_\delta$ and that $k_{ab} \in H^{s-1}_{\delta+1}$ for s sufficiently large and $-3/2 < \delta < -1/2$. If a spacetime is asymptotically flat then Theorem 1.1 can be sharpened to say that there exists a local solution corresponding to the given initial data and coordinates defined for that solution such that the solution exists on $\mathbf{R}^3 \times [0, T)$ for some $T > 0$ and the data induced on the hypersurfaces $t =$const. satisfy the same type of asymptotic flatness conditions as the initial data on the hypersurface $t = 0$. More precisely, there exist coordinates so that the data induced by the solution on each hypersurface of constant time belongs to the same weighted Sobolev space as the initial data. Moreover the restrictions of the functions $g_{00} + 1$ and g_{0a} of the metric components to any hypersurface of constant time belong to weighted Sobolev spaces. A proof of this is sketched in Section 4.

The local existence theorem can be supplemented by a statement that the solution depends continuously on the data. This is not stated in full generality here; only some

statements for asymptotically flat data are given. Given a family of initial data which is bounded in a weighted Sobolev space with s sufficiently large and is such that the metric is uniformly positive definite, the coordinates above can be chosen so that the solutions corresponding to all data in the family exist on the same time interval $[0, T)$, the weighted Sobolev norm of the data induced by these solutions on a hypersurface of constant time is uniformly bounded and the induced metric on one of these hypersurfaces is uniformly positive definite. Moreover, the restrictions of $g_{00} + 1$ and g_{0a} to each hypersurface of constant time are uniformly bounded in a weighted Sobolev space for all data in the family and g_{00} is uniformly bounded away from zero. See Section 4.

Suppose now that an asymptotically flat initial data set admits a group G of symmetries, i.e. that a Lie group G acts on the manifold S in such a way that g_{ab}, k_{ab} and f_0 are preserved. Then the spacetime in Theorem 1.1 can be chosen so that it admits G as a symmetry group. More precisely, there exists an action of G on M by isometries which preserves f and which restricts to the original action on S. To prove this, consider the geodesic $\gamma(p)$ through a point $p \in S$ orthogonal to S in a spacetime with the given initial data. Let $t(p)$ denote the largest number such that, when $\gamma(p)$ is parametrized by proper time, with p corresponding to the parameter value zero, this geodesic is defined on the interval $(-t(p), t(p))$. By what has been said above, the spacetime can be chosen so that there exists a number $T > 0$ which is smaller than $t(p)$ for each $p \in S$. Since the spacetime is globally hyperbolic, each point can be joined to the initial hypersurface by a timelike geodesic of maximal length. If T is chosen sufficiently small then the geodesic is unique. A new spacetime can be defined as the open subset of the original spacetime where this distance is less than T. It will now be shown that the action of G on S extends to an action on this new spacetime, whose underlying manifold will be denoted by M. Given a point $q \in M$, let p be the point where the unique geodesic γ of maximal length from q to the initial hypersurface S meets S. Let $\phi : G \times S \to S$ denote the action of G on S. For $g \in G$ and $q \in M$, let $\tilde{\phi}_g(q)$ be the point which lies the same distance to the future of $\phi_g(p)$ along the future-directed geodesic starting orthogonal to S at $\phi_g(p)$ as q lies to the future of p along $\gamma(p)$. This defines a mapping $\tilde{\phi} : G \times M \to M$ by $\tilde{\phi}(g, q) = \tilde{\phi}_g(q)$. This mapping $\tilde{\phi}$ is an action of G on M which restricts to ϕ. By the uniqueness part of Theorem 1.1, it must preserve $g_{\alpha\beta}$ and f.

The global theorems to be proved later make use of the concept of maximal hypersurfaces. A spacelike hypersurface in a spacetime is called maximal if its mean curvature $\mathrm{tr}k$ is zero. If an initial data set is given which is maximal in this sense and asymptotically flat it is of interest to know whether the corresponding local solution of the Einstein–Vlasov system, whose existence is guaranteed by the above theorem, can be foliated by maximal hypersurfaces in a neighbourhood of the initial hypersurface. Given what has already been said about the local existence of asymptotically flat spacetimes, this can be proved using the implicit function theorem (cf. [16]). The time coordinate t above can be chosen so that its level hypersurfaces are maximal hypersurfaces and it can also be arranged that as $|x| \to \infty$ the coordinate t agrees asymptotically with proper time along a geodesic which starts normal to the initial hypersurface. (This implies that g_{00} tends to unity as $|x|$ tends to infinity.) When this has been imposed the foliation by maximal hypersurfaces is unique and so this construction gives a unique preferred time coordinate. The restriction

of the solution to any of the maximal hypersurfaces is asymptotically flat. Once again there is a statement of uniformity. For a family of maximal initial data which is bounded in a suitable weighted Sobolev space the maximal foliation can be assumed to exist on a time interval which is uniform for all data in the family. These statements about the existence of maximal foliations are not too dependent on the fact that the matter is described by the Vlasov equation. The essential property which is needed for the existence and uniqueness theorems is the strong energy condition.

The approach to studying the global structure of asymptotically flat solutions of the Einstein–Vlasov system presented in these lectures is closely related to work which has been done in the spatially compact case [25, 4]. The main difference is that in a non-flat spacetime satisfying the strong energy condition with a compact Cauchy hypersurface there exists at most one maximal hypersurface [16]. In this case the maximal foliation of the asymptotically flat case can be replaced by a constant mean curvature (CMC) foliation. This means that each leaf of the foliation has constant mean curvature, while the value of the mean curvature varies monotonically from one leaf to the next.

2. Spherical symmetry. Investigating the global properties of general solutions of the Einstein–Vlasov system is beyond the scope of existing mathematical techniques. For comparison, note that the same comment applies to the Vlasov–Maxwell system (cf. [19] for the most general known results) while general global existence results have been obtained for the simpler Vlasov–Poisson system ([17], [13], [21]). When a system of partial differential equations appears inaccessible to direct attack, a natural strategy is to study the simpler equations obtained by imposing symmetry conditions on the solutions of the original equations. In the case of asymptotically flat solutions of the Einstein–Vlasov system, it seems that there are only two possible symmetry assumptions: spherical and rotational symmetry. In the latter case, where the symmetry group is one-dimensional and has fixed points (on the axis of rotation), the simplification obtained is not sufficient to bring the problem within range of present techniques. Thus in the following treatment of global questions we consider only the spherically symmetric case. Note for comparison that in the spatially compact case a wider variety of tractable symmetry types exists.

A solution $(M, g_{\alpha\beta}, f)$ of the Einstein–Vlasov system is said to be spherically symmetric if there exists an action of $SO(3)$ on M by isometries whose generic orbits are two-dimensional such that the natural lift of this action to the mass shell preserves f. There is of course an analogous definition of a spherically symmetric initial data set. Consider now a spherically symmetric asymptotically flat maximal initial data set. From the last section we know that there exists a corresponding local solution of the Einstein–Vlasov system. Moreover, it can be assumed without loss of generality that $SO(3)$ acts on this local solution as a symmetry group so that it is spherically symmetric. Furthermore, there exists a neighbourhood U of the initial hypersurface which can be foliated by maximal hypersurfaces whose intrinsic geometry is asymptotically flat and this foliation is unique. It follows from the latter fact that each maximal hypersurface is invariant under the action of $SO(3)$. In other words, it is a union of orbits of the action of $SO(3)$ on spacetime. Let r, the area radius, be defined by the condition that on any orbit it takes the constant value $\sqrt{A/4\pi}$, where A is the area of the given orbit. Consider now a fixed

spacelike hypersurface S_0 which is invariant under the action of $SO(3)$. A geodesic of the induced metric on S_0 which starts orthogonal to the orbits remains orthogonal to them. These geodesics will be called radial geodesics. Let θ, ϕ be standard spherical coordinates on one of the orbits. Extend them to be constant along the radial geodesics. Since radial geodesics can never intersect except at the centre this prescription is globally well defined. In the coordinates (r, θ, ϕ) the metric components g_{12} and g_{13} vanish. The metric intrinsic to the orbits takes the standard form $r^2(d\theta^2 + \sin^2\theta d\phi^2)$. Hence, provided the gradient of r does not vanish anywhere, the induced metric on S_0 can be written in the form

$$e^{2\lambda(r)}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.1}$$

This way of defining spatial coordinates was used in [20]. It cannot be used if the restriction of r to the hypersurfaces of constant time has a vanishing gradient somewhere. A point where the gradient vanishes corresponds to a minimal surface. Here we use another type of coordinate system which does not suffer from this difficulty. First note that it is possible to write the metric in the form

$$B^2(x)dx^2 + r^2(x)(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.2}$$

without restriction, for some function B. Since there is always a neighbourhood of the origin without minimal surfaces it can be assumed without loss of generality that $r(x) = x$ near the origin. Since B is smooth when considered as a function on spacetime, it must be a smooth function of x^2. Furthermore, in order that the spacetime be regular at $x = 0$ and not have a conical singularity, $B(0)$ must be equal to unity. This means in particular that it is possible to write $B(x) = 1 + x^2D(x)$, where $D(x)$ is a smooth function of x^2. A new coordinate R will be sought which is a function of x and has the property that, when expressed in terms of the coordinate R, the metric takes the form:

$$A^2(R)(dR^2 + R^2(d\theta^2 + \sin^2\theta d\phi^2)) \tag{2.3}$$

Coordinates of this sort, or more precisely the Cartesian coordinates corresponding to these polar coordinates, are often known as isotropic coordinates. Writing down the coordinate transformation shows that $R(x)$ is a solution of the equation $dR/dx = RB(x)/r(x)$. This equation has a solution which is unique up to a constant scaling. To see this, note first that this is a first order homogeneous linear ordinary differential equation. Hence it has a one-parameter family of solutions which can all be got by multiplying one particular solution by an arbitrary constant. If we know the existence of the solution near the origin then the existence for all values of R follows by the standard existence and uniqueness theorem for ordinary differential equations. Near the origin the equation can be solved, giving $R = Cx\exp\int_0^x x'D(x')dx'$. Near infinity, the asymptotic flatness of the metric implies that A tends to a constant value. The scaling can be chosen so that $\lim_{R\to\infty} A(R) = 1$ and then the solution is determined uniquely. Now let t be a time coordinate which is constant on each leaf of the preferred foliation by maximal hypersurfaces and which agrees asymptotically with proper time. Introducing a coordinate R as above on each leaf puts the metric in the form:

$$-\alpha^2(t, R)dt^2 + A^2(t, R)[(dR + \beta(t, R)dt)^2 + R^2(d\theta^2 + \sin^2\theta d\phi^2)] \tag{2.4}$$

As a consequence of the choice of coordinates the functions A and α tend to unity as $R \to \infty$ for each fixed t while $\beta \to 0$. The smoothness of the spacetime metric together with spherical symmetry implies that α, A, and $R^{-1}\beta$ are smooth functions of R^2.

For a metric of the form (2.4) which satisfies the extra condition that the hypersurfaces of constant time are maximal hypersurfaces the field equations and coordinate conditions take the following form:

$$(R^2(A^{1/2})')' = -\tfrac{1}{8}A^{5/2}R^2(\tfrac{3}{2}K^2 + 16\pi\rho) \tag{2.5}$$

$$\alpha'' + 2\alpha'/R + A^{-1}A'\alpha' = \alpha A^2[\tfrac{3}{2}K^2 + 4\pi(\rho + \mathrm{tr}S)] \tag{2.6}$$

$$K' + 3(A^{-1}A' + 1/R)K = 8\pi Aj \tag{2.7}$$

$$\beta' - R^{-1}\beta = \tfrac{3}{2}\alpha K \tag{2.8}$$

$$\partial_t A = -\alpha K A + (\beta A)' \tag{2.9}$$

$$\partial_t K = -A^{-2}\alpha'' + A^{-3}A'\alpha' + \alpha[-2A^{-3}A'' + 2A^{-4}A'^2 - 2A^{-3}A'/R - 8\pi S_R + 4\pi \mathrm{tr}S - 4\pi\rho] + \beta K' \tag{2.10}$$

The notation used in these equations will now be explained. A prime denotes a derivative with respect to R. The quantity K is that obtained by contracting the second fundamental form of the hypersurface t =const. twice with the unit vector $A^{-1}\partial/\partial R$ while S_R is obtained in the corresponding way from the energy-momentum tensor. The quantity $\mathrm{tr}S$ is the trace of the spatial part of the energy-momentum tensor, i.e. if n^α is the unit future-pointing normal vector to the hypersurfaces of constant time, then $\mathrm{tr}S = T_{\alpha\beta}(g^{\alpha\beta} + n^\alpha n^\beta)$. The quantity j is obtained by contracting $T_{\alpha\beta}$ once with n^α and once with the vector $A^{-1}\partial/\partial R$ and ρ is the energy density $T_{\alpha\beta}n^\alpha n^\beta$. In the standard terminology of the $3+1$-decomposition of Einstein's equations, α is the lapse function and β is the one non-trivial component of the shift vector. Equation (2.5) is the explicit form of the Hamiltonian constraint (1.7) in this class of spacetimes with this kind of coordinate condition. The maximal slicing condition is expressed by the lapse equation (2.6). The one non-trivial component of the momentum constraint (1.8) in these spacetimes is (2.7). Equation (2.8) is a consequence of the coordinate condition chosen while (2.9) follows from the definition of the second fundamental form. Finally, (2.10) is the one non-trivial Einstein evolution equation in this class of spacetimes. This form of the field equations has been used by Shapiro and Teukolsky for numerical calculations (see [27]).

What has been shown above implies that given asymptotically flat spherically symmetric maximal initial data for the Einstein–Vlasov system, there exists a corresponding local solution and some $T_1 > 0$ such that this spacetime can be covered by coordinates which cast it in the form (2.4) and for which the time coordinate ranges in the interval $(-T_1, T_1)$ and the initial hypersurface is given by $t = 0$. In fact we are only interested in evolution to the future and hence only consider the part of spacetime on the interval $[0, T_1)$. The quantities describing the metric and matter then satisfy the equations (2.5)-(2.10). The Vlasov equation is of course also satisfied. It turns out to be useful for some purposes to write it in Cartesian coordinates. Let x^a be the coordinates $(R\sin\theta\cos\phi, R\sin\theta\sin\phi, R\cos\theta)$. Define a related orthonormal frame by $e_i = A^{-1}\partial/\partial x^i$. Then if the mass shell is coordinatized by (t, x^a, v^i) , where v^i denote the components of a vector in the given orthonormal

frame the Vlasov equation takes the explicit form:

$$\frac{\partial f}{\partial t} + \left(\alpha A^{-1} \frac{v}{\sqrt{1+|v|^2}} - \beta \frac{x}{R} \right) \cdot \frac{\partial f}{\partial x} + \left[-A^{-1} \alpha' \sqrt{1+|v|^2} \frac{x}{R} - \frac{1}{2} \alpha K \left(v - 3 v_r \frac{x}{R} \right) \right.$$
$$\left. -\alpha A^{-2} A' \left(v v_r - |v|^2 \frac{x}{R} \right) \frac{1}{\sqrt{1+|v|^2}} \right] \cdot \frac{\partial f}{\partial v} = 0 \tag{2.11}$$

Here a dot denotes the usual inner product in $\mathbf{R}^3$, $|v| = \sqrt{v \cdot v}$ and $v_r = (v \cdot x)/R$. For given initial data there exists a solution on an interval $[0, T_1)$ of the equations (2.5)-(2.11) supplemented by the definitions of the matter quantities. The whole system of equations will be referred to as the 'reduced Einstein–Vlasov system'. This solution of the reduced system is uniquely determined by its restriction to $t = 0$, as follows easily from the general uniqueness theorem for solutions of the Einstein–Vlasov system and the uniqueness of the coordinate system used. As a consequence there exists a greatest value of T_1 (finite or infinite, call it T_*) for which a solution of the reduced equations with the given initial data exists on the time interval $[0, T_1)$. The interval $[0, T_*)$ is called the maximal interval of existence. The global existence question, which is the main theme of these lectures, is the question under what circumstances $T_* = \infty$.

One possible strategy for proving global existence theorems will now be outlined. Suppose that in some way it were possible to show for given initial data that for any corresponding solution on a finite interval $[0, T_1)$ the metric components, the distribution function and all their derivatives of all orders with respect to t and x were bounded. Then global existence for these initial data would follow. For the metric components, the distribution function and all their derivatives of all orders would be uniformly continuous. By a standard theorem on metric spaces they would all extend to continuous functions on the closed interval $[0, T_1]$. By another standard theorem, this time of real analysis, the extensions are C^∞ and each derivative of each extension is equal to the extension of the corresponding derivative. In this way smooth initial data are defined on the hypersurface $t = T_1$. Provided these new initial data are asymptotically flat, the local existence theorem can be applied again to show that the original solution has an extension to an interval $[0, T_2)$ with $T_2 > T_1$. Hence $T_1 \neq T_*$. But since T_1 was an arbitrary positive number this only leaves the possibility that $T_* = \infty$ and global existence is proved. In the following a situation is exhibited where bounds similar to those which are assumed in this argument can actually be obtained.

To get closer to the situation which has just been described, consider a solution of the reduced equations defined on some interval $[0, T_1)$. It will now be shown that many quantities can be bounded by using the Einstein–Vlasov system. By the dominant energy condition $\rho \geq 0$. Hence equation (2.5) shows that $R^2(A^{1/2})'$ is a non-increasing function of R for each fixed t. When $R = 0$ it is zero and hence $R^2(A^{1/2})' \leq 0$. It follows that $A' \leq 0$. The boundary condition that $A \to 1$ as $R \to \infty$ then gives $A \geq 1$. Next, from (2.6),

$$(R^2 A \alpha')' = \alpha A^3 R^2 [\tfrac{3}{2} K^2 + 4\pi(\rho + \mathrm{tr}S)] \geq 0 \tag{2.12}$$

The expression on the right hand side of (2.12) is non-negative, as follows from the dominant energy and non-negative pressures conditions. Since $R^2 A \alpha'$ vanishes for $R = 0$

it can be seen that $R^2 A\alpha' \geq 0$ and $\alpha' \geq 0$. Using the boundary condition that $\alpha \to 1$ as $R \to \infty$ gives $\alpha \leq 1$.

Next an estimate of Malec and Ó Murchadha [15] will be used. The expansions of the null geodesics which start normal to the orbits are given by

$$\theta = 2(A^{-2}A' + (AR)^{-1}) + K \quad \theta' = 2(A^{-2}A' + (AR)^{-1}) - K \tag{2.13}$$

The area radius is given by $r = AR$. The theorem of [15] states that, if the dominant energy condition holds, the quantities $r\theta$ and $r\theta'$ are bounded in modulus by two. Adding and subtracting these estimates and using the explicit expressions in (2.13) gives the inequalities $RA|K| \leq 2$ and $|RA^{-1}A' + 1| \leq 1$. Since $A \geq 1$ it can be concluded from the first of these inequalities that $|K| \leq 2R^{-1}$. The second inequality gives $|A^{-1}A'| \leq 2R^{-1}$. In particular this gives pointwise bounds for K and $A^{-1}A'$ away from the centre. Integrating (2.8) gives

$$R_2^{-1}\beta(R_2) - R_1^{-1}\beta(R_1) = \tfrac{3}{2}\int_{R_1}^{R_2} (\alpha K/s) ds \tag{2.14}$$

Asymptotic flatness implies that $K = O(R^{-1})$ as $R \to \infty$. Hence it is possible to let R_2 tend to infinity in this equation to get an expression for $R^{-1}\beta(R)$ as an integral from R to ∞. Using the bounds already obtained for K and α shows that

$$|\beta| \leq \tfrac{3}{2} R \int_R^\infty (2/s^2) ds \leq 3 \tag{2.15}$$

and so β is bounded. A bound for α' can be obtained by analysing the inequality $(R^2 A\alpha')' \geq 0$, which was already used to show that $\alpha \leq 1$. Integrating this between the radii R_1 and R_2 with $R_1 < R_2$ gives:

$$\alpha'(R_2) \geq (R_1/R_2)^2 (A(R_1)/A(R_2))\alpha'(R_1) \tag{2.16}$$

Integrating again then gives:

$$\alpha(R_2) \geq \alpha'(R_1)(R_1^2 A(R_1)) \int_{R_1}^{R_2} R^{-2}(A(R))^{-1} dR \tag{2.17}$$

Now use the facts that $\alpha(R_2) \leq 1$ and that $A(R) \leq A(R_1)$ for $R \geq R_1$ to see that:

$$1 \geq \alpha'(R_1) R_1^2 \int_{R_1}^{R_2} R^{-2} dR \tag{2.18}$$

This holds for all $R_2 \geq R_1$ and so it is permissible to replace the upper limit in the integral by infinity. Evaluating the integral gives $\alpha'(R) \leq R^{-1}$ for all $R > 0$. Note that in deriving all these estimates, the only properties of the matter fields used were the dominant energy condition and the inequality $\rho + \mathrm{tr}S \geq 0$. The latter follows from the strong energy condition. Thus all these estimates hold not only for the Einstein–Vlasov system, but also for the Einstein equations coupled to any matter model which satisfies the dominant and strong energy conditions. This includes perfect fluids with reasonable equations of state, matter described by the Boltzmann equation, the massless scalar field (or more generally wave maps), any of these matter models combined with an electromagnetic field in such a way that the total energy-momentum tensor is the sum of the individual energy-momentum tensors, and the Yang-Mills equations for any semi-simple gauge group.

These estimates give good information on the solution away from the centre but, except in the case of the estimate for β, give no control at the centre. Pointwise estimates for α', A' and K which also give useful information at the centre can be obtained by optimization arguments. For any fixed radius R_0 we have

$$|K(R)| \leq 4\pi A^4(0)\|\rho\|_\infty R_0 \tag{2.19}$$

for any $R \leq R_0$. Here $\| \ \|_\infty$ denotes the L^∞ norm in space, i.e. the maximum value of a function on a hypersurface of constant time. This inequality is obtained by integrating equation (2.7) and using the dominant energy condition to bound the modulus of j by ρ. On the other hand, if $R \geq R_0$ then $|K(R)| \leq 2R_0^{-1}$. Thus for any value of R it is true that

$$|K(R)| \leq 4\pi A^4(0)\|\rho\|_\infty R_0 + 2R_0^{-1} \tag{2.20}$$

This can be optimized by choosing R_0 so that the function of R_0 occurring on the right hand side of this last inequality has a critical point. This occurs when R_0 is equal to $[2\pi A^4(0)\|\rho\|_\infty]^{-1/2}$. It follows that $\|K\|_\infty \leq CA^2(0)\|\rho\|_\infty^{1/2}$ for some constant C. A similar procedure can be used to estimate α'.

$$\alpha'(R) = A^{-1}R^{-2}\int_0^R \alpha A^3 s^2(\tfrac{3}{2}K^2 + 4\pi\rho + 4\pi \mathrm{tr}S)ds \leq CA^7(0)\|\rho\|_\infty R_0 \tag{2.21}$$

for $R \leq R_0$. Combining this with the previous pointwise estimate for α' gives

$$\alpha' \leq C(A^7(0)\|\rho\|_\infty R_0 + R_0^{-1}) \tag{2.22}$$

Doing an optimization as above leads to an estimate of the form $\alpha' \leq CA^{7/2}(0)\|\rho\|_\infty^{1/2}$. From the equation for A:

$$A'(R) = \tfrac{1}{4}R^{-2}A^{1/2}\int_0^R A^{5/2}s^2(\tfrac{3}{2}K^2 + 16\pi\rho)ds \tag{2.23}$$

Thus

$$|A'| \leq CA(0)[\tfrac{1}{4}A^6(0)\|\rho\|_\infty R_0 + 2R_0^{-1}] \tag{2.24}$$

Optimizing gives $\|A'\|_\infty \leq CA^4(0)\|\rho\|_\infty^{1/2}$. Starting from the same equations we can also bound $R^{-1}K$, $R^{-1}\alpha'$ and $R^{-1}A'$ pointwise in terms of $\|\rho\|_\infty$ and $A(0)$. In particular $R^{-1}K$ can be bounded by a constant multiple of $A^4(0)\|\rho\|$. Equations (2.6) and (2.7) then allow α'' and K' to be bounded. Solving (2.5) for A'' shows that it too can be bounded. Equation (2.14) and the bounds for β and $R^{-1}K$ imply that $R^{-1}\beta$ can be bounded by a constant multiple of $A^2(0)\|\rho\|_\infty^{1/2}$. Using equation (2.8) then gives a similar bound for β'.

In the proof of variants of these estimates discussed in the next section, the conservation of the total (ADM) mass plays a role and it is convenient to say something about this conservation law at this point. It is a quite general property of asymptotically flat spacetimes. It is particularly easy to see in the situation considered here, where outside a compact set the spacetime is vacuum and spherically symmetric. Equation (2.7) can be rearranged to give $(A^3R^3K)' = 8\pi R^3A^4j$. In vacuum this integrates to give $K = K_0(t)R^{-3}A^{-3}$ for some function $K_0(t)$. Putting this in (2.5) and using the vacuum condition again gives $(R^2(A^{1/2})')' = O(R^{-4})$. This can be integrated to give:

$$A(t,R) = (1 + A_0(t)R^{-1})^2 + O(R^{-4}) \tag{2.25}$$

It follows from (2.8) that $(R^{-1}\beta)' = \frac{3}{2}R^{-1}\alpha K = O(R^{-4})$. Hence $\beta = O(R^{-2})$ and $\beta' = O(R^{-3})$. It then follows from (2.9) that $\partial_t A = O(R^{-3})$. Integrating this last equation in time from 0 to t shows that $A(t,R) = A(0,R) + O(R^{-3})$, so that $A_0(t)$ is in fact independent of t. The ADM mass is given by $\lim_{R\to\infty}(-R^2A')$ and so is also time independent. From (2.5) we can calculate that the ADM mass is equal to

$$m_{ADM} = \tfrac{1}{8}\int_0^\infty A^{5/2}R^2(\tfrac{3}{2}K^2 + 16\pi\rho)dR \tag{2.26}$$

In all these estimates only the dominant and strong energy condtions have been used.

To go beyond the results of the last paragraphs it is necessary to use the specific nature of the matter model. For the Einstein–Vlasov system a continuation criterion will be proved. It is formulated in terms of a quantity $P(t)$, which is defined to be the largest momentum of any particle at time t. In other words

$$P(t) = \sup\{|v| : f(t,x,v) \neq 0 \text{ for some } x\} \tag{2.27}$$

Before stating the continuation criterion it is necessary to take some time to discuss the relation between the differentiability of the functions of R describing the spacetime and the differentiability of the corresponding objects in spacetime. The simplest case is that of the scalar functions α and A. They are C^∞ in the spacetime sense if and only if they are C^∞ as functions of R and all the derivatives of odd order vanish at the origin. This follows from Lemma A1 of the appendix with $m = 0$. There is also a quantitative version of this, which follows from Lemma A2. Consider next β. By definition $\beta = \beta^a x_a/R$ and $\beta^a = \beta x^a/R$, where β^a is the shift vector. Because of spherical symmetry β must vanish at the origin. Hence we can apply Lemma A3 to β^a. Lemma A4 gives quantitative results for β^a. Consider next K. By definition $A^2K = k_{ab}x^a x^b/R^2$. The maximal hypersurface condition and spherical symmetry together imply that k_{ab} vanishes at the origin. Hence Lemma A2 can be applied with $m = 2$. Similar considerations apply to the matter quantities j and S_R, whereby in the latter case it is necessary to write $S_R = \tilde{S}_R + \frac{1}{3}\mathrm{tr}S$, with $\tilde{S}_R$ being the contribution to S_R of the trace free part of T_{ab}. Because of spherical symmetry $\tilde{S}_R$ vanishes at $R = 0$. The following expressions for some of the quantities occurring in the Vlasov equation are also significant:

$$\begin{aligned} &\beta x_a/R = \beta_a \qquad k_{ab} = -\tfrac{1}{2}KA^2(\delta_{ab} - 3x_a x_b/R^2) \\ &\alpha' x_a/R = \nabla_a\alpha \\ &A' x_a/R = \nabla_a A \end{aligned} \tag{2.28}$$

THEOREM 2.1. *If a solution of the reduced Einstein–Vlasov system on the interval $[0,T)$ for some positive real number T is such that $P(t)$ and $A(t,0)$ are bounded then the solution extends to an interval $[0,T_1)$ with $T_1 > T$. In particular, if the maximal interval of existence $[0,T_*)$ is finite then either $P(t)$ or $A(t,0)$ is unbounded there.*

P r o o f. Suppose that $P(t)$ is bounded on the interval $[0,T)$. Then the matter quantities ρ, $\mathrm{tr}S$ and S_R are bounded there. It has been shown above that this, together with the boundedness of $A(t,0)$, implies that the quantities A, A', A'', $R^{-1}A'$, α, α', α'', $R^{-1}\alpha'$, K, K', $R^{-1}K$. β, β', and $R^{-1}\beta$ are bounded. It remains to show that all higher spacetime derivatives of all these quantities are bounded. We have a C^2 bound for α and A and a

C^1 bound for K and β when these are considered as functions of R. These imply a C^2 bound for A, α and a C^1 bound for β^a and k_{ab} in the three dimensional sense, using the results of the appendix. It follows that a C^1 bound for all the coefficients of the Vlasov equation on the support of f is obtained. The equations obtained by differentiating the Vlasov equation with respect to x and v then give the boundedness of the first derivatives of f with respect to x and v. Using the definition of the energy-momentum tensor gives a C^1 bound for its Cartesian components. The results of the appendix then imply a C^1 bound for ρ, j and $\mathrm{tr}S$.

Higher derivatives can now be bounded inductively. Assume that a solution of the reduced equations on a given time interval is such that the C^{k+1} norms of α and A and the C^k norms of K, β, f, ρ, j, $\mathrm{tr}S$ and S_R are bounded and that A^{-1} is also bounded. Note that it has already been shown that under the hypotheses of the theorem this statement holds for $k = 1$ and this suffices to start the induction. Now consider the case of general k. It is convenient to rewrite some of the reduced equations in the following form:

$$(A^{1/2})'(R) = -\tfrac{1}{8}R^{-2}\int_0^R s^2[A^{5/2}(\tfrac{3}{2}K^2 + 16\pi\rho)](s)ds \tag{2.29}$$

$$A(R)\alpha'(R) = R^{-2}\int_0^R s^2[A^2(\tfrac{3}{2}K^2 + 4\pi(\rho + \mathrm{tr}S))](s)ds \tag{2.30}$$

$$A^3(R)K(R) = R^{-3}\int_0^R s^3[4\pi A^4 j](s)ds \tag{2.31}$$

$$\beta'(R) = \beta'(0) + R\int_0^R s^{-1}[\tfrac{3}{2}\alpha K](s)ds \tag{2.32}$$

Applying Lemma A5 to (2.29) and (2.30) gives C^{k+1} bounds for $(A^{1/2})'$ and $A\alpha'$, considered as functions of R. Combining this with the information already available gives C^{k+2} bounds for A and α, considered as functions of R. In a similar way, (2.31) and Lemma A5 give a C^{k+1} bound for K. The quantity $\beta'(0)$ is already known to be bounded. Hence (2.32) and Lemma A6 imply a C^{k+1} bound for β. (In fact it implies a C^{k+3} bound, but that is not relevant here.) Moreover, it can be checked that the derivatives of these functions which are required to vanish in order that the functions have the corresponding differentiability when considered as functions of three variables, according to the results of the appendix, do so. For the same reason bounds for the derivatives of these functions of three variables are obtained. It follows in particular that the coefficients of the Vlasov equation are C^{k+1}. Hence the solution of the Vlasov equation is bounded in the C^{k+1} norm. An immediate consequence is that the Cartesian components of the energy-momentum tensor are bounded in the C^{k+1} norm. Finally, applying the results of the appendix again shows that the C^{k+1} norms of ρ, j and $\mathrm{tr}S$ are bounded and this completes the inductive step.

It was mentioned earlier that in order to prove global existence it would suffice to bound the derivatives of all orders of all quantities of interest with respect to t and R. Here only the derivatives with respect to R have been bounded and it turns out to be difficult to bound the time derivatives of the lapse function directly. Fortunately it is enough, in the present context, to bound the spatial derivatives, as will now be shown.

Let t_n be a sequence of times with $t_n < T$ for each n and $\lim_{n\to\infty} t_n = T$. The initial data induced by the given solution on the hypersurfaces $t = t_n$ define a sequence of initial data which are bounded in the C^∞ topology. In fact they are also bounded in the topology of a weighted Sobolev space. To prove this it suffices to obtain some estimates on an exterior region, say that defined by $R \geq 1$. Equation (2.29) and the conservation of ADM mass shows that A' is $O(R^{-2})$, uniformly in t. Equation (2.31) and the fact that the support of the matter is contained in a region of the form $R \leq R_0$ for all t in the interval $[0, T_1)$ shows that $K = O(R^{-3})$, uniformly in t. Using (2.25) and the fact that under the given circumstances the $O(R^{-4})$ error term there is uniform in t shows that $A - 1 = O(R^{-1})$, uniformly in t. Thus $g_{ab} - \delta_{ab}$ is bounded in H^1_δ and k_{ab} is bounded in $H^0_{\delta+1}$ for $-3/2 < \delta < -1/2$. To apply the more precise version of the local existence theorem for asymptotically flat spacetimes it is necessary to have a similar statement for weighted Sobolev spaces of higher order. This can be proved straightforwardly by induction using the equations (2.5) and (2.7). It follows that the solutions of the Einstein–Vlasov system corresponding to the data on each of the hypersurfaces of constant t exist on some time interval of length ϵ about the initial time where data are given, with ϵ independent of n. Hence the solution extends to the interval $[0, T + \epsilon)$.

With this result in hand, it is easy to show that the first singularity, if one exists, must occur in the centre.

THEOREM 2.2. *If a solution of the reduced Einstein–Vlasov system on the interval* $[0, T)$ *for some positive real number* T *is such that it has a smooth extension to an open neighbourhood of the point with coordinates* $(T, 0)$ *then the solution extends to an interval* $[0, T_1)$ *with* $T_1 > T$. *In particular, if the maximal interval of existence* $[0, T_*)$ *is finite then the solution has a singularity at the point* $(T_*, 0)$

P r o o f. The neighbourhood occurring in the hypotheses of the theorem contains all points with $t > T_1$ and $R \leq R_1$ for some $T_1 < T$ and some $R_1 > 0$. Since the solution is smooth for $t < T_1$ it follows that all unknowns in the reduced system are bounded on the region $R \leq R_1$, $0 \leq t < T$. The value of A at any point of a hypersurface of constant time can be bounded by its value at the centre at the given time and so A is bounded on the interval $[0, T)$. It was shown earlier that on any region of the form $R \geq R_1$ the quantities $A^{-1}A'\ K$ and α' are uniformly bounded. Hence under the present assumptions the quantities A', K and α' are bounded on the interval $[0, T_1)$. It was also shown that β is always bounded everywhere. It can be concluded that all the functions of t and x occurring as coefficients in the Vlasov equation are bounded. The characteristic system of the Vlasov equation then implies an inequality of the form $P(t) \leq P(0) + C\int_0^t 1 + P(s)ds$. By Gronwall's lemma $P(t)$ is bounded on the interval $[0, T_1)$, and applying Theorem 2.1 completes the proof.

R e m a r k. It is clear from the proof that the existence of an extension could be replaced by the assumption that there exists some $R_1 > 0$ such that ρ and A are bounded on the region $R \leq R_1$.

Theorems 2.1 and 2.2 are analogues of Theorem 3.2 of [20] and Theorem 4.1 of [22] respectively. It is instructive to compare the theorems involving maximal-isotropic coor-

dinates proved here with those involving Schwarzschild coordinates proved in [20] and [22]. The continuation criterion of Theorem 2.1 appears at first sight weaker than that of [20] since it is assumed that not only $P(t)$ but also $A(t,0)$ is bounded. However, it is much easier to pass from Theorem 2.1 to Theorem 2.2 than it is to pass from the continuation criterion of [20] to the regularity theorem of [22]. Moreover, the passage from Theorem 2.1 to Theorem 2.2 does not involve any deep analysis of the Vlasov equation, which the proof of Theorem 4.1 of [22] does. Thus it is reasonable to hope that the method of proof used here can more easily be adapted to matter models other than the Vlasov equation than the approach using Schwarzschild coordinates. In the next section it will be seen that the apparently weaker continuation criterion given by Theorem 2.1 is also good enough to be applied in the proof of a global existence theorem for small initial data.

At this point some further remarks on the notion of 'well-behaved' matter models are in order. Consider the case of dust, i.e. a perfect fluid without pressure. In fact (see [24]), smooth solutions of the Einstein-dust equations can be considered as distributional solutions of the Einstein–Vlasov system. Dust satisfies the dominant energy and non-negative pressures conditions. However it cannot be expected that an analogue of Theorem 2.2 holds for dust. The reason is the occurrence of so-called shell-crossing singularities, which do not occur at the centre. As has been discussed in [24] and [23] this is of significance for the formulation of the cosmic censorship hypothesis and Theorem 2 can be taken as an indication that the Einstein–Vlasov system is a good starting point for studying the cosmic censorship hypothesis and has advantages over other, superficially simpler matter models, such as a perfect fluid. This is one of the main motivations for investigating the global properties of solutions of these differential equations.

3. Global existence for small data. In the last section a continuation criterion was given for solutions of the reduced equations. Now it will be applied to obtain a global existence theorem in a particular situation, namely that of small data. The notion of smallness of initial data is defined in the present context in terms of three quantities which characterize the size of the data. Let $F_0 = \|f(0)\|_\infty$, $P_0 = P(0)$ and let R_0 be the smallest value of R such that $f(0,R)$ vanishes for $R > R_0$.

THEOREM 3.1. *Let K be a fixed positive constant and consider initial data for the reduced equations with $R_0 \leq K$ and $P_0 \leq K$. Then there exists an $\epsilon > 0$ such that for all data of this type which, in addition, satisfy $F_0 < \epsilon$ the corresponding solution exists globally in time and the spacetime which it defines is timelike and null geodesically complete.*

Remarks 1. The spacetimes of the theorem are also spacelike geodesically complete but this will not be proved here. It is the completeness of timelike and null geodesics which is most interesting physically, since these represent the wordlines of particles.

2. The statement on geodesic completeness is an important part of the theorem since a theorem on global existence in some coordinate time does not necessarily imply any interesting invariant information.

In fact more detailed information concerning the asymptotic behaviour of the spacetimes covered by the theorem will be obtained. In particular, information will be obtained

on the decay of the curvature as $t \to \infty$. A good understanding of the decay properties of the curvature is also important for the proof of the theorem and for this reason the curvature components in a Cartesian frame will now be examined in some detail. The curvature can be computed using the following relations (which are independent of symmetry assumptions):

$$\begin{aligned} {}^{(4)}R_{abcd} &= R_{abcd} + k_{ac}k_{bd} - k_{ad}k_{bc} \\ {}^{(4)}R_{\sigma abc}n^{\sigma} &= -\nabla_c k_{ab} + \nabla_b k_{ac} \\ {}^{(4)}R_{\sigma a\tau b}n^{\sigma}n^{\tau} &= -8\pi[S_{ab} + \tfrac{1}{2}(\rho - \mathrm{tr}S)g_{ab}] + R_{ab} + \mathrm{tr}k k_{ab} - k_{ac}k_b^c \end{aligned} \tag{3.1}$$

The first of these equations is the Gauss equation, the second the Codazzi equation and the third the Einstein evolution equation. It turns out that the frame components of the curvature tensor which are purely spatial are linear combinations of the quantities $A^{-3}A''$, $A^{-4}A'^2$ and K^2, with coefficients which are homogeneous functions of the Cartesian coordinates. The frame components with two indices corresponding to the time direction are linear combinations of these quantities and frame components of the energy-momentum tensor. The frame components with precisely one index corresponding to the time direction need to be calculated explicitly. They are given by:

$$\hat{R}_{0ijk} = \tfrac{1}{2}R^{-1}(x_j\delta_{ik} - x_k\delta_{ij})(K' + 3KR^{-1} + 3A^{-1}A'K) \tag{3.2}$$

The combination of metric coefficients which occurs is precisely that which is familiar from the momentum constraint.

The idea behind the proof is as follows. In flat space free particles which start in a compact set spread out linearly with time. This causes the associated density to decay. In fact it decays uniformly in space like t^{-3} as $t \to \infty$. It is reasonable to suppose that a spacetime which evolves from 'small data' (i.e. data which are 'close' to data for flat space) has small curvature, so that the behaviour of solutions of the Vlasov equation is similar to that in flat space. Thus, with luck, the density will fall off at the same rate as in flat space. Conversely, it is this fall off of the density which ensures decay of the curvature. The proof which follows makes these intuitive considerations precise and quantitative. It is broken up into a number of lemmas which are arranged in a way which parallels as closely as possible the proof of the analogous theorem in [20].

LEMMA 3.1. *Consider a spherically symmetric solution of the reduced equations on a time interval* $[0, T)$ *with the property that the support of the restriction of* $T^{\alpha\beta}$ *to each hypersurface* $t =$*const. is contained in the ball of radius* $R_0 + t$ *about the centre and* $A(t, 0) \leq 3$ *and suppose that*:

$$\|\rho\|_\infty (1+t)^{2+\delta} \leq K_1 \tag{3.3}$$

for some constants $K_1 > 0$ *and* $\delta \in (0, 1]$. *Then there exists a constant* C, *only depending on* K_1, δ *and the restriction of the solution to the initial hypersurface*, *such that*:

$$\|A''\|_\infty (1+t)^{2+\delta} + \|A'\|_\infty (1+t)^{1+\delta} + \|\alpha'\|_\infty (1+t)^{1+\delta} + \|K\|_\infty (1+t)^{1+\delta} \leq C \tag{3.4}$$

Moreover the L^∞ *norm of the frame components of the curvature tensor can be bounded by a constant times* $(1+t)^{-2-\delta}$. *This constant and that in* (*3.4*) *can be made as small as desired by choosing a sufficiently small value of* K_1.

P r o o f. Equation (2.29) and the conservation of the ADM mass (2.20) imply that $A'(R)$ can be bounded by an expression of the form CR_0^{-2} for some constant C for all $R > R_0$. The constant depends a priori on a pointwise bound for $A(t,0)$ but this is taken care of by the assumption on $A(t,0)$ occurring in the hypotheses of the lemma. An analogous estimate holds for α', as a consequence of (2.30). In the case of K it is necessary, in order to get an estimate which holds inside the matter, to cancel one power of R with a power of s in (2.31). Thus once again the estimate obtained involves R_0^{-2}. In the last section estimates in terms of R_0^{-1} were used in an optimization argument to bound A', α' and K by $\|\rho\|_\infty^{1/2}$. If they are replaced by the estimates in terms of R_0^{-2} just discussed, then the optimization argument allows the L^∞ norms of A', α' and K to be bounded by a constant multiple of $\|\rho\|_\infty^{2/3}$. This is enough to take care of the last three terms in (3.4). It was remarked in the last section that $R^{-1}A'$ can be bounded pointwise by a constant times $\|\rho\|_\infty$ if $A(t,0)$ is known to be bounded. Putting this into equation (2.5) together with the information just obtained gives the desired estimate for the first term in (3.4). The estimate for the components of the curvature tensor follows from the expressions for these components given above.

R e m a r k. The estimates obtained by optimization in the proof of this lemma could have been used in Section 2 instead of the other estimates obtained by the same method which were actually used there. The reason for presenting both types of estimates is that while the estimates in the proof of the lemma are stronger in situations where the solutions are small, the estimates of Section 2 are stronger where the solutions are large. Thus they might be important in other contexts.

The following simple lemma is taken directly from [20].

LEMMA 3.2. *Consider the ordinary differential equation $du/dt = F(t,u)$ for a C^1 function F satisfying the inequality $|F(t,u)| \leq \eta(1+t)^{-1-\delta}(1+|u|)$ for some constants $\delta > 0$ and $\eta > 0$. Then given any initial datum at $t = 0$ the corresponding solution exists on the whole of $[0,\infty)$ and satisfies the inequality $|u(t) - u(0)| \leq \frac{\eta}{\delta} \exp \frac{\eta}{\delta}(1 + |u(0)|)$.*

This lemma is used to control the behaviour of timelike geodesics in a spacetime satisfying certain inequalities. Consider a timelike geodesic γ passing through a point with coordinates (t, x^a) and suppose that it intersects the initial hypersurface $t = 0$. Let τ be proper time measured along γ, starting at $t = 0$. Let $\{e'_\sigma\}$ be an orthonormal frame which is parallelly transported and is such that e'_0 is the tangent vector to γ. Let θ'^σ be the dual coframe. A Jacobi field along γ is the derivative with respect to the parameter of a one-parameter family of geodesics in which γ is embedded. If it is expressed as a linear combination $Z^s e'_s$ of e'_1, e'_2 and e'_3 then Z^s satisfies the equation (see [12], p.96):

$$d^2 Z^s / d\tau^2 = (R^\alpha_{\beta\gamma\delta} \theta'^s_\alpha e'^\beta_0 e'^\gamma_t e'^\delta_0) Z^t \tag{3.5}$$

It turns out to be crucial for the global existence theorem to estimate the Jacobian determinant of the mapping

$$v^i \to X^a(0, t, x^a, v^i) \tag{3.6}$$

for fixed values of t and x^a, where $X^a(s, t, x^a, v^i)$ is part of the solution of the characteristic system, as discussed in Section 1. Note that, as indicated by the use of the notation

v^i, the characteristic system which is of interest here is that of the Vlasov equation written in terms of frame components. This mapping can be described in words as follows. Follow the geodesic γ through (t, x^a) with initial tangent vector $v^\mu e_\mu$ backwards until it meets the hypersurface $t = 0$. The derivative of this mapping takes vectors tangent to the mass shell at the point with coordinates v^i to vectors tangent to the initial hypersurface. A vector tangent to P at e'_0 can be identified with a vector in the tangent space to M at (t, x^a) which is orthogonal to e'_0. This can then be expressed as a linear combination $Y^s e'_s$. The frame vectors e'_s can be fixed uniquely by requiring that they be obtained from the coordinate vectors $\partial/\partial v^i$ by means of the Gram–Schmidt process. The Jacobian of the mapping which transforms from the basis $\partial/\partial v^i$ to the basis e'_s and that of its inverse can be bounded in terms of $|v|$. The derivative of the mapping (3.6) can conveniently be written as a composition of three linear mappings from $\mathbf{R}^3$ to itself. The first mapping L_1 is the one sending the components of a vector tangent to the mass shell in the basis $\{\partial/\partial v^a\}$ to the components Y^s. Let $L_2(Y^s)$ be the value at $t = 0$ of the solution of (3.5) with the initial data $Z^s = 0$ and $dZ^s/d\tau = Y^s$ at the point (t, x^a). Let $L_3(Z^s)$ be the Cartesian components of the vector obtained by projecting the vector $Z^s e'_s$ onto the hypersurface $t = 0$ along the vector e'_0. The components of this vector are given explicitly by:

$$W^a = Z^s[e'^a_s - (e'^0_s/e'^0_0)e'^a_0] \tag{3.7}$$

the derivative of (3.6) at the point of interest is $L_3L_2L_1$.

Lemma 3.3. *Consider spherically symmetric solutions of the Einstein–Vlasov system on intervals* $[0, T)$ *satisfying the following conditions*:

$$\left.\begin{aligned} &\|f_0\|_\infty \leq \epsilon, \\ &P_0 \leq K_2, \\ &R_0 \leq K_3, \end{aligned}\right\} \tag{3.8}$$

the inequality (3.3) and the condition that $A \leq 3$. *Then if* ϵ *and* K_1 *are sufficiently small there exists a constant* K_4, *depending only on* ϵ, K_1, K_2 *and* K_3, *such that* $\alpha \geq 1/2$ *and*:

$$\|T_{\alpha\beta}\|_\infty \leq K_4(1 + t)^{-3} \tag{3.9}$$

P r o o f. Lemma 3.1 provides a lot of information on the decay of geometric quantities as $t \to \infty$. In particular the estimate (3.4) can be combined with Lemma 3.2 to show that if K_1 is small enough

$$P(t) \leq P_0 + 1 \tag{3.10}$$

It follows that $\alpha'(R) \leq C\epsilon \min\{R, R^{-2}\}$. If ϵ is small enough then this can be integrated to show that $\alpha \geq 1 - 3C\epsilon/2$. This implies the first statement of the theorem. The inequality (3.10) implies a uniform bound for the Cartesian components of e'_0 if the tangent vector to the geodesic γ is contained in the support of the distribution function. (Only geodesics of this kind are of interest here.) In order to estimate the Cartesian components of the remaining frame vectors e'_s, note that it suffices, under the assumptions of the lemma, to estimate their components in the frame e_μ. Let these components be denoted by U^μ_s so

that $e'_s = U^\mu_s e_\mu$. These components satisfy:

$$U^\mu_{s,\beta} e'^\beta_0 + \gamma^\mu_{\nu\lambda} U^\nu_s U^\lambda_0 = 0 \tag{3.11}$$

The spatial components of (3.11) can be written as

$$\frac{d}{d\tau}(U^i_s) = -\gamma^i_{\nu\lambda} U^\nu_s U^\lambda_0 \tag{3.12}$$

To profit from this it is necessary to have some information about the relation between proper time and coordinate time along γ. In fact

$$d\tau/dt = \alpha(1+|v|^2)^{-1/2} \tag{3.13}$$

Since $\alpha \leq 1$ the relations (3.12) and (3.13) imply that

$$\left|\frac{d}{dt}(U^i_s)\right| \leq C(|U^i_s| + |U^0_s|)(1+t)^{-1-\delta} \tag{3.14}$$

On the other hand, since e'_s is a unit vector, $|U^0_s| \leq C|U^i_s|$. Consider now the vectors e'_s along all geodesics contained in the support of f. We have a uniform bound for the initial data for equation (3.12) and so (3.14) and Lemma 3.2 imply the global boundedness of U^i_s and hence of U^0_s. (Global boundedness means by definition that they can be bounded by a constant which is independent of T.) As indicated above, this implies a global bound for the Cartesian coordinate components of the frame vectors e'_s. The detour through the quantities U^μ_s is necessitated by the fact that while good bounds are available for the rotation coefficients, it is not obvious that the Christoffel symbols, which contain time derivatives of lapse and shift, satisfy analogous bounds.

Consider now once again a timelike geodesic γ whose tangent vector is contained in the support of the distribution function. As a consequence of (3.13) there are positive constants C_1, C_2 such that

$$C_1 t \leq \tau \leq C_2 t \tag{3.15}$$

along γ. We can assume without loss of generality that $C_2 \geq 1$ and then

$$(1+t)^{-2-\delta} \leq C_2^{2+\delta}(1+\tau)^{-2-\delta} \tag{3.16}$$

Thus if $K^s_t = R^\alpha_{\beta\gamma\delta}\theta'^s_\alpha e'^\beta_0 e'^\gamma_t e'^\delta_0$, an estimate of the form

$$|K^s_t(\tau)| \leq C(1+\tau)^{-2-\delta} \tag{3.17}$$

holds along γ. Let τ_0 be the value of τ at the point (t, x^a) and let $Z^s(\tau)$ be the solution of (3.5) with $Z^s(\tau_0) = 0$ and $dZ^s/d\tau(\tau_0) = Y^s$. Let

$$E^s(\tau) = Z^s(\tau) - (\tau - \tau_0)Y^s \tag{3.18}$$

Then by Taylor's theorem with integral remainder

$$E^s(\tau) = \int_{\tau_0}^{\tau} (\tau_0 - \sigma)(\sigma - \tau)K^s_t(\sigma)Y^t d\sigma + \int_{\tau_0}^{\tau} (\tau - \sigma)K^s_t(\sigma)E^t(\sigma)d\sigma \tag{3.19}$$

The first integral can be estimated as follows:

$$\begin{aligned}\left|\int_{\tau}^{\tau_0} (\tau - \sigma)(\sigma - \tau_0)K^s_t(\sigma)Y^t d\sigma\right| &\leq (\tau_0 - \tau)|Y^t| \int_{\tau}^{\tau_0} C\sigma(1+\sigma)^{-2-\delta} d\sigma \\ &\leq C(\tau_0 - \tau)|Y^t|\end{aligned} \tag{3.20}$$

since $\int_0^\infty \sigma(1+\sigma)^{-2-\delta}d\sigma < \infty$. Combining (3.19) and (3.20) and applying Gronwall's inequality gives

$$|E^s(\tau)| \leq C(\tau_0 - \tau)|Y^t| \exp \int_\tau^{\tau_0} C(\sigma - \tau)(1+\sigma)^{-2-\delta}d\sigma \tag{3.21}$$

Hence

$$|E^s(\tau)| \leq C(\tau_0 - \tau)|Y^t| \tag{3.22}$$

If K_1 is chosen small enough then $C < \frac{1}{2}$ and using (3.22) and the definition of E^s shows that $|\det L_2| \geq C\tau_0^3$. As a consequence of (3.15) τ_0 can be replaced by t in this inequality. The determinant of L_1 can be bounded from below uniformly for all geodesics whose tangent vectors are contained in the support of the distribution function. The same is true of the determinant of L_3, as can be seen from (3.7). Thus the statement is obtained that

$$\left|\det\left(\frac{\partial X^a}{\partial v^b}\right)(0, t, x^a, \cdot)\right| \geq Ct^3 \tag{3.23}$$

This has been shown for all v^a in the support of the distribution function. If we knew that the mapping (3.6) was injective on the support of f then we could change variables from v^i to X^a in the definition of the energy-momentum tensor. As a consequence of (3.23) and the boundedness of P this would give an estimate of the form (3.9), completing the proof of the lemma.

It will now be shown that for K_1 sufficiently small this mapping is indeed injective. Suppose that, on the contrary, there are two distinct geodesics γ_0 and γ_1 starting at the point with coordinates (t, x^a) which meet the initial hypersurface at the same point. Let v_0^i and v_1^i be the components of their initial tangent vectors on the mass shell. For $l \in [0,1]$ let $v_l^i = (1-l)v_0^i + lv_1^i$. Let γ_l be the geodesic with initial tangent vector corresponding to v_l^i. Denote the spatial coordinates of the point of intersection of γ_l with the initial hypersurface $t = 0$ by $\xi^a(l)$. Then $\xi^a(0) = \xi^a(1)$ and $\xi^a(l)$ is a closed curve. Let $w^i = v_1^i - v_0^i$. The tangent vector to the curve $\xi^a(l)$ is the image under the mapping $L_3L_2L_1$ of w^i. It can be concluded from (3.23) that the tangent vector to the curve $\xi^a(l)$ can never vanish. It is convenient at this point to make a different choice of the frame e_i'. The change is that instead of the basis $\{\partial/\partial v^i\}$ of the tangent space to the mass shell as above, a basis is chosen which consists of vectors which are linear combinations of the vectors $\partial/\partial v^i$ with constant coefficients with the first vector of the basis being $w^i\partial/\partial v^i$. The vector $L_1(w)$ is proportional to $(1,0,0)$. Using the smallness assumption on the data shows that the components $e_1'^a(0)$ are close to the components $e_1'^a(\tau_0)$. More precisely, for K_1 small enough

$$|e_1'^a(0) - e_1'^a(\tau_0)| \leq \eta|e_1'^a(\tau_0)| \tag{3.24}$$

for any given $\eta > 0$. Similarly

$$|e_0'^a(0) - e_0'^a(\tau_0)| \leq \eta|e_0'^a(\tau_0)| \tag{3.25}$$

In the same sense $L_2L_1(w)$ is close to a vector of the form $(Z^1, 0, 0)$. For Z^s is close to W^s and the frame vectors have also only changed a little between $\tau = \tau_0$ and $\tau = 0$. The tangent vector to the curve $\xi^a(l)$ is given by the projection of $L_2L_1(w)$ on the hypersurface $t = 0$ along the vector $e_0'(0)$. The explicit form of this projection is displayed in equation

(3.7). It will now be shown that, for K_1 sufficiently small the second term in the square brackets in (3.7), with $s = 1$, can be bounded in modulus by a constant $k < 1$ times the modulus of the first term. Before proving this statement it will be shown that it implies the desired result. Z^s is as close as desired to a vector of the form $(Z^1, 0, 0)$ and so W^a is as close as desired to the projection of a vector proportional to $e_1'(0)$. The estimate for the terms in (3.7) with $s = 1$ implies that the projection of a vector proportional to e_1' is contained in a convex cone about $e_1'^s$. Hence W^a is contained in a convex cone about a vector proportional to w^a. This is inconsistent with the fact that $\xi^a(l)$ returns to its starting point. Thus it only remains to prove the above statement about the relative sizes of the terms in (3.7). Now $e_1'^a$ is close to $A^{-1}Bw^a$ for a certain quantity B, $e_1'^0$ is close to $B\alpha^{-1}(w\cdot v_l)(1+|v_l|^2)^{-1/2}$, $e_0'^a$ is close to $A^{-1}v_l^a$ and $e_0'^0$ is close to $\alpha^{-1}(1+|v_l|^2)^{1/2}$. Thus the two terms are close to Bw^a and $B(w\cdot v_l)(1+|v_l|^2)^{-1}v_l^a$. Using the Cauchy–Schwarz inequality the modulus of the second expression can be bounded by $|B||w||v_l|^2(1 + |v_l|^2)^{-1}$ while the modulus of the first is $|B||w|$. Using the fact that $|v_l| \leq P$, so that $|v_l|^2/(1 + |v_l|)^2 \leq P^2/(1 + P^2) \leq k < 1$ completes the argument.

Lemma 3.4. *Suppose that initial data for the reduced system satisfy the inequalities (3.8) and let T_0 be a fixed positive number. Then for fixed K_2, K_3 there exists $\epsilon > 0$ such that the solution corresponding to this data exists on the interval $[0, T_0]$, and $P(t) \leq 2P(0)$, $A(t, 0) \leq 2$ and $\|\rho\|_\infty$ is as small as desired there.*

Proof. First a system of integral inequalities for $P(t)$ and $Q(t) = A(t, 0)$ will be derived. It is elementary that $\|\rho\|_\infty \leq C\|f\|_\infty(1 + P(t))^4$ with a constant C independent of the initial data. It was shown in the last section that the coefficients in the part of the characteristic system controlling the velocities can be bounded by $C\|\rho\|_\infty^{1/2}(1 + Q(t))^q$, for some positive integer q. Hence

$$P(t) \leq P(0) + C\epsilon^{1/2} \int_0^t P(s)(1 + P(s))^2(1 + Q(s))^q ds \tag{3.26}$$

Equation (2.9) and the estimates obtained above imply that

$$Q(t) \leq Q(0) + C\epsilon^{1/2} \int_0^t P(s)(1 + P(s))^p(1 + Q(s))^q ds \tag{3.27}$$

for some positive integer p. Let z_1 and z_2 be the unique solutions of the system of integral equations obtained by replacing the inequalities in (3.26) and (3.27) by equalities and P and Q by z_1 and z_2 respectively, with initial data $z_1(0) = P(0)$ and $z_2(0) = Q(0)$. When ϵ is zero the solution of the integral equations is constant. In particular, it is global in time. Moreover the initial data $z_1(0)$ and $z_2(0)$ can be bounded in terms of K_2, K_3 and ϵ. It follows that for ϵ sufficiently small the solution of these integral equations exists on a time interval $[0, T_1)$ with $T_1 > T_0$ and, by making ϵ smaller if necessary, it can be assumed that $z_1(t) \leq 2P(0)$ and $z_2(t) \leq 2$ for $0 \leq t \leq T_1$. Comparing the integral inequalities with the integral equations shows that $P(t) \leq z_1(t)$ on any interval where both are defined and similarly $Q(t) \leq z_2(t)$. By the above estimates $\|\rho\|_\infty$ will be as small as desired if ϵ is chosen sufficiently small. The continuation criterion given by Theorem 2.1 implies that the solution can be extended to any time interval where the solution (z_1, z_2) of the integral equations exists and this completes the proof of the lemma.

THEOREM 3.2. *Let a non-negative C^∞ compactly supported spherically symmetric maximal initial datum f_0 for the Einstein–Vlasov system be given which satisfies the inequalities (3.8) for some positive constants ϵ, K_2 and K_3. Then if ϵ is small enough the corresponding solution of the reduced system exists globally in time. Moreover, for this solution*

$$\|T_{\alpha\beta}\|_\infty \leq C(1+t)^{-3} \tag{3.28}$$

the metric coefficients α^{-1} and A are bounded and the estimates (3.4) hold with $\delta = 1$.

Proof. Let K_1 and ϵ be positive constants which are small enough so that the conclusions of Lemma 3.3 hold for some $\delta < 1$. Let T_1 be a positive number satisfying $K_4(1+T_1)^{\delta-1} < K_1$. By Lemma 3.4 the constant ϵ can be chosen so small that the solution exists on the time interval $[0, T_1)$. Moreover, it can be arranged that on this interval the solution is as small as desired. In particular, ϵ can be chosen so that (3.3) is satisfied on the interval $[0, T_1)$ and $A \leq 2$ there. Consider a fixed initial datum satisfying (3.8). Define T_* to be the supremum of those positive numbers T such that the solution of the reduced equations corresponding to the given initial data exists on $[0, T)$, and $A \leq 2$ for the solution on this interval and it satifies (3.3) there. Here $T_* = \infty$ is possible. In fact we will show that it is the only possibility. For suppose that $T_* < \infty$. The assumptions already made ensure that $T_* > T_1$. The definitions of T_1 and T_* and Lemma 3.3 then show that the continuation criterion is satisfied on $[0, T_*)$ and that

$$\|T^{\alpha\beta}(T_*)\|_\infty \leq K_1(1+T_*)^{-2-\delta} \tag{3.29}$$

This means that the solution can be extended to an interval $[0, T_2)$ with $T_2 > T_*$. Also T_2 can be chosen so that (3.3) is satisfied there. This contradicts the definition of T_* and so in fact it must be the case that $T_* = \infty$. The inequality (3.3) holds on $[0, \infty)$ for some $\delta < 1$. Applying Lemma 3.1 shows that an inequality of this form also holds for $\delta = 1$. The remaining conclusions of the theorem then follow from Lemma 3.1.

The statement of Theorem 3.2 includes all the conclusions of Theorem 3.1 except that concerning geodesic completeness. Equation (3.15) shows that along a timelike geodesic proper time and coordinate time are equivalent. This, together with global existence in coordinate time shows that timelike geodesics are complete. Similarly, the information which we have on the geometry is more than enough to show that an affine parameter along a null geodesic is equivalent to coordinate time.

At the end of the last section some remarks were made about the relation of the Vlasov equation with dust. For the Einstein-dust system there is no reasonable smallness assumption on initial data which ensures geodesic completeness of the corresponding solutions, as follows from the work of Christodoulou [5]. The only other matter model for which a global existence theorem for spherically symmetric asymptotically flat solutions of the Einstein-matter equations with small initial data has been proved is the massless scalar field [6]. The method of proof used here in the case of the Vlasov equation can not be applied directly to the case of matter models with radiation such as the massless scalar field, since there the flat space fall-off rates for the matter only make the energy-momentum tensor fall off like t^{-2} and do not furnish the faster decay rates used in the above.

4. Local existence in general. The purpose of this section is to present some aspects of the local existence theorem for the Einstein–Vlasov system, without any symmetry assumptions. The standard method for proving such theorems proceeds in several steps. First, some coordinate conditions are imposed, leading to a system of 'reduced equations'. This is similar to what was done in Section 2 in the spherically symmetric case. In the general case different coordinate conditions are used and hence the reduced system is also different. The second step is to prove a local existence theorem for the reduced equations. The third step is to establish the connection between the reduced equations and the full equations. This means showing that if the data for the reduced system satisfies certain gauge conditions, if the constraints are satisfied on the initial hypersurface and if the reduced equations are satisfied everywhere, then the coordinate condition and the constraints are satisfied everywhere. This then implies that the solution of the reduced equations is actually a solution of the Einstein equations. The first and third steps are not discussed further here; the reader is referred to [8] for details. (The treatment which follows uses the harmonic coordinate condition, which is that discussed in [8].)

The reduced equations in harmonic coordinates constitute a system of nonlinear wave equations for the metric coupled to the Vlasov equation for the distribution function f. Solutions of this system have a domain of dependence determined by the light cone and so when proving a local existence theorem spatial boundary conditions play no role. For this reason it is sufficient to consider data of compact support on $\mathbf{R}^3$. The data for the reduced system consists of all components of the metric and their first time derivatives together with the distribution function. It is assumed that $g_{\alpha\beta} - \eta_{\alpha\beta}$, $\partial_t g_{\alpha\beta}$ and f have compact support on the initial hypersurface. Such data can never satisfy the constraints globally on $\mathbf{R}^3$ except in the case of data for flat space but this is not a problem, because of the possibility of using the domain of dependence to localize. To start with only the case of C^∞ data is treated but it will be shown later that it is easy to extend the argument to data of finite differentiability.

Before going further some remarks will be made on the various notations used for derivatives in this section. This is intended to avoid confusion which might arise from the mixture of notation from differential geometry and from analysis which occurs. As in previous sections Greek or Roman indices attached to geometric objects are usually to be thought of as abstract indices although they may also occasionally denote components in a coordinate system. No confusion should arise from this dual role. Indices of this kind are also used to label the coordinate functions themselves and the partial derivatives with respect to the coordinates, denoted by ∂. The ranges of the indices are as before. On the other hand the indices on the operator D denoting differentiation are of a different kind. Here use is made of the multi-index notation which is very effective in handling expressions containing high order derivatives of functions of several variables. These multi-indices always refer to derivatives with respect to the spatial coordinates x^a. The expression $D^\alpha u$ denotes a particular higher order partial derivative of the function u with repect to the spatial variables. The order of this derivative is denoted by $|\alpha|$. Roman indices on D are used in connection with norms as in equation (1.11). We have $\|D^i u\| = \max_{|\alpha|=i} \|D^\alpha u\|$. This notation may be applied to any norm.

In order to have a suitable framework for proving a local existence theorem it is necessary to work with function spaces which are well adapted to the equation being studied. The natural spaces for hyperbolic equations are the Sobolev spaces. In the present context, where the asymptotic behaviour plays no role, it is possible (and convenient) to use Sobolev spaces without weights. If u is a smooth function on $\mathbf{R}^3$ define:

$$\|u\|_{H^s} = \left[\sum_{i=0}^{s} \int |D^i u|^2 dx\right]^{1/2} \tag{4.1}$$

Compare this with the weighted Sobolev norms defined in (1.11).

The reduced Einstein equations take the form:

$$g^{\gamma\delta}\partial_\gamma\partial_\delta g_{\alpha\beta} = F_{\alpha\beta}(g_{\gamma\delta}, \partial_\epsilon g_{\gamma\delta}) + 8\pi[T_{\alpha\beta} - \tfrac{1}{2}g^{\gamma\delta}T_{\gamma\delta}g_{\alpha\beta}] \tag{4.2}$$

It is convenient (although not essential) to write this in first order form by introducing the first derivatives of the metric as additional variables. Let $h_{\alpha\beta\gamma} = \partial_\gamma g_{\alpha\beta}$. Then in terms of the unknowns $g_{\alpha\beta}$ and $h_{\alpha\beta\gamma}$ the equations can be written as follows:

$$\begin{aligned} -g^{00}\partial_0 h_{\alpha\beta 0} - 2g^{0a}\partial_a h_{\alpha\beta 0} &= g^{ab}\partial_a h_{\alpha\beta b} + \dots \\ g^{ab}\partial_0 h_{\alpha\beta a} &= g^{ab}\partial_a h_{\alpha\beta 0} \\ \partial_0 g_{\alpha\beta} &= h_{\alpha\beta 0} \end{aligned} \tag{4.3}$$

The terms which are not written out explicitly are those coming from the right hand side of (4.2). The first of these can be written as a function of $g_{\alpha\beta}$ and $h_{\alpha\beta\gamma}$ without using derivatives. The reason why certain linear combinations of the original equations have been taken when writing (4.3) is to make contact with the notion of a symmetric hyperbolic system. Let u be a function which takes values in an open set U of $\mathbf{R}^k$ for some k. A differential equation for u of the form:

$$A^0(x^\alpha, u)\partial_0 u + A^a(x^\alpha, u)\partial_a u + B(x^\alpha, u) = 0 \tag{4.4}$$

is called symmetric hyperbolic if the matrices $A^0(x^\alpha, u)$ and $A^i(x^\alpha, u)$ are symmetric for all (x^α, u) and if $A^0(x^\alpha, u)$ is positive definite. Here A^0, A^a are functions on the product of U with some open subset of $\mathbf{R}^4$ with values in the $k \times k$ matrices and B is a function on the same domain with values in $\mathbf{R}^k$. If $u = (g_{\alpha\beta} - \eta_{\alpha\beta}, h_{\alpha\beta\gamma})$, with $\eta_{\alpha\beta}$ denoting the components of the Minkowski metric in standard coordinates then the equations (4.3) are of this form, for each fixed $T_{\alpha\beta}$. Moreover the reduction process, which is not described explicitly here, can be (and usually is) done in such a way that $g_{00} = -1$ and $g_{0a} = 0$ on the initial hypersurface. The open set U is defined by the condition that $g_{\alpha\beta}$ have Lorentz signature. The Vlasov equation will be written in the form (1.1). Note that the Christoffel symbols are rational functions of the unknowns u. The full system of reduced equations consists of (4.3), (1.1) (with the Christoffel symbols expressed algebraically in terms of u) and the definition (1.3) of $T_{\alpha\beta}$. The Vlasov equation itself is a symmetric hyperbolic equation of a particularly simple type, with only one unknown. This symmetric hyperbolic equation is defined on $\mathbf{R}^6$ rather than $\mathbf{R}^3$. The main theorem can now be stated:

THEOREM 4.1. *Let (u_0, f_0) be C^∞ compactly supported initial data for the reduced Einstein–Vlasov system written in first order form. Then there exists a $T > 0$ and a C^∞*

solution (u, f) of the reduced system on the interval $[0, T)$ which induces the given initial data. Moreover $f(t)$ has compact support for each fixed t.

Remarks. 1. The domains of definition of the functions u_0, f_0, u and f are $\mathbf{R}^3$, $\mathbf{R}^6$, $\mathbf{R}^3 \times [0, T)$ and $\mathbf{R}^6 \times [0, T)$ respectively.

2. The solution (u, f) is the unique solution with the properties stated in the theorem.

In proving this result an existence theorem for solutions of linear symmetric hyperbolic systems with C^∞ initial data is assumed (see e.g. [9], p. 668). In order to prove Theorem 4.1 a certain iteration is defined and then it is proved that this iteration converges in an appropriate sense. The convergence follows from certain inequalities satisfied by the solutions of linear symmetric hyperbolic equations. The proof makes use of the following Moser-type inequalities [14, 1]. Here these inequalities are only needed in the case of functions which are smooth and compactly supported. The first inequality concerns products of functions and says that if f amd g are smooth functions of compact support on $\mathbf{R}^n$ then

$$\|D^s(fg)\|_2 \leq C(\|D^s f\|_2 \|g\|_\infty + \|f\|_\infty \|D^s g\|_2) \tag{4.5}$$

Here $\| \ \|_2$ denotes the L^2 norm. The second says that under the same conditions for any derivative D^α of order s

$$\|D^\alpha(fg) - fD^\alpha g\|_2 \leq C(\|D^s f\|_2 \|g\|_\infty + \|Df\|_\infty \|D^{s-1} g\|_2) \tag{4.6}$$

The last estimate concerns the composition of a smooth function F with a function in a Sobolev space. Suppose that F is a smooth function defined on an open subset U of $\mathbf{R}^k$ and that f takes values in an open set V of $\mathbf{R}^k$ whose closure is compact and contained in U. Then for $s \geq 1$:

$$\|D^s(F(f))\|_2 \leq C\|F\|_{C^s} \|f\|_\infty^{s-1} \|D^s f\|_2 \tag{4.7}$$

Here the C^s norm of F is taken over $\bar{V}$. Note that $s = 0$ is excluded in (4.7). An inequality for the case $s=0$ can be derived as follows. Choose some fixed $u_0 \in U$. Then there exists a smooth matrix-valued function M on $U \times U$ such that $F(u) = M(u, u_0)(u - u_0)$ (cf. [11], p. 77). It follows in an elementary way that:

$$\|F(u)\|_2 \leq \|F(u_0)\|_2 + \|M(u, u_0)\|_\infty \|u - u_0\|_2$$

Now, for fixed u_0, $\|M(u, u_0)\|_\infty$ can be bounded in terms of the C^1 norm of F on $\bar{V}$ for $\|u\|$ in any open set V of the type introduced above that contains $\|u_0\|$. If in addition U contains the origin and $F(0) = 0$ then the inequality reduces to $\|F(u)\|_2 \leq \|F\|_{C^1} \|u\|_2$, which is a rather close analogue of (4.7). The fundamental tool used in the existence proof for the nonlinear equations is the following standard energy estimate for a linear symmetric hyperbolic system.

Lemma 4.1. *Let u be a smooth solution of the linear hyperbolic equation $A^0(x)\partial_0 u + A^a(x)\partial_a u + B(x) = 0$ whose restriction $u(t)$ to each hypersurface $t =$const. has compact support. Suppose further that the functions $A^\alpha - \bar{A}^\alpha$ have compact support, where $\bar{A}^\alpha$ are*

constant matrices with $\bar{A}^0$ positive definite. Then:

$$\|u(t)\|_{H^s} \leq \|u(0)\|_{H^s} + C\int_0^t [(\|A^\mu\|_{C^1} + \|\partial_t A^0\|_{C^0})\|u\|_{H^s} + (\|Du\|_\infty + \|\partial_t u\|_\infty)\|A^\mu\|_{H^s} + \|B\|_{H^s}](t')dt' \tag{4.8}$$

where the constant C only depends on upper and lower bounds for the quadratic form defined by A^0.

P r o o f. Applying the derivative D^α to the equation gives:

$$\begin{aligned} &A^0\partial_0(D^\alpha u) + A^a\partial_a(D^\alpha u) + D^\alpha B \\ &= -[D^\alpha(A^0\partial_0 u) - A^0 D^\alpha(\partial_0 u)] - [D^\alpha(A^a\partial_a u) - A^a D^\alpha(\partial_a u)] \\ &= Q^\alpha, \text{ say.} \end{aligned} \tag{4.9}$$

This can be used to compute the time derivative of the quantity $\int\langle A^0 D^\alpha u, D^\alpha u\rangle$. Differentiating under the integral and substituting in (4.9) gives a sum of terms of which only one contains derivatives of u of order higher than that of D^α. However this derivative can be eliminated by partial integration:

$$\int\langle -A^a\partial_a(D^\alpha u), D^\alpha u\rangle = \tfrac{1}{2}\int\langle \partial_a A^a D^\alpha u, D^\alpha u\rangle \tag{4.10}$$

The equation which results is:

$$d/dt(\int\langle A^0 D^\alpha u, D^\alpha u\rangle) = \int\langle(\partial_t A^0 + \partial_a A^a)D^\alpha u - 2D^\alpha B - 2Q^\alpha, D^\alpha u\rangle \tag{4.11}$$

Note that the inner product defined by A^0 is uniformly equivalent to the standard inner product on $\mathbf{R}^k$. Hence it follows that, if $N^\alpha = \sqrt{\langle A^0 D^\alpha u, D^\alpha u\rangle}$, then

$$|d/dt\int(N^\alpha)^2| \leq \int[N^\alpha(\|\partial_t A^0 + \partial_a A^a\|_\infty N^\alpha + \|D^\alpha B\|_2 + \|Q^\alpha\|_2)] \tag{4.12}$$

The L^2 norm of Q^α can be estimated with the help of the Moser estimate (4.6). Using this fact and integrating (4.12) with repect to t gives:

$$\begin{aligned} \int(N^\alpha)^2(t) &= \int(N^\alpha)^2(0) \\ &+ \int_0^t\int\{N^\alpha(t')[\|\partial_\nu A^\mu\|_\infty\|u\|_{H^s} + \|\partial_\nu u\|_\infty\|DA^\mu\|_{H^{s-1}} + \|D^\alpha B\|_2]\}dt' \end{aligned} \tag{4.13}$$

Formally, $d/dt((N^\alpha)^2) = 2N^\alpha d/dt(N^\alpha)$, so that one factor N^α can be cancelled in this formula. This formal calculation can be justified. Adding the inequalities (4.13) for all derivatives D^α of order less than or equal to some fixed s completes the proof of the lemma.

P r o o f o f T h e o r e m 4.1. Define u_0 and f_0 to be the time independent functions which agree with the initial data on the hypersurface $t=0$. Then define an iteration recursively as follows. If u_n and f_n have been defined, substitute these for u and f in all places in the reduced Einstein equations (including the energy-momentum tensor) except those where derivatives of u occur. Replace these derivatives by the corresponding derivatives of u_{n+1}. This gives a linear hyperbolic equation for u_{n+1} on an interval $[0, T_{n+1})$ which

can be solved with the initial datum which is to be prescribed for u. Here T_{n+1} is the largest time such that the metric defined by certain components of u_n is non-degenerate. Next substitute u_{n+1} into the Vlasov equation and replace f by f_{n+1}. This gives a linear hyperbolic equation on the interval $[0, T_{n+1})$ which can be solved with the initial datum which is to be prescribed for f. This defines u_{n+1} and f_{n+1} on the interval $[0, T_{n+1})$. In this way approximate solutions (u_n, f_n) of the reduced system are obtained. They all induce the correct initial data on the initial hypersurface. Let $C^k([0,T], H^s(\mathbf{R}^n))$ denote the Banach space of C^k functions on the interval $[0,T]$ with values in the Sobolev space $H^s(\mathbf{R}^n)$. It will be shown that if $s \geq 5$ and $T > 0$ is chosen sufficiently small then $T_n \geq T$ for all n, the sequence u_n is bounded in the space $C^0([0,T], H^s(\mathbf{R}^3))$ and the sequence f_n is bounded in the space $C^0([0,T], H^s(\mathbf{R}^6))$. The essential point is that for $s \geq 5$ we have the Sobolev embedding theorem which says that in $\mathbf{R}^3$ and $\mathbf{R}^6$ any H^s function is C^1 and there is a constant C such that $\|u\|_{C^1} \leq C\|u\|_{H^s}$. Hence, if the estimate of Lemma 4.1 is applied to the equation for u_{n+1}, the pointwise norms of A^a and u occurring there can be estimated in terms of the H^s norms of the same quantities. To estimate the norm of $\partial_t A^0(u_n)$ which occurs, first apply the chain rule and then substitute for the term $\partial_t u_n$ which comes up using the equation. Finally, apply the Sobolev embedding theorem to the result. The H^s norms of the coefficients $A^0(u_n)$, $A^a(u_n)$ and the part of $B(u_n)$ which does not involve the matter quantities can be bounded by a polynomial in the H^s norms of u_n, using the Moser estimate (4.7), on any interval where u takes values in a compact subset of U. (Note that this part of B maps the origin to itself and so an estimate is also obtained for the undifferentiated quantity, as in the discussion following (4.7).) It is also straightforward to show that, in the presence of a bound for the maximum momentum $P(t)$ of any particle in the support of the distribution function, the H^s norm of the energy-momentum tensor can be estimated by a constant depending on the H^s norm of u times the H^s norm of f. The result of all this is that if $U_n(t) = \sup_{k \leq n} \|u_k(t)\|_{H^s}$ and $F_n(t) = \sup_{k \leq n} \|f_k(t)\|_{H^s}$ then inequalities are obtained of the form

$$\begin{aligned} U_n(t) &\leq U_n(0) + \int_0^t G(U_n(t'), F_n(t'))dt' \\ F_n(t) &\leq F_n(0) + \int_0^t H(U_n(t'), F_n(t'))dt' \end{aligned} \tag{4.14}$$

where G and H are polynomials with non-negative coefficients. These inequalities are obtained under the assumption that the u_k with $k \leq n$ take values in a compact subset K of U and that the P_k are all bounded by some constant $\bar{P}$. The functions F and G may depend on K and $\bar{P}$. The inequalities are reminiscent of (3.26) and (3.27) and once again the solutions of the integral inequalities can be compared with the solutions of the corresponding integral equations. Choose the number T that the solution of the integral equations with initial data $(\|u\|_{H^s}, \|f\|_{H^s})$ exists on the interval $[0,T]$. For $n \geq 1$ the validity of these inequalities for the functions u_n and f_n depends on knowing that the u_k take values in a fixed compact set K of U for all $k \leq n$. It can be proved by induction that these inequalities are valid for T sufficiently small. Note first that they obviously hold for $n = 0$. Now suppose they hold up to some given value n of the index. Then they give a bound on $\|\partial_t u_n\|_\infty$. This in turn can be used to show that for T small enough (with

a smallness condition which does not depend on n) u_n takes values in a fixed compact set K and $P_{n+1}(t)$ is less than a fixed constant $\bar{P}$. This completes the inductive step. It follows that U_n and F_n are uniformly bounded on the interval $[0, T]$.

Next note that if the coefficients A^α of a symmetric hyperbolic system are smooth there exist smooth functions A_1^α such that $A^\alpha(u) - A^\alpha(u') = A_1^\alpha(u, u')(u - u')$. A similar statement applies to the function $B(u)$. Hence the difference $u_{n+1} - u_n$ satisfies a linear symmetric hyperbolic equation whose coefficients involve $A_1^\alpha(u_n, u_{n-1})$. The difference $f_{n+1} - f_n$ satisfies a similar equation. Using the estimate of Lemma 4.1, the uniform bounds for U_n and F_n and the fact that the initial data are the same for all iterates leads to an estimate of the form:

$$\|u_{n+1} - u_n\|_{H^{s-1}} + \|f_{n+1} - f_n\|_{H^{s-1}} \leq CT(\|u_n - u_{n-1}\|_{H^{s-1}} + \|f_n - f_{n-1}\|_{H^{s-1}}) \quad (4.15)$$

If T is chosen small enough this shows that the sequences u_n and f_n are Cauchy sequences in the Banach space $C^0([0, T], H^{s-1})$ and hence converge to some limits (u, f) in that space. It should be noted that the analogue of the estimate (4.15) with H^{s-1} replaced by H^s does not follow from Lemma 4.1. The reason is that the equation satisfied by the differences of iterates contain first derivatives of the iterates as inhomogeneous terms. To get some statements about convergence in H^s, a little functional analysis will be used. This uses once again the fact that the iteration is bounded in the H^s norm. The Banach–Alaoglu theorem ([18],p.115) and the fact that $L^\infty([0, T], H^s)$ is the dual of a Banach space implies that there is a subsequence which converges weakly in $L^\infty([0, T], H^s)$. It can only converge to (u, f). Thus the limiting functions u and f are in $L^\infty([0, T], H^s)$. If s is chosen to be at least seven then u_n and f_n and their spacetime derivatives converge uniformly and hence the limiting functions satisfy the equations. It must still be checked that if the initial data satisfy the condition $\partial_\gamma g_{\alpha\beta} = h_{\alpha\beta\gamma}$, then the solution also has this property. This holds because the equations (4.3) imply the equation $\partial_0(h_{\alpha\beta\gamma} - \partial_\gamma g_{\alpha\beta}) = 0$. The time of existence depends only on the H^s norm and an argument used in the proof of Theorem 2.1 shows that as long as the H^s norm, $s \geq 7$, is bounded on some interval the solution can be extended to a longer time interval. If, in deriving a differential inequality for the solution, we keep the C^1 norms instead of eliminating them using the Sobolev embedding theorem, we see that the resulting inequalities are linear in the H^s norm. What this means is the following. If any H^s norm, $s \geq 7$, is bounded then the C^1 norm is bounded and then the H^s norm is bounded for every s. Hence the solution can be extended in the space H^s for every s. This implies that a solution corresponding to C^∞ initial data is itself C^∞ as long as it exists in H^7. The statement concerning the support of f was already proved in Section 1 and hence the theorem is proved.

The estimates used in this proof can also be used to prove a local existence and uniqueness theorem for data of finite differentiability. The idea is to approximate the data of finite differentiability by a sequence of C^∞ data which converge to the original data in some Sobolev space H^s. By Theorem 4.1 there is a C^∞ solution corresponding to each of these initial data sets and it can be assumed without loss of generality that all these solutions exist on a common time interval $[0, T]$, since the sequence of data is bounded in H^s. If u and u' are solutions of a symmetric hyperbolic system then their difference satisfies a homogeneous linear symmetric hyperbolic system. The argument is

the same as that used to estimate the difference of iterates above. The estimate of Lemma 4.1 can be applied to this equation to show that if u and u' are bounded in H^s it is possible to bound the H^{s-1} norm of $u-u'$ by a constant multiple of the corresponding norm of the initial data. A similar estimate can be obtained for the reduced Einstein–Vlasov system. It follows that the sequence of C^∞ solutions introduced above converges to a solution corresponding to the initial data of finite differentiability. It may be remarked in passing that the uniqueness of the solution whose existence is asserted by Theorem 4.1 can also be proved by this method.

It is possible to prove energy estimates in weighted Sobolev spaces with a weight $\delta > -3/2$ in a way which is similar to that which has been done above for ordinary Sobolev spaces. The weighted Moser estimates which are necessary can be obtained using the techniques of Bartnik [2] from the Moser estimates for a bounded region in $\mathbf{R}^n$. These estimates imply the boundedness of a sequence of approximating solutions in $L^\infty([0,T], H^s_\delta)$. Applying the Banach–Alaoglu theorem once more shows that the previously obtained solution is in this space if the data are in H^s_δ. This is still not the most precise result on the propagation of aymptotic flatness which is desirable. For it is usually assumed that in initial data sets for the Einstein equations, the second fundamental form falls off faster than the spatial metric and it is desirable that this property be preserved by the time evolution. For a general symmetric hyperbolic system, there is no obvious reason why this should be true. For if one tries to use the equation directly to prove a property of this kind, the term $B(u)$ intervenes and in general there is no reason that this should fall off faster than u itself. However if B at least quadratic it is true. Here this is only sketched briefly and we do not even give a precise definition of the phrase 'at least quadratic'. If B is at least quadratic and $u \in H^s_\delta$, and $\delta > -3/2$, then $B(u)$ belongs to $H^s_{\delta'}$ for some $\delta' > \delta$. If $\delta' < \delta + 1$ then it follows from the equation that $\partial_t u$ is in $H^{s-1}_{\delta'}$. If $\delta' \geq \delta + 1$ then $\partial_t u$ is in $H^{s-1}_{\delta+1}$. If necessary this procedure can be repeated and after finitely many steps the result is obtained that $\partial_t u \in H^{s-1}_{\delta+1}$. In the case of the Einstein–Vlasov system, the matter term in the Einstein equations is irrelevant for this discussion since its support is in a known compact set. On the other hand, the part of B only involving the geometry is quadratic in the Christoffel symbols. The energy estimates show that the time of existence of the solution in a weighted Sobolev space only depends on a weighted Sobolev norm of the initial data.

Appendix

LEMMA A1. *Let $f : [0,\infty) \to \mathbf{R}$ be a function, q a polynomial on $\mathbf{R}^n$ which does not vanish identically and is homogeneous of degree $2m$, where m is an integer, and let F be the function on $\mathbf{R}^n$ defined by $F(x) = f(|x|)q(x)/|x|^{2m}$. If $m \geq 1$ suppose further that f is C^{2m-2} and $f^{(2l)}(0) = 0$ for all integers l with $l \leq m-1$. Then for any $k \geq 0$ the following conditions are equivalent*:
(*i*) *f is C^k and $f^{(2l+1)}(0) = 0$ for all integers l such that $2l+1 \leq k$.*
(*ii*) *F is C^k.*

P r o o f. If (ii) holds consider $g(x^1) = F(x^1, 0, \ldots, 0) = f(|x^1|)[q(x^1, 0, \ldots, 0)/(x^1)^{2m}]$. The second factor is a constant. If this constant is non-zero it can be seen immediately

that f is C^k. If the constant is zero it can be made non-zero by a rotation. Since also $g(x^1) = g(-x^1)$, the statement (i) holds. Conversely, suppose that (i) holds. Let p_k be the Taylor polynomial of order k of f at the origin. Then we can write:

$$F(x) = p_k(|x|)q(x)/|x|^{2m} + (f(|x|) - p_k(|x|))q(x)/|x|^{2m} \tag{A1}$$

Taylor's theorem applied to f shows that the second term is C^k with all its derivatives up to order k vanishing at the origin. To complete the proof it only remains to show that under the hypothesis (i) the first term is a polynomial. Under this hypothesis the first $2m-1$ derivatives of p_k at zero vanish, as well as higher derivatives of odd order. Hence $p_k(|x|)$ is a polynomial in $|x|^2$ which is divisible by $|x|^{2m}$ and the first term in A1 is a polynomial.

LEMMA A2. *Suppose that the hypotheses of Lemma A1 are satisfied and that the derivatives of f of order $\leq k$ are bounded. Then the partial derivatives of F of order $\leq k$ are bounded and there exists a constant C independent of f such that $\|F\|_{C^k} \leq C\|f\|_{C^k}$.*

Proof. On the region $|x| \geq 1$ the statement is obvious. To see that it holds when $r \leq 1$ is included, first split F into a sum of two terms as in $A1$ above. The first term and its derivatives can easily be bounded in terms of derivatives of f. A derivative of the second term of order $\leq k$ is a sum of terms, each of which is the product of a derivative of $(f-p_k)(|x|)$ with a function which is homogeneous of degree $\geq -k$ and is independent of f. The derivative of $(f-p_k)(|x|)$ of order $l \leq k$ can be bounded on the region $|x| \leq 1$ in terms of the $\|f\|_{C^l}|x|^l$. The resulting powers of $|x|$ are sufficient to compensate the functions which are homogeneous of negative degree and this gives the desired estimate.

LEMMA A3. *Let $f : [0,\infty) \to \mathbf{R}$ be a function with $f(0) = 0$ and let F^a be the vector field on $\mathbf{R}^n$ defined by $F^a(x) = f(|x|)x^a/|x|$ for $x \neq 0$ and $F(0) = 0$. Then for any $k \geq 0$ the following conditions are equivalent*:
(i) f is C^k and $f^{(2l)}(0) = 0$ for all l such that $2l \leq k$.
(ii) F^a is C^k.

Proof. $F^1(x^1, 0, \ldots, 0) = f(x^1)$ for $x^1 \geq 0$ and this immediately shows that (ii) implies that f is C^k. Then the relation $F^1(-x^1) = -F^1(x^1)$ shows that the derivatives of even order vanish, proving (i). Conversely, suppose that (i) holds. The vector field F^a can be written as a sum of two terms, as in the proof of Lemma A1, and the second of these terms can be handled just as in the proof of that lemma. As for the first term, it is the product of $x^a/|x|$ with a polynomial in $|x|$ where only the coefficients of odd powers are non-zero. Hence it can be written in the form $x^a q(|x|^2)$, for some polynomial q.

LEMMA A4. *Suppose that the hypotheses of Lemma A3 are satisfied and that the derivatives of f of order $\leq k$ are bounded. Then the partial derivatives of F^a of order $\leq k$ are bounded and there exists a constant C independent of f such that $\|F^a\|_{C^k} \leq C\|f\|_{C^k}$.*

Proof. The method used in the proof of Lemma A2 also applies in this case.

LEMMA A5. *Let $f : [0,\infty) \to \mathbf{R}$ be a measurable function and define*

$$g_{v,w}(r) = r^{-v} \int_0^r s^w f(s) ds$$

If w and v are integers with $w \geq 0$ and $v \leq w$ and f is C^k then $g_{v,w}$ is C^{k+1}. Moreover, on any interval of the form $[0, R]$ there exists a constant C, independent of f, such that $\|g_{p,q}\|_{C^{k+1}} \leq C\|f\|_{C^k}$.

P r o o f. Write $f = p_k + (f - p_k)$, where p_k is the Taylor polynomial of f of order k at the origin. Now $r^{-v}\int_0^r s^w p_k(s)ds$ can be bounded by a constant times $\|f\|_{C^k}$ under the given assumptions. Thus it can be assumed that without loss of generality that $p_k = 0$. Now:

$$d^l g_{v,w}/dx^l = d^l(r^{-v})/dx^l \int_0^r s^w f(s)ds + \sum h_j d^{j-1} f/dx^{j-1}$$

where h_j is homogeneous of degree $-v + w + j$. Under the assumption that $p_k = 0$ the integral in the first term is bounded by Cr^{k+w+1}, where the constant C only depends on the C^k norm of f. Since the other factor is $O(r^{-v-l})$ this suffices to bound the first term. The other terms can be bounded using the fact that $d^{j-1}f/dx^{j-1}$ is bounded by an expression of the form $C|x|^{k-j+1}$.

Lemma A6. *Let $f : [0, \infty) \to \mathbf{R}$ be a continuous function with $f(0) = 0$ and define*

$$h(r) = r\int_0^r s^{-1} f(s)ds$$

If f is C^k for some $k \geq 1$ then h is C^{k+1}. Moreover, on any interval of the form $[0, R]$ there exists a constant C, independent of f, such that $\|h\|_{C^{k+1}} \leq C\|f\|_{C^k}$.

P r o o f. Since $f(0) = 0$ and f is C^k for some $k \geq 1$ it is possible to write $f(r) = rf_1(r)$ for some C^{k-1} function f_1. Then $h(r) = r\int_0^r f_1(s)ds$. Clearly $r^{-1}h$ is C^k and its C^k norm can be estimated by the C^k norm of f. It follows that h itself is C^{k+1} and that its C^{k+1} norm can be bounded by the C^k norm of f.

Acknowledgements. Many of the ideas reported here arose in the course of my collaboration with Gerhard Rein. I am also indebted to Francis Cagnac for helpful suggestions.

References

[1] S. Alinhac and P. Gérard, *Opérateurs pseudo-différentiels et théorème de Nash-Moser*, InterEditions, Paris, 1991.

[2] R. Bartnik, *The mass of an asymptotically flat manifold*, Commun. Pure Appl. Math. 34 (1986), 661–693.

[3] J. Binney and S. Tremaine, *Galactic dynamics*, Princeton University Press, Princeton, 1987.

[4] G. A. Burnett and A. D. Rendall, *Existence of maximal hypersurfaces in some spherically symmetric spacetimes*, Class. Quantum Grav. 13 (1996), 111–123.

[5] D. Christodoulou, *Violation of cosmic censorship in the gravitational collapse of a dust cloud*, Commun. Math. Phys. 93 (1984), 171–195.

[6] D. Christodoulou, *The problem of a self-gravitating scalar field*, Commun. Math. Phys. 105 (1986), 337–361.

[7] Y. Choquet-Bruhat, *Problème de Cauchy pour le système intégro différentiel d'Einstein-Liouville*, Ann. Inst. Fourier 21 (1971), 181–201.

[8] Y. Choquet-Bruhat and J. York, *The Cauchy problem*, in: General Relativity and Gravitation, Vol. 1 , A. Held (ed.), Plenum, New York, 1980 99–172.

[9] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. 2*, Wiley, New York, 1989.

[10] J. Ehlers, *Survey of general relativity theory*, in: Relativity, Astrophysics and Cosmology, W. Israel (ed.), Reidel, Dordrecht, 1973, 55–89.

[11] R. S. Hamilton, *The inverse function theorem of Nash and Moser*, Bull. Amer. Math. Soc. 7 (1982), 65–222.

[12] S. W. Hawking and G. F. R. Ellis, *The large-scale structure of space-time*, Cambridge University Press, Cambridge, 1973.

[13] P.-L. Lions and B. Perthame, *Propagation of moments and regularity for the three-dimensional Vlasov-Poisson system*, Invent. Math. 105 (1991), 415–430.

[14] A. Majda, *Compressible fluid flow and systems of conservation laws in several space variables*, Springer, New York, 1984.

[15] E. Malec and N. ÓMurchadha, *Optical scalars and singularity avoidance in spherical spacetimes*, Phys. Rev. D50 (1994), 6033–6036.

[16] J. E. Marsden and F. J. Tipler, *Maximal hypersurfaces and foliations of constant mean curvature in general relativity*, Phys. Rep. 66 (1980), 109–139.

[17] K. Pfaffelmoser, *Global classical solutions of the Vlasov-Poisson system in three dimensions for general initial data*, J. Diff. Equations 95 (1992), 281–303.

[18] M. Reed and B. Simon, *Methods of modern mathematical physics*, Academic Press, New York, 1972.

[19] G. Rein, *Generic global solutions of the relativistic Vlasov-Maxwell system of plasma physics*, Commun. Math. Phys. 135 (1990), 41–78.

[20] G. Rein and A. D. Rendall, *Global existence of solutions of the spherically symmetric Vlasov-Einstein system with small initial data*, Commun. Math. Phys. 150 (1992), 561–583. Erratum: Commun. Math. Phys. 176 (1996), 475–478.

[21] G. Rein and A. D. Rendall, *Global existence of classical solutions to the Vlasov-Poisson system in a three dimensional, cosmological setting*, Arch. Rat. Mech. Anal. 126 (1994), 183–201.

[22] G. Rein, A. D. Rendall and J. Schaeffer, *A regularity theorem for solutions of the spherically symmetric Vlasov-Einstein system*, Commun. Math. Phys. 168 (1995), 467–478.

[23] A. D. Rendall, *On the choice of matter model in general relativity*, Approaches to Numerical Relativity, R. d'Inverno (ed.), Cambridge University Press, Cambridge, 1992, 94–102.

[24] A. D. Rendall, *Cosmic censorship and the Vlasov equation*, Class. Quantum Grav. 9 (1992), L99–L104.

[25] A. D. Rendall, *Crushing singularities in spacetimes with spherical, plane and hyperbolic symmetry*, Class. Quantum Grav. 12 (1995), 1517–1533.

[26] D. G. Swanson, *Plasma Waves*, Academic Press, Boston, 1989.

[27] S. L. Shapiro and S. A. Teukolsky, *Relativistic stellar dynamics on the computer. I. Motivation and numerical method*, Astrophys. J. 298 (1985), 34–57.

[28] R. M. Wald, *General Relativity*, Chicago University Press, Chicago, 1984.