Caustics in Gravitational Theories

Damon John Ridgley Swatton

Ph.D.

Faculty of Mathematical Studies

February 1999

UNIVERSITY OF SOUTHAMPTON
FACULTY OF MATHEMATICAL STUDIES

ABSTRACT

DOCTOR OF PHILOSOPHY

CAUSTICS IN GRAVITATIONAL THEORIES

by Damon John Ridgley Swatton

The gravitational collapse of a spherically symmetric, pressure-free dust is an interesting problem in General Relativity for it can lead, under certain initial conditions, to the situation where infinitesimally neighbouring shells approach and cross each other. The curve generated by these points of crossing generate a particular case of a caustic. In the situation where we have matter associated with each shell the density becomes unbounded on the caustic and, in a General Relativistic framework, we have a singularity.

The interest in these types of singularity is two-fold: they present a possible mechanism for galaxy formation and they represent a counter example to both the strong and weak versions of the cosmic censorship hypothesis. In fact, Yodzis and collaborators prove that an event horizon is generated to the future of the initial singularity, so that these types of singularity are naked. If, however, a solution to the field equations describing a spacetime with a caustic represent points that are internal (rather than being excluded as is generally the case for singularities), then this would stop these solutions as being counter examples to the simplest form of the cosmic censorship hypothesis. In addition, this would reinforce the idea that only strong singularities are censored.

The characteristic feature of shell crossing singularities is that at some point the world lines of shells coincide, meaning that the fluid flow vector becomes non-unique. If, however, we lift the geodesics that our shells follow onto the tangent bundle, then the vector tangent to these curves is unique. This indicates that we might be able to use the methods of Rendall and collaborators as a way to obtain existence to a solution of the field equations but, unfortunately, the unbounded nature of the density functions which arises in our formulation precludes this. We are forced, instead, to take the

direct approach and consider solving the equations that model several superimposed dusts.

The critical factor in any existence proof is to determine the shape of the caustic close to the point of cusp formation. In Newtonian theory or General Relativity this becomes the question of whether or not gravity alters the shape that is predicted by the simple cubic which is well known from catastrophe theory. We shall refer to this as the zero gravity solution. In this thesis we present a rigorous investigation of the limiting behaviour of both the Newtonian and General Relativistic pictures, showing in both cases that it can be represented by a similarity solution. We also relate the Newtonian to the Relativistic case. To further our understanding we also investigate the dynamics of the situation by constructing a computer model based on the Relativistic formulation. This numerical solution corroborates the results previously obtained.

In the Newtonian analysis we show that the similarity solution (based on simple scaling transformations) obtained in the limit as we approach the cusp describes unbounded densities on the axis of symmetry. To correct this we suppose that the Newtonian constant G must also be scaled. We find that the solution now obtained in the limit is one where $G = 0$ which describes the zero gravity case. Moreover, if the initial conditions are described by a cubic, then we find that the asymptotic shape of the caustic does not differ from that of the generic caustic. We check for any other, more general transformation group that leaves the Newtonian differential equations invariant whilst reducing to the gravity free equations in this asymptotic limit. The conclusion is that, subject to an arbitrary Galilean transformation, the scaling transformations are the only transformations that fit this description.

A similar analysis is performed with the General Relativistic equations. In this case, to enable asymptotic solutions to exist, we find that c must also be scaled. The result is that the geodesic and conservation of matter equations reduce again to the gravity free case. Thus even in the General Relativistic formulation of caustic formation we have gravity playing no part.

In the latter parts of this thesis, work is presented that goes some way towards an existence proof for the Newtonian problem. We formulate the differential equations using a Lagrangian coordinate system and then discuss the set-up of a contraction mapping proof of existence of the solution to these equations. In the set-up of the existence proof, we prove that the solution must be C^2. We assume that any solution corresponding to $G \neq 0$ cannot deviate from the zero gravity solution by more than a certain parameter which we are able to chose. By considering a small neighbourhood containing the cusp, we write the solution as a double iterated integral in time away from $t = 0$. We find that the integrand is not integrable through the cocaustic thus excluding any proof of existence of an initial value problem using a contraction mapping type of argument. It did, however, prove possible to show existence for a family of solutions parameterised by two arbitrary functions based on using the Arzela-Ascoli theorem. This approach which has been published in collaboration with C.J.S. Clarke.

CONTENTS.

LIST OF FIGURES.

Page

ACKNOWLEDGEMENTS.

Many people have helped me to complete this Ph.D. thesis, all of whom I am indebted to. I would especially like to thank my supervisors, Prof. Chris Clarke and Prof. Ray d'Inverno, for their support and encouragement throughout this research, but particularly Chris for having the patience to answer the same question many times. I would also like thank Stewart and Sarah Swatton for helping to proof read this manuscript and my mother and father who always had faith in me and my ability to complete this thesis.

I am very grateful to the University of Southampton for the scholarship that supported me during my stay at Southampton and also to John Brown, Penny Endersby and Paul Ratcliff for their kindness in allowing me to concentrate on writing this thesis.

CHAPTER 1. THESIS OVERVIEW.

§1.1. Introduction.

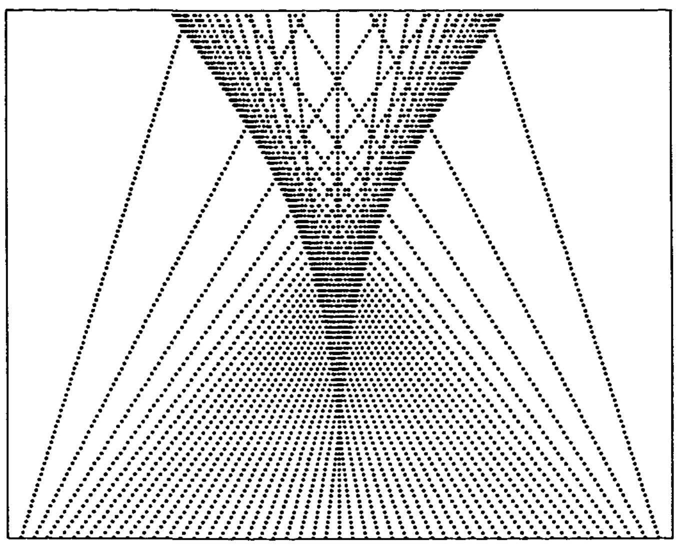
The gravitational collapse of a spherically symmetric dust (pressure-free perfect fluid) is an interesting problem in any gravitational theory for it leads to the formation of two types of singularity known as shell focusing and shell crossing singularities. Shell focusing singularities occur when the dust geodesics focus to a point and are essential in the sense that these singularities cannot be eliminated via an extension of the metric. Shell crossing singularities or *caustics* on the other hand, are non-central and are formed by the piling up of dust trajectories at some finite radius.

The unbounded behaviour of the density in the second example is of particular interest in both cosmology and General Relativity since it provides a model for galaxy formation and points to the existence of naked singularities respectively. Yodzis et al. [YMS] describe what happens in the region external to the caustic but conclude that no analytical continuation can be made through the singularity. The reason for this is that the solution exhibits unphysical behaviour in the sense that pressures become negative, forcing us to consider the proper caustic described by multi-dust spacetimes, [CO] and [C] (ref. §1.3 and chapter 4). Now, although Clarke and O'Donnel have succeeded in showing that an extension through the singularity can be made self consistent, as yet no exact solution exists in either the Relativistic or the simpler, Newtonian case. This thesis presents a rigorous analysis of the equations that describe caustic formation in both theories with the aim being to try to understand the essential physics of caustic formation. We finish by presenting in the final two chapters the beginnings of an existence proof for the Newtonian case. This is completed in a joint paper, [SC].

§1.2. Caustics, cocaustics and tangent bundle surfaces.

As mentioned, caustics represent those regions of unbounded density caused by the crossing of shells. Fig. 1 illustrates what this means. The plane of the paper represents a single spacial coordinate along the horizontal axis and time along the vertical and therefore it can illustrate the General Relativistic description of either a two dimensional spacetime or a four dimensional spacetime with symmetry conditions, or the Newtonian description of one dimensional motion parameterised by time. The lines represent the trajectories of particles and at points where adjacent lines cross, the density becomes unbounded and a singularity forms. By way of notation, we shall call these types of singularities caustics and the point at which the left and the right caustics meet, we shall call the cusp.

Fig. 1. Newtonian caustic.



For realistic reasons, we shall only concern ourselves with higher dimensional spacetimes simplified by assuming certain symmetry conditions: spherical symmetry for General Relativity and planar symmetry for Newtonian theory. To take the first case and relate this to fig. 1 as an example, we impose two dimensional motion in a four dimensional spacetime by assuming that there is no θ or ϕ dependence in any of our dust variables, v, representing the fluid's flow vector and ρ, its density, and that v has only radial and temporal components. With this picture of 'two dimensional gravity' in mind, we realise that the converging lines in fig. 1 represent geodesics followed by shells rather than particles. In addition, since each trajectory is straight, we conclude that the spacetime is Euclidean and that in actual fact, the diagram represents the motion of massless particles.

The fact that Fig. 1 illustrates caustics being formed by particles of zero mass does not preclude it from our discussion. In fact, since exact solutions exist for this situation (catastrophe theory, [A]), a lot of the work we shall present concerning gravity will at least involve the same initial conditions but also, in general, will be based on the assumption that any solution for the gravitational case can be thought of as a perturbation on the solution obtained for G (the gravitational constant) equal to zero. This is important to realise for it becomes an underlying assumption throughout most of this work. Indeed for the final two chapters where we present ideas towards an existence proof, this underlying theme is brought out and made a fundamental assumption upon which the whole theorem will hinge. In other words, we will look for the existence of a solution to the differential equations describing caustic formation that are *near* to that generated when $G = 0$.

In the following chapter we spend time introducing a mathematical description that encompasses caustic formation within the two formulations of gravity as well as any Euclidean spacetime. Without going into much detail, we construct manifolds, M and N, and a map, $f: N \longrightarrow TM$, such that $\pi \circ f(N) = M$ where π is the projection of TM onto M. We suppose that (t, x) and (t, v) represent local coordinates to M and N respectively, and relate the x and v by $\pi \circ f(t, v) = (t, x(t, v))$. So far the t's, x's and v's have no meaning, however, as soon as we place a dust in M, we can begin to interpret these quantities. For M, t and x adopt the labels time and position respectively and if $x_i(t)$ represents a geodesic in M then for N, $v(t, x)$ is defined by $v = v_i(t) = dx_i/dt(t)$ whenever $x = x_i(t)$, and represents the velocity.

The problem we have is that if we allow particle trajectories in M to cross, then we must be prepared to accept the fact that $v(t, x)$ must be multi-valued (corresponding to different geodesics, x_1, x_2 etc. being coincident at (t, x)). It is this behaviour of v that one way or another, encompasses all of the technical problems that are associated with trying to understand caustic formation.

To continue, $f(N)$ represents an embedded surface in the tangent bundle associated with M. We shall construct this surface in such a way so that $f(N)$ is ruled by curves that project down onto geodesics in M. Now, the multi-valued nature of $v(t, x)$ and the continuity of the geodesics in M imply that $f(N)$ looks like a sheet of paper that has been folded twice to make an 'S' shape. The projection of this surface onto M acts in such a way so as to squash these folds into creases, which correspond to the caustic set in M. In terms of our mathematical construction, we have the caustic set corresponding to the image, $\pi \circ f(t, v_c)$, of those points, $(t, v_c) \in N$, where $\pi \circ f$ becomes singular. We will show, in terms of our local coordinates for M and N, that this condition translates into the following statement: the caustic in M corresponds to $x(t, v_c)$ where $v_c(t)$ is the solution to $\partial x/\partial v(t, v) = 0$. We shall call $v = v_c(t)$ the equation of the caustic in TM and its image under π, namely $x = x(t, v_c(t))$, the equation of the caustic in M.

To complete this section we introduce the idea of another curve in $f(N)$, which we call the *cocaustic*. Again, since $v(t, x)$ is multi-valued, it is clear that there will exist points, $(t, v_{cc}, x(t, v_{cc}))$ in $f(N)$, such that $v_{cc} \neq v_c$ but with $x(t, v_{cc}) = x(t, v_c)$. In other words, given $v_c(t)$ we define v_{cc} to be the solution of $x(t, v) = x(t, v_c(t))$ and call $v = v_{cc}(t)$ the equation of the cocaustic. It follows from its definition that the images of the caustic and cocaustic coincide in M and so it is unnecessary to define a cocaustic in M and redundant if we specify the cocaustic to be a curve in TM.

Finally, since the above definitions of the caustic and cocaustic are possibly a little abstract, we illustrate their significance in terms of particles moving on M. We recall that $f(N)$ can be constructed by lifting geodesics on M into the tangent bundle. It follows, therefore, that as particles move along trajectories in M a corresponding point moves along a curve in $f(N)$, which must at some stage cross the caustic and cocaustic. Now refer to fig. 1. As we follow a particular trajectory from the initial time slice, through the external region and into the region to the future of the caustic set, we can say that the lift of this geodesic crosses the cocaustic when the particle on M first crosses the caustic set, and crosses the caustic in TM when the particle 'touches' the caustic in M for the second time. Note that it is only when this particle 'touches' the caustic for the second time does it cross neighbouring geodesics, and so only at these points does the density associated with this particular trajectory become unbounded.

§1.3. Multi-dust regions.

The other thing that we wish to mention before we launch into a description of each chapter is the concept of a multi-dust spacetime. In order to at least attempt to obtain an exact solution for the region in M to the future of the caustic, we have to work with a unique fluid flow vector and for reasons that we have already stated, this is certainly not the case. To solve this problem we are forced to introduce the added complication of a multi-dust region [CO] where we expect shells to cross. In M, this corresponds to a region, bounded by the caustic, where we have several superimposed dusts, each with a *unique* flow vector whose x-component is given by $v_i(t,x)$. This system can be seen to represent a special case of the Einstein-Vlasov equations for a collisionless gas, however, the singular nature of the problem precludes the use of the existence results obtained by Rendall [R].

The equations that illustrate what we mean by this are, for the Newtonian case,

$$\frac{\partial F}{\partial x} = -\sum_{i=1}^{k} \rho_i, \qquad [4.1.1]$$

$$v_i \frac{\partial v_i}{\partial x} + \frac{\partial v_i}{\partial t} = F \qquad [4.1.2]$$

and

$$v_i \frac{\partial \rho_i}{\partial x} + \rho_i \frac{\partial v_i}{\partial x} + \frac{\partial \rho_i}{\partial t} = 0. \qquad [4.1.3]$$

With $k = 1$, equations [4.1.2] and [4.1.3] represent the standard conservation equations for a single dust whereas equation [4.1.1] expresses the Newtonian law of gravity. For $k > 1$ we notice two things, firstly that only variables describing the ith dust appear in the corresponding conservation equations, and secondly, that from the force equation, the gravitational field is dependent on the total density at a point and is constructed by the sum of the three ρ_i. This illustrates what we mean by superimposed dusts, i.e. we model our fluids in such a way so that they are essentially invisible to each other except via their gravitational interaction.

In TM the situation is somewhat different. Because of the now distinct v_i, $f(N)$ can be considered as the union of three parts corresponding to those points, $(t, v_i, x(t, v_i))$, satisfying, $v_1 < -v_c$, $-v_c \leq v_2 \leq v_c$ and $v_c < v_3$. Thus in TM, the multi-dust region corresponds to the union of three disjoint surfaces characterised by those points satisfying, $-v_{cc} \leq v_1 < -v_c$, $-v_c \leq v_2 \leq v_c$ and $v_c < v_3 \leq v_{cc}$. In addition to this, we often find that working in the tangent bundle makes the situation that we are trying to understand a lot clearer. We can express this if we reformulate the above equations in terms of tv coordinates. Equations [4.1.2] and [4.1.3] simply transform into their tv space equivalents and we shall not restate them. The force equation, however, becomes

$$\frac{\partial F}{\partial v_i} = (-1)^{i+1}\sigma_i + \frac{\partial x}{\partial v_i} \sum_{j \neq i} (-1)^{j+1} \left(\sigma_j \left(\frac{\partial x}{\partial v_j} \right)^{-1} \circ \phi_j \right), \qquad [4.5.3]$$

with arguments, (t, v_i). The function, ϕ_j, defined by

$$\phi_j(t, v_i) = \{(t, v_j) \mid v_j \neq v_i,\ x(t, v_i) = x(t, v_j)\}\,,$$

illustrates what we have been trying to say all along: namely that in order for us to determine the force at any point, (t, v_i) we have to be aware of contributions from (t, v_j) $(i \neq j)$ that correspond to points elsewhere on $f(N)$. Understanding $v_j(t, v_i)$ summarises all the problems associated with caustic formation in any gravitational theory

§1.4. Thesis overview.

This thesis is structured in three parts. The first considers the problem of understanding caustics within the Newtonian framework. This constitutes chapters 2–5. The next part (chapters 6 and 7) considers the application of the techniques developed so far to the General Relativistic case. Finally, chapters 8 and 9 present the ground work for an existence proof for the solutions to the Newtonian differential equations.

Chapter 2 formulates the idea of caustic formation in terms of surfaces in the tangent bundle. The argument is of a general nature for it describes the generation of caustics as the projection of this surface onto M in terms of arbitrary m-dimensional manifolds. Gravity is not a requirement, however, we specify how both the Newtonian and General Relativistic pictures sit within this formulation.

Chapter 3 applies the above ideas to the case where caustics are formed in Euclidean spacetimes. This chapter forms an important part in constructing the foundations upon which we build ideas that are used to discuss the more general picture. It first of all describes the conditions that the flow vector must satisfy in order for the cusp of a caustic to be formed. It then proceeds to discover the solution, $x = q(t, v)$, that can be obtained for the case when $G = 0$. This process is equivalent to defining the surface, S_q, in the tangent bundle. Then, by actually projecting S_q onto M, we proceed to determine the shape of the caustic set. It also introduces the reader to the idea of the cocaustic.

Chapter 4 tries to understand the Newtonian formulation of caustics on a spacetime with local coordinates (t, x). That is to say, we investigate the solution of the equations that describe 1-D Newtonian gravity within the context of a multi-dust spacetime exhibiting planar symmetry ([4.1.1]–[4.1.3]). The procedure that we adopt is to look for similarity solutions for the general case and then check our results by setting $G = 0$ and solving for the gravity free scenario. We then repeat this procedure for the Newtonian equations written in terms of (t, v) coordinates.

Chapter 5 takes this analysis further. One of the problems we obtained from using simple similarity solutions is that unbounded densities are predicted on the axis of symmetry. This is clearly unrealistic and to try to solve this, we consider asymptotic solutions of the tx space equations. These type of solutions are based on a 'stretching' or scaling transformation of the form,

$$g(\varepsilon; t, x, F, v_i, \rho_i) = (\varepsilon^{k_t} t, \varepsilon^{k_x} x, \varepsilon^{k_F} F, \varepsilon^{k_v} v_i, \varepsilon^{k_\rho} \rho_i).$$

The idea is to consider a new coordinate system, $(\tilde{t}, \tilde{x})$ say, whose length and time scales increase, as $\varepsilon \longrightarrow \infty$, relative to that of the original and fixed coordinate system, (t, x). During this magnification process, the dependant variables are also scaled by an amount determined by the similarity degrees which were obtained in the previous chapter. The result of this analysis is that a fixed region in our $\tilde{t}\tilde{x}$ coordinate system which contains the cusp, increases in size exponentially so that points relative to the original coordinate system approach the cusp asymptotically. Moreover, the differential equations based on this coordinate system are transformed so that only those terms that are significant during cusp formation remain. We find that in order for us to obtain asymptotic solutions that are bounded on the symmetry axis, the Newtonian constant must also be scaled and we do so in such a way so as to ensure that $\tilde{G}$ asymptotically approaces zero. The implication of this is profound for, as our differential equations transform, we can answer the question, 'does gravity play a role in cusp formation?'

Chapter 6 changes tack for we now begin to consider the General Relativistic formulation of caustics. We start by defining the concept of a multi-dust spacetime within this theory. This follows ideas presented by Clarke and O'Donnel [CO] and essentially replaces each Newtonian equation by its Relativistic analogue. In other words we solve,

$$G^{ij} = -\kappa T^{ij} = -\kappa \sum_{p=1}^{k} T^{ij}{}_{(p)} = -\kappa \sum_{p=1}^{k} \mu_{(p)} v^i_{(p)} v^j_{(p)}, \qquad [6.1.1]$$

$$v^i_{(p);j}v^j_{(p)} = 0 \qquad [6.1.2]$$

and

$$T^{ij}{}_{(p);j} = 0. \qquad [6.1.3]$$

We consider these equations within Synge's formulation of spherically symmetric spacetimes [S] and look for a numerical solution. This involves time evolving a finite set of points that are initially spaced at regular intervals throughout the spacetime using the Euler numerical scheme. These points should be considered as *reference points* within our dust that move along geodesics. The key problem with this approach concerns the definition of the density function. It turns out that the conservation of matter equation ([6.1.3]) allows us to define $\mu_{(p)}$ in terms of conserved masses that can be associated with each reference point. The numerical intricacies are primarily concerned with tracking each geodesic and ensuring that these conserved masses are treated correctly, particularly when trajectories cross. These problems will be described.

Chapter 7 continues our discussion on General Relativistic caustics by looking at firstly the Newtonian limit and secondly at the asymptotic limit of [6.1.1]–[6.1.3]. There are three reasons for doing this work. The first is to check that the Newtonian limit corresponds to a Newtonian formulation using tr coordinates so as to ensure that the two models are consistent. The second reason concerns the asymptotic analysis; we hope to determine, within the context of General Relativity, whether or not gravity plays a role in cusp formation. Again we seek correlation between this and the result obtained in the Newtonian case. The other and final reason is that we expect the processes, take the Newtonian limit and take the asymptotic limit, to commute. The results concerning this particular aspect are surprising for we obtain an unexpected link between the planar symmeric Newtonian problem and the spherically symmetric Relativistic case.

Chapter 8 returns to the Newtonian discussion of caustic formation and begins the construction of an existence proof. We first of all reformulate the Newtonian equations in terms of a comoving or *Lagrangian* coordinate system, (t, X). The reason for doing this is because a remarkable simplicfication takes place: the equations describing the motion of our dust particles get completely decoupled from those that determine how the density functions change with time and a solution for the latter readily presents itself. The next stage is to formulate the solution for $x(t, X)$ in terms of an integral equation, which we write as $x(t, X) = J[x](t, X)$. The idea behind the existence proof is to then show that J is a contraction mapping on some suitable space of functions. We then appeal to the fixed point theorems of such maps to conclude uniqueness.

To proceed with this proof we need to complete the specification of J. In other words we need to define the metric space upon which it acts. This consists of two parts. The first concentrates on determining the exact differentiability of x, the second part involves defining the class of functions in which we look to prove existence. This is where the underlying theme of the general solution being approximately equal to $x = q(t, X)$ is brought out and made the main assumption. In fact, it becomes the main driver behind the specification of the particular metric space that we aim to use.

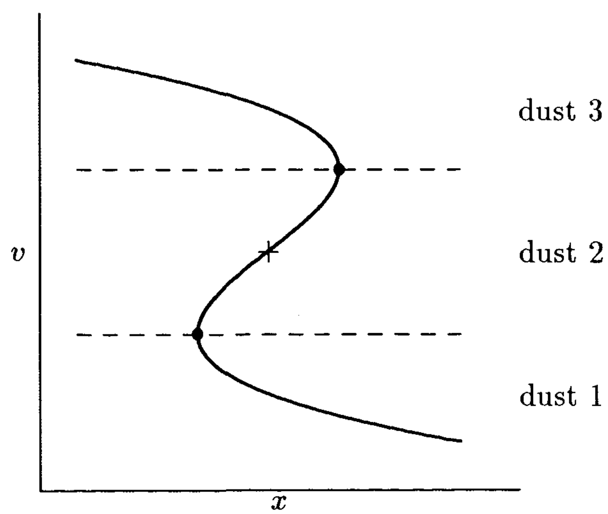
The final chapter takes these ideas further. The final stage in our contraction mapping proof involves showing that J, our candidate, is in fact a contraction. To do this we need to estimate the size of $J[f]$ with respect to the norm that we have previously defined. The only difficulty arises from the second X-derivative and so we concentrate on this aspect of the work. The problem manifests itself as the occurrence of terms that look like $[f'(t, Y(t, X))]^{-1}$ within the integrand of $\partial^2 x/\partial X^2 = \partial^2/\partial X^2 J[x]$. Since these become unbounded as $Y(t, X)$ tends towards the caustic, estimating these quantities then becomes the last stumbling block for our theorem. In this chapter, therefore, we perform this calculation considering all X within the multi-dust region.

CHAPTER 2. CAUSTICS IN GRAVITATIONAL THEORIES.

§2.1. Introduction.

To facilitate the understanding of caustics and how they are generated it is useful to lift the discussion from the spacetime, M, to its tangent bundle. The main reason for this is that the congruence of geodesics that are particular to the type of caustic being formed, generates a surface, $S \subset TM$. This surface gives a clearer understanding of how the dust particles self-interact when gravitational effects are considered.

Fig. 2. Illustration of the surface, $S_q \in TM$, generated by geodesics satisfying $x = q(t, v)$. The + represents the axis of symmetry of the caustic.



The means by which this surface provides this picture can be illustrated by first of all considering the simple example of converging dust particles in zero gravity. Fig. 1 shows the formation of a caustic under these conditions. It represents a spacetime of only two dimensions; position x and time t. It can also be used to represent a spacetime of four dimensions. In this example the y and z directions are suppressed by the assumption that all motion occurs in the x direction and that there is no variation with respect to y or z (planar symmetry). In either case, the spacetime is Euclidean so that all geodesics are straight lines with an associated velocity, $v = dx/dt$. Now, at any point on this diagram there is either one or three geodesics passing through implying a relationship between x, v and t. For the simple caustic (ref. chap. 3), which has slightly different initial conditions to that illustrated in fig. 1 but is still a zero gravity solution, we obtain $x = q(t,v) := vt - v^3$. This equation is defining a surface in the tangent bundle and in order to distinguish this from the general S, we shall denote this by S_q.

Fig. 2 shows this surface, S_q, for some fixed $t > 0$. It illustrates how x varies with velocity on this time slice. Now as is the case for ordinary differential equations, the shape of the caustic must be dependent on the initial conditions. For any zero gravity solution, however, no matter what its initial conditions are, provided a small enough region enclosing the point of caustic formation is considered, a section through its tangent bundle surface will always look like that illustrated by fig. 2 (ref. §3.2). This last statement can be carried over into the case where gravitational effects are included. In fact, for the general case it becomes of fundamental importance as it gives a handle by which an existence theorem for Newtonian caustics can be constructed.

Suppose now a different spacetime is considered with converging dust particles which are allowed to interact via the gravitational force. The effect that this may have on the surface can be thought of in either of two ways. Firstly, that starting from the same initial conditions as those for the simple caustic, the gravitational forces generate a surface that can be thought of as a perturbed form of S_q. Obviously the perturbation becomes greater as t increases so this picture is only valid for early times. Alternatively, that with different initial conditions, a section through the surface is identical to a section through S_q at that time. In either case S_q can be thought of as a *representation* of S at t.

The surface for the general case can now be visualised showing how S clarifies the way in which gravity acts. Consider a point, (t,x), in a spacetime that contains a caustic set. Then knowing S, the velocities, v_i $(i = 1,2,3)$, which represent the trajectories of all particles that are coincident at that point, can be determined. For gravitational theories where a degree of symmetry is involved such as spherically symmetric or planar spacetimes, we can think of the region in M containing dust particles to be composed of a series of shells that move along geodesics. In addition to this we can show that the acceleration of any shell is governed by the integrated mass between the origin and its position at that time. This is where the surface in TM becomes of use for we can now see that the integrated mass between (t,x)

and the symmetry axis is also a function of the velocities v_i. If the velocity at some arbitrary time can be used to parameterise the geodesics in M, then this acceleration is determined by three disjoint regions in TM centred on the trajectories labelled by v_i. It is this concept of different, isolated regions influencing the motion of each particle that determines the complexity of the gravitational interaction.

§2.2. The Σ-formulation of the tangent bundle surface, S.

To understand further the importance of S, how it is generated and how it relates to caustic formation within a spacetime, we suppose the existence of a manifold, N, of dimension $m = \dim(M)$ and a map, f, such that $f\colon N \longrightarrow TM$ with $f(N) = S$. This simple statement gives the overall picture for the general case where a surface is embedded in TM. However, if N and f are constructed in a particular way then a structure to S is given that makes caustic formation a lot clearer.

In order to make this construction, let us suppose that a nowhere zero vector field, Z, on TM and a smoothly embedded surface, Σ, in TM of dimension $m-1$ and transverse to Z, i.e. $Z \notin T^*_{(p,X_p)}\Sigma$, is given. Then the family of integral curves of Z, $C_z\colon \mathbb{R} \longrightarrow TM$, such that $C_z(0) = z \in \Sigma$, defines an immersion, $f\colon \mathbb{R} \times \Sigma \supset N \ni (t,z) \longrightarrow C_z(t) \in TM$, the image of which we call S. A consequence of this particular construction is that the union of all integral curves that pass through Σ, i.e. $\bigcup_{z\in\Sigma}\{C_z(t) \mid t \in \mathbb{R}\}$, is equal to S. This is the picture that we are trying to emphasise and we shall describe this by saying that the surface is *ruled* by the curves C_z. In a moment we shall discuss particular cases of this construction whereby making refinements to the definition of Z, each C_z projects onto a curve in the spacetime which has specific properties. In the meantime, however, we shall complete the definition of f by stating that the converse to the above also holds. That is to say, given Z, a surface, S, of dimension m everywhere tangent to Z and an $m-1$ dimensional subsurface, Σ, transverse to Z, then the integral curves as defined above remain in S.

The general situation specialises to the particular case we wish to consider where dust particles are allowed to move along geodesics in M. This specialisation can be summarised in two steps. The first is the case where for $(p, X_p) \in f(N) \subset TM$, $(\pi_* Z)_{\pi(p,X_p)} = X_p$ and the integral curves of Z project down onto curves in M that are solutions to 2nd order differential equations. The second specialises further to when these integral cures project onto *geodesics* in M. This refinement requires, in addition to the above, that the vertical part of Z in natural coordinates is a homogeneous polynomial of degree 2.

To explain these comments, let us take an arbitrary vector field evaluated at a point $p \in M$, i.e. X_p say. Now p and X_p define a point $(p, X_p) \in TM$ which we can suppose lies on a curve. We let this curve be C_z as defined above so that $C_z(t) = (p, X_p)$ and $Z = C_{z*}d/dt$ has components, $(dC_z^1/dt(t) \ldots dC_z^{2m}/dt(t))$. Now let us consider the restriction on Z that forms the first specialisation mentioned above, namely $(\pi_* Z)_{\pi(p,X_p)} = X_p$. Since local coordinates to M and TM can be $(x^1, \ldots, x^m)$ and $(x^1, \ldots, x^m, y^1, \ldots, y^m)$ respectively, this means that the components of both sides of this equation can be equated to give

$$\sum_{i=1}^{m} Z^i(p, X_p) \left.\frac{\partial}{\partial x^i}\right|_{(p,X_p)} (x^k \circ \pi) + \sum_{j=1}^{m} Z^{j+m}(p, X_q) \left.\frac{\partial}{\partial y^j}\right|_{(p,X_p)} (x^k \circ \pi) = X^k(p)$$

where $k = 1, \ldots, m$. This simplifies to

$$\sum_{j=1}^{m} \frac{dC_z^j}{dt}(t)\delta_j^k = X^k(C_z^1(t), \ldots, C_z^m(t))$$

$$\Longrightarrow \frac{dC_z^k}{dt}(t) = X^k(C_z^1(t), \ldots, C_z^m(t)).$$

The above equation tells us that there exists a one-to-one relationship between the first m components of $Z_{(p,X_p)}$ and those of X_p. Moreover, we have, by the definition of C_z, $C_z^{j+m}(t) = X^j(C_z^1(t), \ldots, C_z^m(t))$ $(j = 1, \ldots, m)$ and thus $C_z^{j+m} = dC_z^j/dt$, meaning that (p, X_p) has local coordinates, $(C_z^1(t), \ldots, C_z^m(t), dC_z^1/dt(t), \ldots, dC_z^m/dt(t))$, and that the vertical part of $Z_{(p,X_p)}$ (i.e. the components $Z^{j+m}(p, X_p)$) is given by $(d^2C_z^1/dt^2(t), \ldots, d^2C_z^m/dt^2(t))$. These components are unrestricted and we could choose many different vector fields that satisfy the above conditions. To express this we introduce arbitrary functions, $f^j(t)$, so that one choice for Z might be such that $d^2C_z^j/dt^2(t) = f^j(t)$. It follows therefore, that since $\pi(C_z(t))$ has local coordinates $(C_z^1(t), \ldots, C_z^m(t))$, we have C_z projecting onto a curve in M, which is a solution to a 2nd order differential equation.

This argument can be continued to include the second step where the projected curves are geodesics. In this case the requirement that the vertical part of Z in natural coordinates is a homogeneous polynomial of degree 2 means that $f^j(t) = -\Gamma^j_{kl} dC^k_z/dt(t) dC^l_z/dt(t)$. The coefficients, Γ^j_{kl}, are as yet undetermined functions of t. Within this definition it follows that $d^2C^j_z/dt^2(t) = -\Gamma^j_{kl}(t) dC^k_z/dt(t) dC^l_z/dt(t)$, which is precisely the geodesic equation.

§2.3. Caustic formation in gravitational theories.

Given the above construction for the surface in the tangent bundle, it is now possible to see how caustics in M are generated by the knowledge of f (or more precisely, the knowledge of C_z). Since f is essentially taking N, folding it twice and embedding the result in TM, it follows that when projecting this surface onto the spacetime we obtain points in N where the map describing this process ($\pi \circ f$) becomes singular[1]. It is the image of these points that form the caustic set. The reason as to why this occurs can be clearly seen if we again use the zero gravity situation as an example. In this case we are basically taking a folded two dimensional surface and mapping this from a three to a two dimensional manifold (we shall see in the next chapter that all points on S_q have local coordinates $(t, x, 1, v)$). The corresponding reduction in dimension means that the folds get pressed into creases, implying a reduction in the degree of differentiability of $(\pi \circ f)^{-1}$ and hence that $\pi \circ f$ is singular.

[1] In this context a function is singular if there exists a point where at least one of its derivatives is non-invertible.

CHAPTER 3. CAUSTICS IN EUCLIDEAN SPACETIMES.

§3.1. Introduction.

In the previous chapter the surface in the tangent bundle was constructed for the most general case and it was shown how this specialises to when the curves which rule S project onto geodesics in M. These ideas are also applicable to spacetimes where the geometry is Euclidean corresponding to zero gravitational force.

To put this within the context of the previous chapter's definition of S, we need to stipulate a further refinement on Z. This requirement is, quite simply, that Z must satisfy $(\pi_* Z)_{\pi(p,X_p)} = X_p$ and $Z^{i+m}(p, X_p) = 0$ $(i = 1, \ldots, m)$. If C_z represents some integral curve of Z as defined in the previous chapter, then the first constraint implies that points along this curve have local coordinates which look like $(C_z^1(t), \ldots, C_z^m(t), dC_z^1/dt(t), \ldots, dC_z^m/dt(t))$. The second states that $d^2C_z^i/dt^2(t) = 0$. Since $\pi \circ C_z$ is a curve in M with local coordinates, $(C_z^1(t), \ldots, C_z^m(t))$, this is equivalent to saying that the connection on M is flat or that the geodesics are straight lines, implying a Euclidean geometry.

In order for these straight lines to produce caustics the initial conditions need to be determined. Of course this amounts to defining the surface in TM. Now there are an infinite number of different caustic types that can be generated depending on how we define the initial conditions. An example that is of particular importance, because of its relevance in the discussion of Newtonian caustics encountered later on, is that where the surface, $S_q \subset TM$, defined by $x = q(t, v)$ generates caustics on a manifold that has planar symmetry with respect to the plane described by $x = 0$. For this reason, we will concentrate this chapter on discussing S_q in some detail and finish by using it as an example to show how, in general, the surface in TM generates caustics in M.

To begin this discussion S_q must be derived and the next two sections will be dedicated to doing this. The aim here will be to provide an indication that this simple surface generates caustics rather than a formal proof. In the first section we will start with the flow and obtain the surface. That is to say, an approximation to S_q will be obtained based on certain assumptions regarding the geodesic congruence in M. The second section will start with the surface and obtain the flow. This means that given S_q, we can prescribe a vector field, Z, on TM which is everywhere tangent to this surface so that a subset of its integral curves lie in S_q. It will then be shown how Z can be related to straight lines in M using the ideas discussed in the early parts of this introduction.

In the final section of this chapter we will demonstrate how caustics in M are generated by the knowledge of f. This will be illustrated by using the caustics produced by S_q as an example. Since in this case both Σ (which we shall identify with $\mathbb{R}$ via the v coordinate) and the immersion, $f_q : \mathbb{R} \times \Sigma \subset N \longrightarrow TM$, can be defined, the function $\pi \circ f_q$ is known. Hence it becomes a relatively simple matter to determine the set $N_c \subset N$ and its image, $\pi \circ f_q(N_c)$, such that $\pi \circ f_q$ is singular. In terms of local coordinates this process is defining two relationships; one between t and v known as the equation of the caustic in TM, and the other between t and x which we call the equation of the caustic in M. By way of providing the reader with an indication as to the shape of these curves for the general case, and also because we shall refer to this in the latter part of this thesis, we shall perform these calculations and determine the aforementioned equations.

Within the tangent bundle, the caustic set becomes of greater significance for it defines a set of points $N_{cc} \subset N$ such that $N_{cc} \cap N_c = \emptyset$ known as the cocaustic. By definition, each point n_{cc} in this new set corresponds to a single point $n_c \in N_c$ via the relation $\pi \circ f(n_{cc}) = \pi \circ f(n_c)$. Thus, working locally in TM it is possible to obtain a different relation between t and v for the case when geodesics in M generate S_q. We shall call this relation the equation of the cocaustic. Since this plays a role in the latter part of this thesis where an existence proof for the Newtonian equations of motion is considered, this will also be done.

§3.2. Starting with the flow and getting the surface.

In this section an approximation to the equation defining S_q based on certain assumptions regarding the geodesic flow will be given. These assumptions allow us to determine which straight lines produce caustics in a Euclidean spacetime. It will be supposed, in the first instance, that geodesics in M are parameterized by time. Hence, the relationship that determines which points lie on a particular line is quite simply

$x = vt + x_0$. On any time slice the position of points which lie on these lines varies as a function of v. Thus instead we can consider $x = vt + x_0(v)$ which corresponds to a continuum of curves rather than a single one characterised by v and x_0. At this stage $x_0(v)$ can be an arbitrary function, however, if $x_0(v)$ is smooth then a surface is generated in TM according to the previous equation. It follows that the problem of finding an equation which defines S_q becomes one of finding which initial conditions generate caustics in M. The next restriction imposed on the geodesic congruence allows this to be done.

We suppose that there exists a point $p \in M$ such that the velocity field satisfies $\partial v/\partial x \neq 0$, $\partial^2 v/\partial x^2 = 0$ and $\partial^3 v/\partial x^3 \neq 0$. These constraints are the criteria for cusp formation and are applicable in the general case where gravitational effects are considered. To understand what these equations are saying let us consider, as an example, fig. 1 which is an illustration of caustic formation in the zero gravity case. Now the most obvious thing is that in a region surrounding the cusp all trajectories have either crossed or are travelling towards the axis of symmetry. The rate of this convergence is determined by the velocity of each particle. However, a single particle cannot produce caustics on its own meaning that it is not the velocity that determines caustic formation. Instead, we need to consider how v varies with respect to neighbouring trajectories. Specifically, we are interested in how this rate of change of convergence varies as we hop from one line to its neighbour on some chosen time slice.

The rate of change of v with respect to x represents this important quantity and having identified this as such, it is now possible to discuss certain attributes we must assign it for caustic formation to occur. Let us consider the time slice that intersects the caustic at its cusp and denote this point of intersection to be p. We can see from fig. 1 that neighbouring geodesics at the cusp must converge. If this is not the case then a cusp will not be formed and the caustic set will constitute two intersecting curves in M which are symmetric about the time axis. This implies that $\partial v/\partial x \neq 0$ at p. Secondly, $\partial v/\partial x$ must increase as x increases or decreases away from the spatial origin if it is required that three adjacent trajectories are to cross near a common point. This implies that $\partial v/\partial x$ must have a minimum at p and so $\partial^2 v/\partial x^2 = 0$.

These are the two conditions required for a caustic to be formed and as has been explained there is a clear physical meaning to each of the above statements. There is a third condition that the velocity field must satisfy and that is $\partial^3 v/\partial x^3 \neq 0$. This ensures that any perturbation on $v(x)$ also satisfies the above conditions allowing for the fact that the position of p may change slightly. To explain, suppose the converse is true and that $v(x)$ satisfies the first two of the conditions with $\partial^3 v/\partial x^3 = 0$. This implies a turning point with respect to x in $\partial v/\partial x$ at p. If v' is a perturbation on v then even though $\partial v'/\partial x \neq 0$ and $\partial^2 v'/\partial x^2 = 0$ at p' close to p, it is *possible* that $\partial^3 v'/\partial x^3$ is non-zero at p'. This means that $\partial v'/\partial x$ has a maximum or minimum at p' rather than a turning point and the graphs of these two functions, v and v', will then be very different. Mathematically speaking, the non-vanishing of the third derivative is necessary for all functions, v' say, sufficiently close to v in any C^2 topology to satisfy the above conditions on its first and second derivatives if they hold for v itself.

To begin the derivation of S_q let us consider the point p such that $\partial v/\partial x \neq 0$, $\partial^2 v/\partial x^2 = 0$ and $\partial^3 v/\partial x^3 \neq 0$. Now by a simple change of coordinates this point can be made to occur at $x = t = 0$. In addition, a Galilean transformation allows us to consider a co-ordinate system that is stationary with respect to the cusp. This means that at the origin, $v = 0$. Now let us consider the time slice $t = 0$. Clearly the velocity of those particles on this time slice varies as a function of x_0 and so close to the cusp, $v(x_0) = \alpha x_0 + \beta x_0^3 + O(x_0^4)$. The inverse function can also be written as a Taylor approximation and this is given by $x_0(v) = a + bv + cv^2 + dv^3 + O(v^4)$. Combining these two relationships gives,

$$(\alpha a + \beta a^3) + (\alpha b + 3\beta a^2 b - 1)v + (\alpha c + 3\beta a^2 c + 3\beta a b^2)v^2 \\ + (\alpha d + 3\beta a^2 d + 6\beta abc + \beta b^3)v^3 + O(v^4) = 0.$$

and since this is true for all v, we can suppose that the coefficients of v vanish. Taking the first term, $\alpha a + \beta a^3 = 0$ implies that $a = 0$ or $a = \sqrt{-\alpha/\beta}$. The condition that v is zero at the origin means that the second solution must be discarded. Thus with $a = 0$ the above equation becomes

$$(\alpha b - 1)v + \alpha c v^2 + (\alpha d + \beta b^3)v^3 + O(v^4) = 0,$$

which implies that $b = 1/\alpha$, $c = 0$ and $d = -\beta/\alpha^4$. The initial conditions are then $x_0 = v/\alpha + \beta v^3/\alpha^4 + O(v^4)$ so that $x = vt - v/\alpha - \beta v^3/\alpha^4 + O(v^4)$. Finally we may change coordinates in order to simplify this equation. If $t \longrightarrow \beta/\alpha^4 t - 1/\alpha$ and $x \longrightarrow \beta/\alpha^4 x$ then dropping all terms of order higher than 3 finally results in $x = vt - v^3$.

§3.3. Starting with the surface and obtaining the flow.

In this section one will show that S_q is ruled by curves which produce Newtonian geodesics in the spacetime under the projection map, π. The argument to be adopted is quite concise for it uses ideas that were discussed in the introduction concerning the Σ-formulation of S. Firstly S_q will be defined in terms of local coordinates. Then, a vector field, Z, will be prescribed on TM and we will show that Z is everywhere tangent to this surface. This simple calculation demonstrates that a subset of its many integral curves rule S_q. The next part will show that for some point $(p, X_p) \in S_q$, Z satisfies $(\pi_* Z)_{\pi(p,X_p)} = X_p$. This result coupled with the fact that the vertical components of Z are zero, implies that those curves through (p, X_p) project onto straight lines in M.

The previous section derived S_q given assumptions on the geodesic congruence in M. With this result we can write S_q as the set of points in TM having local coordinates $(t, vt - v^3, 1, v)$ for all $t, v \in \mathbb{R}$. The fact that all points in S_q have local coordinates of the form $(t, x, 1, v)$ is a direct consequence of the assumption that geodesics in M are parameterized by the time coordinate. We say that S_q lies in a reduction of the tangent bundle to $T^1M = \{(p, X_p) | p \in M,\ X_p \in T_pM, \text{ and } X^1(p) = 1\}$. To see this we suppose that $C: \mathbb{R} \longrightarrow M$ is a curve with local coordinates $(t, x(t))$. Then for each t, $C(t)$ defines a tangent vector,

$$\begin{aligned} Y_{C(t)} &= C_* \left.\frac{d}{dt}\right|_t \\ &= \left.\frac{\partial}{\partial t}\right|_{C(t)} + \frac{dx}{dt}(t) \left.\frac{\partial}{\partial x}\right|_{C(t)}, \end{aligned}$$

so that points $(C(t), Y_{C(t)}) \in TM$ have local coordinates $(t, x(t), 1, dx/dt(t))$ implying that they also lie in T^1M.

An alternative formulation is to suppose that

$$S_q = \{(p, X_p) \mid (p, X_p) \in T^1M \text{ and } g(p, X_p) = 0\}$$

where $g: TM \longrightarrow \mathbb{R}$ is defined by $g(t, x, 1, v) = x - q(t, v)$. This is a far more useful definition for it allows us to proceed quite easily with defining Z and proving that it is tangent to S_q. We do this by first of all noting that the one-form, $dg(p, X_p)$, is in the nullifier of the tangent space to the surface in TM: its components are exactly those of ∇g in normal coordinate geometry. Next we choose Z to be the vector field $\partial/\partial t + v\partial/\partial x$. With dg and Z defined we perform the contraction $dg(Z)$ and show that this equals zero, viz.,

$$\begin{aligned} dg(Z) &= Zg \\ &= \partial_t(x - vt + v^3) + v\partial_x(x - vt + v^3) \\ &= -v + v \\ &= 0. \end{aligned}$$

This calculation shows that $Z \in T_{(p,X_p)}S^*$ and that therefore a subset of all its integral curves rule this surface.

Let us now complete this section and show how Z is related to straight lines in M. This will essentially explain in mathematical terms why $Y_{C(t)}$ and $Z_{(C(t),Y_{C(t)})}$ have the same components if C is a straight line in M with coordinates $(t, vt + x_0)$ say. We shall prove that if $(p, X_p) \in T^1M$ then $(\pi_* Z)_{\pi(p,X_p)} = X_p$ so that with the fact that the vertical part of Z is zero, we can conclude that integral curves of Z project onto straight lines in M. The calculation goes as follows; if p and (p, X_p) have coordinates (t, x) and $(t, x, 1, v)$ respectively then locally,

$$\begin{aligned} (\pi_* Z)_{\pi(p,X_p)} &= \partial_t(t, x, 1, v) \left.\frac{\partial}{\partial t}\right|_p + v\partial_x(t, x, 1, v) \left.\frac{\partial}{\partial x}\right|_p \\ &= \left.\frac{\partial}{\partial t}\right|_p + v \left.\frac{\partial}{\partial x}\right|_p \\ &= X_p \end{aligned}$$

as required.

§3.4. The caustics produced by S_q.

In this section we shall demonstrate how the surface in the tangent bundle generates caustics in M for the general case by considering S_q as an example. Since in the case of zero gravity those curves in TM which project onto this kind of singularity in M are known, both Σ and f_q can be determined. This means that the function $\pi \circ f_q$ can be written down explicitly allowing us to find its singular points in terms of local coordinates to either M or N. The result is two equations, one in terms of local coordinates to N and the other in terms of local coordinates to M. Both of these represent the equation of the caustic, and since these will be used extensively throughout this thesis, we shall take the time to derive them.

In N, the set of singular points define a different set known as the cocaustic. This was defined in the introduction and by way of an illustration as to what this set looks like in the case where gravitational effects are considered, this set will be determined for S_q. Since $\pi \circ f_q$ is known and an equation for the caustic has been determined, a relation between t and v, local coordinates to N, representing the equation of the cocaustic can be determined and this will also be done.

In order to begin we must first define $\pi \circ f_q$. Now from results presented in the previous section it follows that an arbitrary point $z \in \Sigma \subset TM$ has local coordinates $(0, -v^3, 1, v)$. Thus Σ corresponds to the $t = 0$ time slice through S_q. This is a one-dimensional surface in TM parameterized by v and so it follows that locally, $f_q\colon \mathbb{R} \times \Sigma \supset N \ni (t, v) \longrightarrow (t, vt - v^3, 1, v) \in TM$. This implies that $\pi \circ f_q(t, v) = (t, vt - v^3)$.

To find the singular points of $\pi \circ f_q$ we follow standard practice and construct its Jacobian matrix. This is given by

$$\begin{pmatrix} 1 & v \\ 0 & t - 3v^2 \end{pmatrix}.$$

Clearly this matrix has rank less than 2 whenever

$$t = 3v^2 \qquad [3.4.1]$$

(or more generally, $\partial x / \partial v = 0$), which implies that points in N with local coordinates that satisfy this equation correspond to points where $\pi \circ f_q$ is singular. Furthermore, since S_q is a two dimensional surface parameterized by t and v, this relation tells us which points on this surface project onto caustics in M. For this reason we shall call this the equation of the caustic in TM.

To find the points in M which lie on the caustic we can simply find the image under $\pi \circ f_q$ of those points in N which satisfy $t = 3v^2$. So, inserting $v = \pm\sqrt{t/3}$ into the above expression for $\pi \circ f_q$ results in an equation relating t and x by $x = q(t, \pm\sqrt{t/3})$. We find that

$$\begin{aligned} x &= \pm t\left(\frac{t}{3}\right)^{1/2} \mp \left(\frac{t}{3}\right)^{3/2} \\ &= \pm t^{3/2}\left(\frac{3-1}{3^{3/2}}\right) \\ &= \pm\frac{2}{3^{3/2}}t^{3/2}, \end{aligned} \qquad [3.4.2]$$

which we shall call the equation of the caustic in M.

Finally, to complete this section we shall determine the equation relating points on the cocaustic in TM. Again, working in N we see that $\pi \circ f_q$ is not one-one. That is to say there are points on N other than those that satisfy $t = 3v^2$ which map onto the caustic set in M. To find these points we need to solve $\pi \circ f_q(n_{cc}) = \pi \circ f_q(n_c)$ in order to find n_{cc} given that n_c has local coordinates which satisfy the equation of the caustic in TM. This amounts to solving $q(t, k_c^{-1}(t)) = q(t, k_{cc}^{-1}(t))$ where $t = k_c(v)$ and $t = k_{cc}(v)$ are the equations of the caustic and cocaustic respectively. Inserting our relation, $k_c^{-1}(t) = \pm(t/3)^{1/2}$ into the left hand side of this equation we obtain a cubic, $vt - v^3 \mp 2(t/3)^{3/2} = 0$, where now, with an abuse of notation, v represents the velocity of the geodesic that intersects the cocaustic at time t, i.e. $v = k_c^{-1}(t)$. We can solve this equation using the following known algebraic recipe.

Suppose that $x^3 + a_1x^2 + a_2x + a_3 = 0$ and define $Q = (3a_2 - a_1^2)/9$, $R = (9a_1a_2 - 27a_3 - 2a_1^3)/54$, $S = R + \sqrt{Q^3 + R^2}$ and $T = R - \sqrt{Q^3 + R^2}$. Then the solutions to this general cubic equation are:

$$x_1 = S + T - \frac{1}{3}a_1,$$

$$x_2 = -\frac{1}{2}(S+T) - \frac{1}{3}a_1 + \frac{1}{2}i\sqrt{3}(S-T)$$

and

$$x_3 = -\frac{1}{2}(S+T) - \frac{1}{3}a_1 - \frac{1}{2}i\sqrt{3}(S-T).$$

Applying this formula to our equation we obtain $Q = -t/3$, $R = -x/2$, $S = \mp(t/3)^{1/2}$ and $T = \mp(t/3)^{1/2}$ so that $v_1 = \mp 2(t/3)^{1/2}$ and $v_2 = v_3 = \pm(t/3)^{1/2}$ are the solutions we are looking for. The relations containing v_2 and v_3 can be identified as describing points that lie on the caustic. The fact that two solutions correspond to the caustic is to be expected for the curve $x = \pm 2(t/3)^{3/2}$ must be tangent to S_q at t because $\partial v/\partial x$ is unbounded for points on the caustic in TM. The solution corresponding to v_1 therefore corresponds to points in N (or equivalently TM) which we know as the cocaustic. Thus $t = 3/4v^2$ and we shall call this the equation of the cocaustic in TM.

CHAPTER 4. NEWTONIAN CAUSTICS.

§4.1. Introduction.

In order to fully understand caustics within the framework of General Relativity, it is helpful to consider the simpler case of converging dust particles in Newtonian theory. The idea is that by considering caustic formation in the low velocity limit of General Relativity we might be able to gain an insight into how the gravitational interaction behaves in a more general system. To begin we must formulate the problem that we wish to solve. We shall consider a spacetime that has planar symmetry with respect to the plane $x = 0$, and place in it a dust which can move according to the acceleration prescribed by Newton's law. This means that particles within our four dimensional spacetime move in groups known as shells. In general these shells can be thought of as volumes with an infinitesimally small thickness and whose shape is determined by the symmetry of our manifold, M. In the case we are considering and in the zero gravity case described in the previous chapter, these shells are planes described by $x = x(t)$ and having thickness dx, moving towards or away from the plane $x = 0$.

As in fig. 1, if we choose our initial conditions carefully, caustics are formed which act as boundaries between two regions. Clearly, to the past of the caustic set there is a region where all geodesics are non-intersecting. This means that the velocity field as a function of t and x is well defined and we can therefore model this region in terms of a single dust. The region to the future of the caustic, however, is far more complicated because particle trajectories are now allowed to cross. If we were to model this situation using a single dust then clearly the velocity field would be multi-valued at any point in this region. To resolve this problem we will adopt the approach taken by Clarke and O'Donnell [CO]. This means that we shall split the dust into several parts and consider a region that contains a number of superimposed dusts. This insures that v_i $(i = 1, \ldots, k)$, which represents the velocity field for the ith dust, is now unique at any point to the future of the caustic. We shall call this part of our spacetime the multi-dust region.

In the case of a multi-dust spacetime the equations governing the motion of the previously mentioned shells are

$$\frac{\partial F}{\partial x} = -\sum_i \rho_i, \tag{4.1.1}$$

$$v_i \frac{\partial v_i}{\partial x} + \frac{\partial v_i}{\partial t} = F \tag{4.1.2}$$

and

$$\rho_i \frac{\partial v_i}{\partial x} + v_i \frac{\partial \rho_i}{\partial x} = -\frac{\partial \rho_i}{\partial t}, \tag{4.1.3}$$

where $i = 1, \ldots, k$. These equations are simply the Newtonian law of gravity, the conservation of momentum equation and the conservation of mass equation respectively. Their solution, which we look for, defines time parameterised geodesics in M. Here F represents the gravitational force per unit mass, v_i represents the three velocity fields associated with each dust such that $v_1 \leq -v_c \leq v_2 \leq v_c \leq v_3$ and ρ_i is the density associated with that velocity. The quantity v_c represents the velocity of the particle whose trajectory is tangent to the caustic at time t. These equations, along with boundary conditions for each dependant variable, can be thought of as defining the concept of a multi-dust spacetime. To explain what we mean by this we notice two things: firstly that only v_i and ρ_i appear in the equations defining the flow of the ith dust and secondly that F depends on the sum of all densities. This means that the dusts interact only via gravity and in this sense can be thought of as being superimposed. The last thing to mention is the choice of k. Following on from our hypothesis that S looks like S_q, we assume that $\pi\colon f(N) \longrightarrow M$ is a simple fold catastrophe (as is $\pi(f_q(N))$). This implies that $k = 3$.

Of course we must now concern ourselves with *joining conditions* which describe how regions with different values of k join. This, in general, can be quite intricate because of two reasons. The first reason is that any point, p say, on the caustic in M represents the end point of two geodesics as well as the initial point of two different geodesics. Specifically, for the left caustic in M, any point such as p is the end point to a geodesic with velocity v say, which originates in the 1-dust region, and an initial point of a geodesic with velocity v_3 say, which proceeds into the 3-dust region. These trajectories must be joined in the correct manner for they represent a particle being transferred from one dust to another. In addition to this we have the other case where p also represents the end point of a 3-dust geodesic with velocity

v_1 say, and the initial point of a different 3-dust geodesic with velocity v_2 say. Again these represent particle trajectories and must be joined in the right way. We achieve this by prescribing boundary conditions for these velocities at the caustic.

The second reason concerns the boundary conditions for the ρ_i. At any point on the caustic we have the situation where neighbouring geodesics converge and eventually cross. This means that the density becomes infinite and the mass flux is not defined at the caustic in M. We can solve this particular problem if we can find a quantity that tends towards zero as we approach the caustic. If we can do this then the mass flux can be defined as the limit of the product of the density with this quantity. The velocity component transverse to the caustic set tends to zero as any particle approaches the caustic. This quantity therefore represents a possible candidate with which we can define the mass flux. Alternatively, we can lift the problem to TM where the mass flux is well defined. We shall discuss how to do this in the relevant section (ref. §4.5.).

In the next section we shall define, in general terms, the concept of a similarity solution. The tools developed will be used to transform equations [4.1.1]–[4.1.3] into ordinary differential equations. The hope is that with this simplification the problem might be soluble. Having done this we then consider the case where the gravitational force is switched off in order to try to recover the solution $x = q(t, v)$, described in the previous chapter. It will, however, take the discussion further for we shall consider how the density functions vary across M. Following this we shall discuss reformulating the problem in terms of (t, v_i) coordinates rather than the (t, x) coordinates of M. This is equivalent to working in the tangent bundle. Finally, we finish by seeing how the criteria for cusp formation (ref. §3.2.) allows the solution $x = q(t, v)$ to emerge naturally from the more general solution that we obtain describing the zero gravity case.

§4.2. Similarity solutions.

When modelling physical systems analytically it is often the case that we must choose a simplified version of the reality that we are interested in. In most cases we look for possible symmetries that we can impose on our system. This means that under certain transformations the differential equations that govern the system we are interested in remain unchanged in form. In other words the equations are identical in either $\mathbf{xf}$ or $\tilde{\mathbf{x}}\tilde{\mathbf{f}}$ coordinates where, given an $\varepsilon \in \mathbb{R}$, $(\tilde{\mathbf{x}}, \tilde{\mathbf{f}}(\tilde{\mathbf{x}})) = g(\varepsilon; \mathbf{x}, \mathbf{f}(\mathbf{x}))$. We say that the differential equations are *invariant* under such a g. The above transformation actually represents a set of transformations and to clarify, we shall list them. We have, for n coordinates, x_i, and m dependant variables, f_j,

$$\tilde{x}_i = g_{x_i}(\varepsilon; \mathbf{x}) \quad \text{and} \quad \tilde{f}_j(\tilde{\mathbf{x}}) = g_{f_j}(\varepsilon; \mathbf{x}, \mathbf{f}(\mathbf{x})).$$

To describe what we mean by invariance, we notice that the Newtonian equations of motion given in [4.1.1]–[4.1.3] are invariant under Galilean transformations. In the following analysis we shall not exploit this symmetry, however, since it is a simple example, we shall use this as an illustration. Suppose that O and $\tilde{O}$ are the origins of two coordinate systems such that a point relative to O has coordinates (t, x) and relative to $\tilde{O}$ has coordinates $(\tilde{t}, \tilde{x})$. If $\tilde{O}$ moves with a velocity ε in the negative x direction with respect to O then we have the relations, $\tilde{t} = t$ and $\tilde{x} = x + \varepsilon t$. Furthermore, due to the relative motion, we have $\tilde{v}_i = v_i + \varepsilon$. Since there is no relative acceleration $\tilde{F} = F$ and because the density is simply a scalar field, $\tilde{\rho}_i = \rho_i$. Our transformation function, g, can then be represented by

$$\tilde{t} = g_t(\varepsilon; \mathbf{x}) = t, \quad \tilde{x}_i = g_{x_i}(\varepsilon; \mathbf{x}) = x_i + \varepsilon t,$$

$$\tilde{F}(\tilde{\mathbf{x}}) = g_F(\varepsilon; \mathbf{x}, F(\mathbf{x}), v_i(\mathbf{x}), \rho_i(\mathbf{x})) = F(\mathbf{x}),$$

$$\tilde{v}_i(\tilde{\mathbf{x}}) = g_{v_i}(\varepsilon; \mathbf{x}, F(\mathbf{x}) v_i(\mathbf{x}), \rho_i(\mathbf{x})) = v_i(\mathbf{x}) + \varepsilon,$$

$$\tilde{\rho}_i(\tilde{\mathbf{x}}) = g_{\rho_i}(\varepsilon; \mathbf{x}, F(\mathbf{x}), v_i(\mathbf{x}), \rho_i(\mathbf{x})) = \rho_i(\mathbf{x})$$

and these define Galilean transformations. Under g, [4.1.1] becomes

$$\frac{\partial \tilde{F}}{\partial \tilde{x}} \frac{\partial \tilde{x}}{\partial x} = -\sum_i \tilde{\rho}_i$$

$$\Longrightarrow \frac{\partial \tilde{F}}{\partial \tilde{x}} = -\sum_i \tilde{\rho}_i,$$

equation [4.1.2] becomes,

$$(\tilde{v}_i - \varepsilon) \frac{\partial}{\partial \tilde{x}} (\tilde{v}_i - \varepsilon) \frac{\partial \tilde{x}}{\partial x} + \frac{\partial}{\partial \tilde{x}} (\tilde{v}_i - \varepsilon) \frac{\partial \tilde{x}}{\partial t} + \frac{\partial}{\partial \tilde{t}} (\tilde{v}_i - \varepsilon) \frac{\partial \tilde{t}}{\partial t} = \tilde{F}$$

$$\Longrightarrow (\tilde{v}_i - \varepsilon)\frac{\partial \tilde{v}_i}{\partial \tilde{x}} + \frac{\partial \tilde{v}_i}{\partial \tilde{x}}\varepsilon + \frac{\partial \tilde{v}_i}{\partial \tilde{t}} = \tilde{F}$$

$$\Longrightarrow \tilde{v}_i\frac{\partial \tilde{v}_i}{\partial \tilde{x}} + \frac{\partial \tilde{v}_i}{\partial \tilde{t}} = \tilde{F},$$

and finally, [4.1.3] becomes

$$\tilde{\rho}_i\frac{\partial}{\partial \tilde{x}}(\tilde{v}_i - \varepsilon)\frac{\partial \tilde{x}}{\partial x} + (\tilde{v}_i - \varepsilon)\frac{\partial \tilde{\rho}_i}{\partial \tilde{x}}\frac{\partial \tilde{x}}{\partial x} = -\frac{\partial \tilde{\rho}_i}{\partial \tilde{t}}\frac{\partial \tilde{t}}{\partial t} - \frac{\partial \tilde{\rho}_i}{\partial \tilde{x}}\frac{\partial \tilde{x}}{\partial t}$$

$$\Longrightarrow \tilde{\rho}_i\frac{\partial \tilde{v}_i}{\partial \tilde{x}} + (\tilde{v}_i - \varepsilon)\frac{\partial \tilde{\rho}_i}{\partial \tilde{x}} = -\frac{\partial \tilde{\rho}_i}{\partial \tilde{t}} - \frac{\partial \tilde{\rho}_i}{\partial \tilde{x}}\varepsilon$$

$$\Longrightarrow \tilde{\rho}_i\frac{\partial \tilde{v}_i}{\partial \tilde{x}} + \tilde{v}_i\frac{\partial \tilde{\rho}_i}{\partial \tilde{x}} = -\frac{\partial \tilde{\rho}_i}{\partial \tilde{t}}.$$

By comparison with [4.1.1]–[4.1.3], the above differential equations clearly have the same form. This is what we mean by the invariance of a differential equation under a given set of transformations. It can be described mathematically if we define $F(x_i, f_j, p_j^{i_1}, \ldots, p_j^{i_1\ldots i_m}) = 0$, where $p_j^{i_1\ldots i_m} = \partial^m f_j/\partial x_{i_1}\ldots\partial x_{i_m}$, as a representation of our mth order partial differential equation. Then, by invariance we mean that

$$F(\tilde{x}_i, \tilde{f}_j, \tilde{p}_j^{i_1}, \ldots, \tilde{p}_j^{i_1\ldots i_m}) = A(\varepsilon)F(x_i, f_j, p_j^{i_1}, \ldots, p_j^{i_1\ldots i_m}) \qquad [4.2.1]$$

where the arguments on the left hand side are as defined by g and A is an arbitrary function.

Let us now continue our construction of the transformations denoted collectively by g. We shall again consider an arbitrary transformation function so as to keep the discussion general, however, in the next section the ideas that we shall develop here will be applied directly to equations [4.1.1]–[4.1.3] together with a new symmetry transformation. Now, the fact that $\varepsilon \in \mathbb{R}$ implies that we have, in actuality, a 1-parameter family of transformations. This set can be given a group structure under the composition of maps, i.e. $g(\varepsilon_1; g(\varepsilon_2; \mathbf{x}, \mathbf{f}(\mathbf{x}))) = g(\varepsilon_1\varepsilon_2; \mathbf{x}, \mathbf{f}(\mathbf{x}))$, with the identity corresponding to $\varepsilon = 1$, i.e. $(\mathbf{x}, \mathbf{f}(\mathbf{x})) = g(1; \mathbf{x}, \mathbf{f}(\mathbf{x}))$. We can now ask how this affects our prototype solution, $\mathbf{f}$. If the form of our differential equations remains the same under g then this implies that our solution must also be unchanged in form under g. In mathematical terms, this means that

$$\mathbf{f}(\tilde{\mathbf{x}}) = g_{\mathbf{f}}(\varepsilon; \mathbf{x}, \mathbf{f}(\mathbf{x})) \qquad [4.2.2]$$

(or equivalently, $\mathbf{f}(\tilde{\mathbf{x}}) = \tilde{\mathbf{f}}(\tilde{\mathbf{x}})$). In this equation, the right hand side represents the transformed solution in $\tilde{\mathbf{x}}\tilde{\mathbf{f}}$ space; the left hand side is simply our original solution with the independent variables renamed. The equals sign is therefore telling us that $\mathbf{f}$ and $\tilde{\mathbf{f}}$ are of the same form. We shall call solutions which satisfy [4.2.2] for a given group of transformations, *similarity solutions.*

The group structure that we have defined provides us with a way of finding all possible candidates for a similarity solution. In summary, we will show that given a set of transformations, g, f can be written as a function of a single x_i and $n-1$ other quantities which are constant under g. We shall first of all prove this statement for the general case. In the next section, however, these ideas will be illustrated by considering a particular set of transformations that act upon F, v_i, and ρ_i in equations [4.1.1]–[4.1.3]. Now, considering only one component of the vector equation, [4.2.2], we have,

$$f(g_{\mathbf{x}}(\varepsilon;\mathbf{x})) = g_f(\varepsilon;\mathbf{x}, f(\mathbf{x})). \tag{4.2.3}$$

By differentiating this equation with respect to ε and evaluating the result at $\varepsilon = 1$ we obtain,

$$\frac{\partial f}{\partial x_i}(\mathbf{x})\left.\frac{dg_{x_i}}{d\varepsilon}(\mathbf{x})\right|_{\varepsilon=1} = \left.\frac{dg_f}{d\varepsilon}(\mathbf{x}, f(\mathbf{x}))\right|_{\varepsilon=1}. \tag{4.2.4}$$

This is a linear partial differential equation and we can solve this using the method of characteristics. The characteristic system is

$$\frac{dx_i}{\partial g_{x_i}/\partial\varepsilon|_{\varepsilon=1}} = \frac{df}{\partial g_f/\partial\varepsilon\big|_{\varepsilon=1}}.$$

Integrating the first pair of equations gives

$$s_2 = h_2(x_1, x_2) = \text{const},$$

whereas the rest, bar one, give

$$s_i = h_i(x_1, x_i) = \text{const} \quad (i = 3, \ldots, n).$$

This leaves

$$\frac{dx_1}{\partial g_{x_1}/\partial\varepsilon|_{\varepsilon=1}} = \frac{df}{\partial g_f/\partial\varepsilon\big|_{\varepsilon=1}}.$$

Letting $x_i = X_i(x_1, s_i)$, where $i = 2, \ldots, n$ enables us to integrate (in theory) this last equation to give

$$h(x_1, s_i, f) = \text{const}.$$

Consequently, the general solution to [4.2.4] is given by $F(s_i, h(x_1, s_i, f)) = 0$ from which we can obtain $f(\mathbf{x})$ and thus candidates for similarity solutions.

§4.3. Similarity solutions for Newtonian equations.

We shall illustrate the construction of g, the group of transformations, and f, the set of all possible similarity solutions, of the previous section by considering [4.1.1]–[4.1.3] in conjunction with a new symmetry. We suppose that in the case of caustic formation with planar symmetry, equations [4.1.1]–[4.1.3] are also invariant under a simple scaling transformation. Consider a region surrounding the cusp, an example might be the set $U = [0, T] \times [-X, X]$. We might wish to magnify this region by multiplying the t coordinate by $\alpha > 1$ and the x coordinate by α^β so that we now have $U_\alpha = [0, \alpha T] \times [-\alpha^\beta X, \alpha^\beta X]$. Suppose now that f is some function defined on U_α as might be F, v_i or ρ_i. Then by invariance under the above transformation we really mean that there exists an α_f such that for $(t, x) \in U \subset U_\alpha$, we have an $(\alpha t, \alpha^\beta x) \in U_\alpha$ and $f(t, x) = \alpha_f f(\alpha t, \alpha^\beta x)$.

We now write our scaling transformations in the following concise form,

$$g(\alpha; t, x, F, v_i, \rho_i) = (\alpha t, \alpha^\beta x, \alpha^{k_F} F, \alpha^{k_{v_i}} v_i, \alpha^{k_{\rho_i}} \rho_i), \qquad [4.3.1]$$

where $\alpha \in \mathbb{R}$ and β and k_f are fixed constants. The fact that $g(1; t, x, F, v_i, \rho_i)$ corresponds to the identity transformation and that

$$\begin{aligned} g(\alpha_1; g(\alpha_2; t, x, F, v_i, \rho_i)) &= g(\alpha_1; \alpha_2 t, \alpha_2^\beta x, \alpha_2^{k_F} F, \alpha_2^{k_{v_i}} v_i, \alpha_2^{k_{\rho_i}} \rho_i)) \\ &= (\alpha_1 \alpha_2 t, \alpha_1^\beta \alpha_2^\beta x, \alpha_1^{k_F} \alpha_2^{k_F} F, \alpha_1^{k_{v_i}} \alpha_2^{k_{v_i}} v_i, \alpha_1^{k_{\rho_i}} \alpha_2^{k_{\rho_i}} \rho_i) \\ &= g(\alpha_1 \alpha_2; t, x, F, v_i, \rho_i) \end{aligned}$$

implies that g, for all $\alpha \in \mathbb{R}$, has a group structure. Similarity solutions are determined by the solution of equation [4.2.4], which for our case has the characteristic system,

$$\frac{dt}{t} = \frac{dx}{\beta x} = \frac{df}{k_f f}. \qquad [4.3.2]$$

The first pair can be integrated to give

$$\xi = h_x(t,x) = \frac{x}{t^\beta} = \text{const},$$

whereas the second pair gives

$$h(t,f) = \frac{f}{t^{k_f}} = \text{const}.$$

The general solution to [4.3.2] can then be written as

$$F\left(\xi, \frac{f}{t^{k_f}}\right) = 0.$$

Consequently, one possible candidate for a similarity solution is then,

$$f(t,x) = t^{k_f}\bar{f}(\xi). \qquad [4.3.3]$$

Equation [4.3.3] illustrates nicely the invariance of f for under such transformations discussed above, ξ is clearly constant and so the *shape* of f is preserved under g. Looking for solutions of this kind simplifies the mathematics because the number of independent variables in [4.1.1]–[4.1.3] is reduced from two to one. From a physical point of view this particular type of similarity solution suggests a kind of magnification invariance; if points on the caustic are related by $t = x^r$ where $r \in \mathbb{R}$, then for the correct choice of r, the caustic structure is preserved under $x \longrightarrow \alpha^\beta x$ and $t \longrightarrow \alpha t$.

Substituting [4.3.3] into [4.1.1]–[4.1.3] the 3 equations transform to

$$t^{k_F-\beta}\frac{d\bar{F}}{d\xi} = -\sum_i t^{k_{\rho_i}}\bar{\rho}_i,$$

$$(t^{k_{v_i}}\bar{v}_i)\left(t^{k_{v_i}-\beta}\frac{d\bar{v}_i}{d\xi}\right) + k_{v_i}t^{k_{v_i}-1}\,\bar{v}_i - \beta t^{k_{v_i}-1}\xi\frac{d\bar{v}_i}{d\xi} = t^{k_F}\bar{F},$$

and

$$(t^{k_{v_i}}\bar{v}_i)\left(t^{k_{\rho_i}-\beta}\frac{d\bar{\rho}_i}{d\xi}\right) + (t^{k_{\rho_i}}\bar{\rho}_i)\left(t^{k_{v_i}-\beta}\frac{d\bar{v}_i}{d\xi}\right) = -k_{\rho_i}t^{k_{\rho_i}-1}\,\bar{\rho}_i + \beta t^{k_{\rho_i}-1}\xi\frac{d\bar{\rho}_i}{d\xi}.$$

These equations simplify further; as a result we obtain,

$$\frac{d\bar{F}}{d\xi} = -\sum_i t^{k_{\rho_i}+\beta-k_F}\bar{\rho}_i,$$

$$\bar{v}_i\frac{d\bar{v}_i}{d\xi} + k_{v_i}t^{\beta-k_{v_i}-1}\,\bar{v}_i - \beta t^{\beta-k_{v_i}-1}\xi\frac{d\bar{v}_i}{d\xi} = t^{\beta+k_F-2k_{v_i}}\bar{F},$$

and

$$\bar{v}_i\frac{d\bar{\rho}_i}{d\xi} + \bar{\rho}_i\frac{d\bar{v}_i}{d\xi} = -k_{\rho_i}t^{\beta-k_{v_i}-1}\,\bar{\rho}_i + \beta t^{\beta-k_{v_i}-1}\xi\frac{d\bar{\rho}_i}{d\xi}.$$

For the above set of equations to be invariant under our scaling transformations, the coefficients of each term must be time independent. We are thus required to set all powers of t to be zero, i.e.

$$k_{\rho_i} + \beta - k_F = 0, \quad \beta - k_{v_i} - 1 = 0, \quad \beta + k_F - 2k_{v_i} = 0.$$

This implies for the similarity degrees,

$$k_{v_i} = \beta - 1, \quad k_F = \beta - 2, \quad k_{\rho_i} = -2,$$

and equations [4.1.1]–[4.1.3] become

$$\bar{F}' = -\sum_i \bar{\rho}_i, \qquad [4.3.4]$$

$$\bar{v}_i\bar{v}_i' + (\beta-1)\bar{v}_i - \beta\xi\bar{v}_i' = \bar{F}, \qquad [4.3.5]$$

$$\bar{v}_i\bar{\rho}_i' + \bar{\rho}_i\bar{v}_i' - \beta\xi\bar{\rho}_i' = 2\bar{\rho}_i, \qquad [4.3.6]$$

where $'$ denotes differentiation with respect to ξ.

We now specialise by stipulating boundary conditions. Rather than setting up conditions on some $t = 0$ time slice for example, we define the equation of the caustic in M. We require that the three dust region is bounded by

$$\xi = \pm 1 \qquad [4.3.7]$$

for $\beta > 1$. This forms part of our joining conditions that was mentioned earlier for it tells us where in the tx plane this occurs. To complete these instructions we define how particles are transferred from one dust to another. That is to say, we require that at $\xi = \pm 1$, the trajectories with velocity v_2 and v_3, or respectively, with velocity v_1 and v_2, to be tangent to the caustic. This defines a limit for v_i as it approaches the caustic. At $\xi = 1$ we have $x = t^\beta$ and hence $dx/dt = \beta t^{\beta-1}$. Thus

$$\bar{v}_i = t^{-k_{v_i}} v_i = t^{1-\beta}(\beta t^{\beta-1}) = \beta \quad i = 2, 3.$$

Similarly at $\xi = -1$,

$$\bar{v}_i = t^{-k_{v_i}} v_i = t^{1-\beta}(-\beta t^{\beta-1}) = -\beta \quad i = 1, 2.$$

This also defines the domains for the functions $\bar{v}_i$ and $\bar{\rho}_i$. For $\bar{v}_1$, $\bar{\rho}_1$; $\bar{v}_2$, $\bar{\rho}_2$ and $\bar{v}_3$, $\bar{\rho}_3$, we have the sets $(-1, \infty)$, $(-1, 1)$ and $(-\infty, 1)$ respectively. Boundary conditions for the $\bar{\rho}_i$ are needed for a complete specification but we shall consider this in the next section.

The important feature of equations [4.3.4]–[4.3.6] is that having specified boundary conditions for each v_i, we find that the latter two become singular on the caustic. In other words the coefficient of the highest derivative vanishes. To highlight this behaviour in $\bar{v}_i$, we transform the equations using $w_i = \bar{v}_i - \beta\xi$. In this case the boundary conditions become

$$w_1(-1) = w_2(-1) = w_2(1) = w_3(1) = 0, \tag{4.3.8}$$

and if we include the symmetry about the plane $x = 0$, we have

$$\bar{F}(0) = 0$$

and

$$w_1(\xi) = -w_3(-\xi), \quad w_2(\xi) = -w_2(-\xi), \quad w_2(0) = 0. \tag{4.3.9}$$

Transforming [4.3.5] and [4.3.6] using the above substitution gives

$$w_i w_i' + (2\beta - 1)w_i + \beta(\beta - 1)\xi = \bar{F} \tag{4.3.10}$$

and

$$w_i \bar{\rho}_i' + \bar{\rho}_i w_i' - (2 - \beta)\bar{\rho}_i = 0. \tag{4.3.11}$$

With the original equations rewritten in this final form, we can clearly see that [4.3.5] and [4.3.6] become singular when the boundary conditions are imposed.

§4.4. Zero gravity case.

Having obtained the differential equations involving the variables w_i, $\bar{\rho}_i$ and $\bar{F}$, we specialise further by considering the case when $\bar{F} = 0$. Integrating [4.3.11] gives for the density

$$\bar{\rho}_i = \frac{1}{|w_i|} \exp\left\{ \int_{a_i}^{\xi} \frac{(2-\beta)}{w_i} d\bar{\xi} \right\}, \qquad [4.4.1]$$

where $a_i \in \mathbb{R}$ represents the constants of integration. For the transformed velocity, dividing equation [4.3.10] by w_i gives a homogeneous differential equation which when solved yields,

$$w_i + \beta\xi = C_i(w_i + (\beta - 1)\xi)^{(\beta-1)/\beta} \qquad [4.4.2]$$

where $\xi \neq 0$ and $C_i \in \mathbb{C}$ are constants. We should say at this point that [4.4.2] is in fact four solutions. We obviously have the cases where $i = 1$ and $i = 3$ but also we have the cases where $i = 2$, $\xi < 0$ and $i = 2$, $\xi > 0$. The reason why for $i = 2$ we have two solutions is because [4.3.10] is singular at the origin and thus is essentially two differential equations defined on the domains $(-1, 0)$ and $(0, 1)$ respectively. This gives the two solutions that we mention above. To determine w_2 we simply 'glue' the solutions together at the origin.

We now impose the boundary conditions given above in order to verify that solutions [4.4.2] are in fact consistent with this analytic model of a caustic. To apply the four conditions of [4.3.8], we take limits of both sides of equation [4.4.2], noting that x^r is continuous for all $r > 0$ in $\mathbb{R}$. We obtain,

$$\lim_{\xi \to -1} w_i - \beta = C_i \left(\lim_{\xi \to -1} w_i - (\beta - 1) \right)^{(\beta-1)/\beta} \qquad i = 1 \text{ and } i = 2 \text{ with } \xi < 0$$

$$\Longrightarrow C_i = -\frac{\beta}{(1-\beta)^{(\beta-1)/\beta}} \qquad i = 1 \text{ and } i = 2 \text{ with } \xi < 0.$$

Similarly,

$$\lim_{\xi \to 1} w_i + \beta = C_i \left(\lim_{\xi \to 1} w_i + (\beta - 1) \right)^{(\beta-1)/\beta} \qquad i = 2 \text{ with } \xi > 0 \text{ and } i = 3$$

$$\Longrightarrow C_i = \frac{\beta}{(\beta-1)^{(\beta-1)/\beta}} \qquad i = 2 \text{ with } \xi > 0 \text{ and } i = 3.$$

The fact that w_2 has two different constants of integration is an artifact of what we discussed above: the different C_i arise from the two solutions for w_2.

Having set the constants C_i, we must now check that the final three constraints of [4.3.9] (recall that $\bar{F} = 0$) agree with this choice. To verify that $\lim_{\xi \to 0} w_2(\xi) = 0$ let us define two new variables, $p_i = w_i + (\beta - 1)\xi$ and $q_i = w_i + \beta\xi$. We then obtain from [4.4.2],

$$q_i = C_i p^{(\beta-1)/\beta}, \qquad [4.4.3]$$

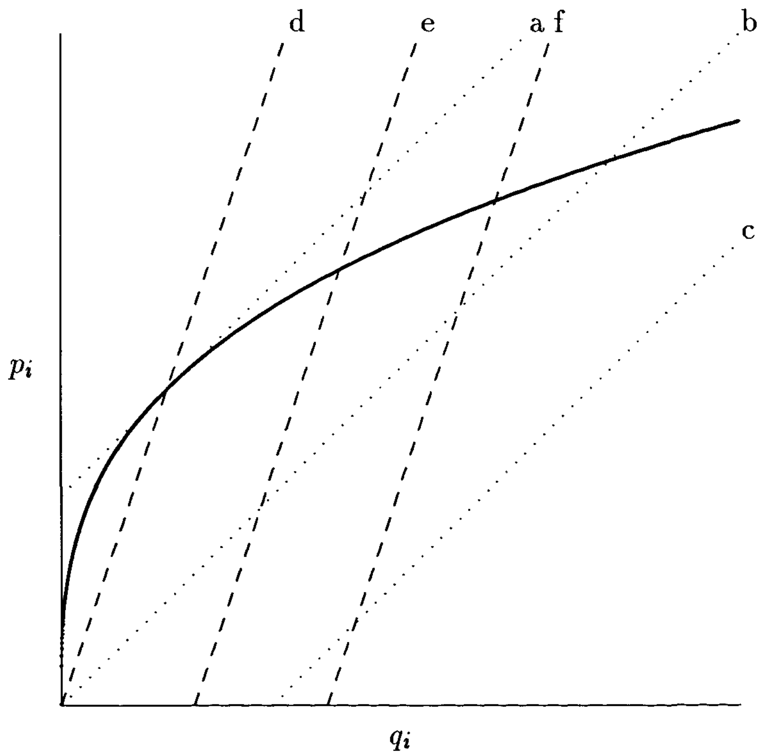
whereas from the definitions of p_i and q_i,

$$\xi = q_i - p_i, \qquad [4.4.4]$$

and

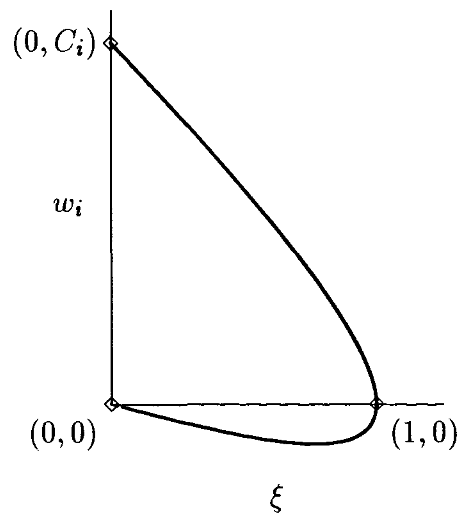
$$w_i = \beta p_i - (\beta - 1)q_i. \qquad [4.4.5]$$

Fig. 3. Plot of $p = w_i + \beta\gamma\xi$ against $q = w_i + (\beta + 1)\gamma\xi$ for $q > 0$, $\beta = 3/2$ and $\gamma = 2^{2/3}/3$. Here the lines a, b, ..., f represent $\xi = 1$, $\xi = 0$, $\xi = -1$, $w_i = 0$, $w_i = 1/2$ and $w_i = 1$ respectively.



On a pq graph, lines of constant ξ and w_i can be plotted using equations [4.4.4], [4.4.5] and assumed values for β. This is shown in fig. 3. Equation [4.4.3] is also plotted. Since this is a representation of the solution given by [4.4.2] it is possible, using this information, to sketch $w_i(\xi_i)$ and this is given in fig. 4. Note that these graphs must be interpreted with care for they do not consider how the i-value might change as different points, (p, q) or ξ, are chosen. In the latter case it is easy for the changes from w_1 to w_2 and from w_2 to w_3 occur when ξ equals -1 and 1 respectively (recall the boundary conditions of [4.3.8]). For the former we need to find the set of points in $\mathbb{R}^2$ such that $q - p = -1$ and $q - p = 1$. These lines then identify where points corresponding to $\xi = \pm 1$ lie on our pq graph and thus where the i-value changes.

Fig. 4. Plot of w_i against ξ for $\xi > 0$ and $\bar{F} = 0$.



Now one can see from fig. 3 that for any $\beta > 1$ there will be two values for $\lim_{\xi\to 0} p_2$ and $\lim_{\xi\to 0} q_2$ and thus possibly four values for $\lim_{\xi\to 0} w_2(\xi)$ by [4.4.5]. Hence, to prove that $\lim_{\xi\to 0} w_2(\xi) = 0$, we specify the direction along the curve $q_i = C_i p_i^{(\beta-1)/\beta}$ with which we approach $\xi = 0$. This is done by removing the second limit point and we do this in the following manner. We have from our pq graph that the line corresponding to $\xi = 0$ has equation $q = p$. Define $\bar{p}$ to be the point where $q = kp$ ($k \in \mathbb{R}$) intersects [4.4.3]. In other words $\bar{p}$ is the solution to $kp = C_i^{1/\beta} p^{(\beta-1)/\beta}$. Then if we start from the origin and move along the curve given by [4.4.3], we reach the point $(\bar{p}, k\bar{p})$ before the curve crosses $\xi = 0$ again provided $q = kp$ has steeper gradient than $q = p$. In other words, provided the constant k

is greater than one. This means that by restricting ourselves to $0 \leq p_2 \leq \bar{p}$ we have defined a domain where only a single limit point exists. Thus we have for all $0 \leq p_2 \leq \bar{p}$ that $kp_2 \leq q_2$ and

$$\xi = q_2 - p_2 \geq (k-1)p_2 \geq 0$$

$$\Longrightarrow \lim_{\xi \to 0} p_2 = 0$$

$$\Longrightarrow \lim_{\xi \to 0} q_2 = 0$$

by equation [4.4.4]

$$\Longrightarrow \lim_{\xi \to 0} w_2(\xi) = 0$$

by equation [4.4.5].

The final check to be made is that $w_i(\xi)$ satisfies the boundary condition, $w_i(\xi) = -w_j(-\xi)$, where the i and j take the appropriate values determined by [4.3.9]. Now suppose $\xi \in (-\infty, 1)$ and $-\xi \in (1, \infty)$. Then the solutions to [4.3.10] in these regions with $\bar{F} = 0$ are $w_3(\xi)$ and $w_1(-\xi)$. By [4.4.4], $\xi = q_3(\xi) - p_3(\xi)$ and $-\xi = q_1(-\xi) - p_1(-\xi)$. Thus,

$$q_3(\xi) + q_1(-\xi) = p_3(\xi) + p_1(-\xi)$$

$$\Longrightarrow C_3(p_3(\xi))^{(\beta-1)/\beta} + C_1(p_1(-\xi))^{(\beta-1)/\beta} = p_3(\xi) + p_1(-\xi)$$

$$\Longrightarrow \frac{\beta}{(\beta-1)^{(\beta-1)/\beta}}(p_3(\xi))^{(\beta-1)/\beta} + \frac{\beta}{(\beta-1)^{(\beta-1)/\beta}}(-p_1(-\xi))^{(\beta-1)/\beta} = p_3(\xi) + p_1(-\xi).$$

Since this equation must hold for all β and hence for all C_3 we can equate coefficients. This gives

$$p_3(\xi) = -p_1(-\xi)$$

$$\Longrightarrow w_3(\xi) + (\beta-1)\xi = -(w_1(-\xi) - (\beta-1)\xi)$$

$$\Longrightarrow w_3(\xi) = -w_1(-\xi)$$

as required. To show that w_2 is an odd function we repeat the above analysis for $-1 < \xi < 1$.

Now that we have an expression for the solution to w_i that satisfies all the necessary constraints (equations [4.3.8] and [4.3.9]), let us check that this reduces to the cubic equation of §3.2. Referring back to equations [3.4.2] and [4.3.7], the appropriate choice for β in order to match the exponents of the two equations would be $\beta = 3/2$. This value of β is in fact justified in a more rigorous manner in section 4.7 of this chapter. However, with these values [4.4.2] becomes for $\xi = x/t^\beta \in (-\infty, 1)$,

$$\bar{v}_3 = \frac{3/2}{(1/2)^{1/3}}\left(\bar{v}_3 - \frac{x}{t^{3/2}}\right)^{1/3}$$

$$\Longrightarrow \bar{v}_3^3 = 2\frac{27}{8}\left(\bar{v}_3 - \frac{x}{t^{3/2}}\right)$$

$$\Longrightarrow \frac{v_3^3}{t^{3/2}} = \frac{27}{4}\left(\frac{v_3}{t^{1/2}} - \frac{x}{t^{3/2}}\right)$$

$$\Longrightarrow v_3^3 = \frac{27}{4}\left(v_3 t - x\right).$$

So if we transform x and t such that $x \longrightarrow 27x/4$ and $t \longrightarrow 27t/4$ we obtain $x = q(t, v_3)$ as required. This same procedure can be repeated for the regions $(-1, 1)$ and $(-1, \infty)$ giving $x = q(t, v_2)$ and $x = q(t, v_1)$ respectively.

Having discussed the analytic solution to [4.3.10] for $\bar{F} = 0$, let us reconsider the solution for the density given in equation [4.4.1]. Now as of yet, no boundary conditions have been specified for the $\bar{\rho}_i$. In fact we shall not bother to define any conditions except to say that the density should be finite everywhere apart form points on the caustic. We will show that with the solution of the form given by [4.4.1] this simple requirement is not satisfied.

Let us investigate the limit of $\bar{\rho}_i$ as ξ tends towards zero. To do this, we need to calculate two quantities, namely $\lim_{\xi \to 0} \xi / w_2$ and $\lim_{\xi \to 0} w_2'$. From [4.4.2] we have for $\xi \neq 0$,

$$w_2\left(1 + \frac{\beta\xi}{w_2}\right) = C_2 w_2^{(\beta-1)/\beta}\left(1 + \frac{(\beta-1)\xi}{w_2}\right)^{(\beta-1)/\beta}$$

$$\Longrightarrow w_2^{1/\beta}\left(1 + \frac{\beta\xi}{w_2}\right) = C_2\left(1 + \frac{(\beta-1)\xi}{w_2}\right)^{(\beta-1)/\beta}$$

$$\Longrightarrow 0 = \lim_{\xi \to 0} \left(1 + \frac{(\beta - 1)\xi}{w_2}\right)^{(\beta-1)/\beta}$$

$$\Longrightarrow \lim_{\xi \to 0} \frac{\xi}{w_2} = \frac{1}{(1-\beta)} \qquad [4.4.6]$$

and hence from [4.3.10],

$$\lim_{\xi \to 0} w_2' = \beta(1-\beta) \lim_{\xi \to 0} \frac{\xi}{w_2} - (2\beta - 1)$$

$$\Longrightarrow \lim_{\xi \to 0} w_2' = (1-\beta). \qquad [4.4.7]$$

Now provided the constant a_i in [4.4.1] is small, we can use [4.4.7] to make the approximation, $w_2 \approx (1-\beta)\xi$, so that

$$\begin{aligned} \int_{a_i}^{\xi} \frac{(2-\beta)}{w_2} \, d\bar{\xi} &= \int_{a_i}^{\xi} \frac{(2-\beta)}{(1-\beta)\xi} \, d\bar{\xi} \\ &= \frac{2-\beta}{1-\beta} \ln \frac{\xi}{a_i}. \end{aligned}$$

Thus, using l'Hopital's rule,

$$\begin{aligned} \lim_{\xi \to 0} \bar{\rho}_2 &= \lim_{\xi \to 0} \left\{ \frac{1}{|w_2|} \left(\frac{\xi}{a_i}\right)^{(2-\beta)/(1-\beta)} \right\} \\ &= \lim_{\xi \to 0} \left\{ \pm \frac{1}{w_2'} \cdot \frac{2-\beta}{1-\beta} \left(\frac{\xi}{a_i}\right)^{(2-\beta)/(1-\beta)-1} \frac{1}{a_i} \right\} \\ &= \lim_{\xi \to 0} \left\{ \pm \frac{(2-\beta)}{(1-\beta)^2} \left(\frac{\xi}{a_i}\right)^{1/(1-\beta)} \frac{1}{a_i} \right\} \end{aligned}$$

and since $\beta > 1$ we conclude that the limit cannot exist, i.e. $\bar{\rho}_2 \longrightarrow \infty$ as $\xi \longrightarrow 0$. This is unfortunate as it predicts unbounded behaviour for the density on the axis of symmetry, contradicting our requirement that ρ_i must be finite everywhere other than at points on the caustic.

We finish this section by raising two important points. The first is an observation on the above solution for ρ_i. If we refer back to where values for the similarity degrees were derived, we can see that k_{ρ_i} is fixed only by the force equation, [4.3.4]. In other words, by fixing $\bar{F} = 0$, [4.3.4] can effectively be discarded so that we are able to freely choose k_{ρ_i}. If this is the case, then by following the above method and replacing the constant $(2-\beta)$ by $(-k_{\rho_i}-\beta)$, we find that the limit of $\bar{\rho}_i$ as ξ tends to zero exists if $k_{\rho_i}=-1$. The second point is that had we removed all force terms at the onset, we surely would have obtained a realistic solution. In other words, it seems that the presence of F is supplying information that is corrupted when we "turn off the force". Clearly these two points are intrinsically linked and we shall expand on these in the next chapter.

§4.5. Similarity solutions in tv space.

Having given a detailed discussion on similarity solutions to equations [4.1.1]–[4.1.3] in tx space, this section digresses from the natural flow to analyse the same equations but in tv space. This is equivalent to working in the tangent bundle. The benefit of this approach is that, as eluded to in the introduction of this chapter, the mass flux is well defined even at the caustic where the density becomes infinite.

Now to formulate the problem we need to transform equations [4.1.1]–[4.1.3] and the boundary conditions we wish to consider. Appendix 1 provides us with the equations which help us to do this. Care must be taken, however, when defining the mass flux in TM since although it overcomes the problems that we have if we formulate the problem in tx space, it does have an extra subtlety. This extra complication arises from the fact that the projection of the tangent bundle surface onto M is not orientation preserving. Before we discuss how the mass flux shall be defined, let us explain exactly why this is so.

The formation of S can be visualised by folding a piece of paper. We begin with a flat sheet which represents N (see chapter 2). An orientation can be considered as a chosen direction for any vector that traverses the paper. Thus we have two possible orientations for N. Suppose that we make a fold, without creasing, in the sheet of paper and then another so that the two folds are parallel. The paper is now in an 'S'-shape if looked at edge on. This action represents the map $f\colon N \longrightarrow TM$ of §2.2. Note that the orientation is preserved for we can still define a unique way of traversing $S=f(N)$. It does not matter that a vector, $\mathbf{n}$ say, normal to the original flat sheet of paper, would first enter, then exit and then re-enter S if we pushed it through the folded region. We now ask ourselves what happens if we collapse our surface onto itself so that the folds in our sheet become creases. In this case our surface, again looked at edge on, becomes like a concertina. This action is equivalent to the map $\pi: TM \longrightarrow M$ and the result represents $\pi(S)$. We immediately see that our orientation is no longer preserved for where we have three layers that coincide, there is not a well defined direction with which we can cross S. In other words, $\mathbf{n}$ can simultaneously enter and exit S if placed at a point on $\pi(S)$ that was formed by the superposition of three separate parts of S.

How does this affect our definition of the mass flux in TM? To answer this, let us suppose that P and Q are orientable manifolds and that there exists $g: Q \longrightarrow P$ which is orientation preserving. Furthermore, let there be some matter, a fluid for example, which moves on both manifolds. The simplest case is where the fluids are identifiable and that P and Q are simply different coordinate systems. In this example, if $U \subset Q$ is some region then the masses in U and $g(U)$ are the same. We can write this mathematically by supposing that

$$M(U) = \int_U g^* \rho = \int_{g(U)} \rho \qquad [4.5.1]$$

where ρ, by assumption, is a closed g_U-form ($g_U = \dim(g(U))$) on P representing the mass flux at some arbitrary point in $g(U)$. The assumption that ρ is closed illustrates the fact that it is conserved. We could show that equation [4.1.3] implies this statement, however, in order to avoid a complicated aside, this will be demonstrated in chapter 6 where the calculation becomes essential to the argument.

Such a definition is fine except when g is *not* orientation preserving. If this is the case then the integral over $g(U)$ is not defined because the integration of forms over the whole of this region requires a continuous orientation. Since this is the situation we expect when we project S onto M for Newtonian caustics, we are forced to modify our definition of mass flux. Instead, we define the mass in U as in [4.5.1] except that we insist that ρ be regarded as a *pseudo-g_U*-form. This is because pseudo-forms can be integrated over non orientable manifolds and they transform between coordinate systems in a very similar manner to forms, i.e. for $q \in U$ and ρ, a pseudo-g_U-form on P,

$$(g^*\rho)_{j_1 \dots j_{g_U}}(q) = \frac{\partial x^{i_1}}{\partial y^{j_1}} \cdots \frac{\partial x^{i_{g_U}}}{\partial y^{j_{g_U}}} \operatorname{sgn}\left[\det\left(\frac{\partial x}{\partial y}\right)\right] \rho_{i_1 \dots i_{g_U}}(g(q))$$

where the x^i and y^j are local coordinates to P and Q respectively.

We can easily calculate the components of the corresponding ρ for the case that we are considering. Thus, Q becomes TM, P becomes M, g becomes π and U is a region in TM such that $\pi(U)$ is a one dimensional surface in M. The reason why we choose $\pi(U)$ to be one dimensional is because ultimately, we wish to define the mass as the integral over a $t =$ const region. Now, since we have a region in M in which there are three superimposed dusts, we consider a pseudo-1-form for each, i.e. we now have $\rho_{(k)}$ for $k = 1, 2, 3$. We also introduce $\sigma_{(k)}$, which is defined on TM, to be the pullback of $\rho_{(k)}$. Suppose now that $\mu_{(k)}$ is a scalar field on M representing the density of the kth dust at any point. Then the mass in $\pi(U)$ say, of the kth dust is also given by

$$M_{(k)}(U) = \int_{\pi(U)} \mu_{(k)} \alpha_{(k)} \qquad [4.5.2]$$

where $\alpha_{(k)}$ is a psuedo-1-form and can be thought of as the volume form on a 1-dimensional surface in M (The equivalence of [4.5.1] and [4.5.2] will be shown in chapter 6.). If $v_{(k)}$ represents the kth dust's flow vector and X a vector such that $(v_{(k)}, X)$ constitutes a basis for M, then if we choose $\pi(U)$ to be orthogonal to $v_{(k)}$ then $\alpha_{(k)}$ can be defined as $(-1)^{k+1} i_{v_{(k)}} \alpha_M$ where α_M is the volume form on M. This quantity represents the restriction of α_M to a surface in M corresponding to those points which contain particles of dust k. The factor of $(-1)^{k+1}$ is important for it takes into account the fact that π is not orientation preserving. If we recall the Euclidean tangent bundle surface, S_q, and project this onto M, we obtain a 'squashed Z'-shaped hypersurface. The horizontal parts of this 'Z' would represent dusts 1 and 3, whereas the diagonal part would represent dust 2. Thus, because a normal to the line formulating our 'Z' flips in direction as it moves form the dust 1 (or dust 3) region into the dust 2 region, it follows that $i_{v_{(2)}} \alpha_M$ has the opposite orientation with respect to $i_{v_{(1)}} \alpha_M$ or $i_{v_{(3)}} \alpha_M$. The $(-1)^{k+1}$ factor accounts for this and therefore insures that the $\alpha_{(k)}$ all have the same orientation with respect to α_M. Now, since in our case the spacetime is flat,

$$\alpha_M = \frac{1}{2!} \varepsilon_{ij} dx^i \wedge dx^j$$

and

$$\begin{aligned} i_{v_{(k)}} \alpha_M(X) &= \frac{1}{2!} \varepsilon_{ij} dx^i \wedge dx^j (v_{(k)}, X) \\ &= \frac{1}{2!} \varepsilon_{ij} \left(dx^i(v_{(k)}) dx^j(X) - dx^i(X) dx^j(v_{(k)}) \right) \end{aligned}$$

$$= \frac{1}{2!}\varepsilon_{ij}\left(v^k_{(k)}\delta^i_k dx^j(X) - dx^i(X)v^k_{(k)}\delta^j_k\right)$$

$$\Longrightarrow i_{v_{(k)}}\alpha_M = \frac{1}{2!}\varepsilon_{ij}\left(v^i_{(k)}dx^j - v^j_{(k)}dx^i\right)$$

$$= \varepsilon_{ij}v^i_{(k)}dx^j.$$

Thus finally we have $\rho_{(k)} = (-1)^{k+1}\mu_{(k)}\varepsilon_{ij}v^i_{(k)}dx^j = (-1)^{k+1}\mu_{(k)}\left(dx - v_{(k)}dt\right)$ if our fluid follows time parameterised geodesics in M. Note that $v_{(k)}$ corresponds to v_i of the previous section when $k = i$. Note also that if we restrict this to a $t =$ const time slice then $\rho_{(k)} = (-1)^{k+1}\mu_{(k)}dx$.

This quantity pulls back onto the tangent bundle. We have, for any basis vector for TM,

$$\sigma_{(k)}\left(\frac{\partial}{\partial y^j}\right) = (\pi^*\rho_{(k)})\left(\frac{\partial}{\partial y^j}\right)$$

$$= \text{sgn}\left[\det\left(\frac{\partial x}{\partial y}\right)\right]\rho_{(k)}\left(\pi_*\frac{\partial}{\partial y_j}\right)$$

$$= \text{sgn}\left[\det\left(\frac{\partial x}{\partial y}\right)\right]\frac{\partial x^i}{\partial y^j}\rho_{(k)}\left(\frac{\partial}{\partial x^i}\right).$$

Thus for the Newtonian case, which we are considering, we have $(x^1, x^2) = (t, x)$ and $(y^1, y^2, y^3, y^4) = (t, x, u, v)$ so that upon restricting $\sigma_{(k)}$ to a surface in TM defined by $(t, x, u, v) = (t, x(t, v), 1, v)$ we have

$$\sigma_{(k)}\left(\frac{\partial}{\partial v_{(k)}}\right) = \text{sgn}\left[\det\left(\frac{\partial x}{\partial y}\right)\right]\frac{\partial x}{\partial v_{(k)}}\rho_{(k)}\left(\frac{\partial}{\partial x}\right)$$

and

$$\sigma_{(k)}\left(\frac{\partial}{\partial t}\right) = \text{sgn}\left[\det\left(\frac{\partial x}{\partial y}\right)\right]\left\{\frac{\partial x}{\partial t}\rho_{(k)}\left(\frac{\partial}{\partial x}\right) + \rho_{(k)}\left(\frac{\partial}{\partial t}\right)\right\}.$$

If we restrict this further to a $t =$ const time slice in TM then

$$\sigma_{(k)} = \sigma_{(k)}\left(\frac{\partial}{\partial v_{(k)}}\right)dv = \text{sgn}\left[\det\left(\frac{\partial x}{\partial y}\right)\right]\frac{\partial x}{\partial v_{(k)}}(-1)^{k+1}\mu_{(k)}dv.$$

We define σ_k and ρ_k as the components of $\sigma_{(k)}$ and $\rho_{(k)}$ respectively. Here the quantity ρ_k represents the density function of §4.1–$4.4. It is the occurrence of the sgn $[\det(\partial x/\partial y)]$ term in the definition of σ_k that constitutes the complexity spoken of at the beginning of this section. We can remove this if we know the sign of $\partial x/\partial v_k$ for different values of k since

$$\det\left(\frac{\partial x}{\partial y}\right) = \begin{vmatrix} \partial x/\partial t & 1 \\ \partial x/\partial v_k & 0 \end{vmatrix} = -\frac{\partial x}{\partial v_k}.$$

If we refer to fig. 2, §2.1, where the surface $S_q \subset TM$ is sketched, then we see that this quantity is positive for $k = 2$ and negative for $k = 1, 3$. In other words we write $\sigma_{(k)} = -\mu_{(k)}\partial x/\partial v_k dv$ or $\sigma_k = (-1)^k \rho_k \partial x/\partial v_k$. Finally, now that both $\rho_{(k)}$ and $\sigma_{(k)}$ are defined, we can see how the mass flux in TM remains bounded even at the caustic, clarifying the statement to this effect towards the end of \$4.1. The reason is simply that $\partial x/\partial v_{(k)}$ tends to zero (ref. §3.4) as we approach the caustic allowing the product, $\mu_{(k)}\partial x/\partial v_{(k)}$, to remain finite.

We are now in a position to convert [4.1.1]–[4.1.3] into equivalent equations with t and v as the independent variables. Now equation [4.1.1] is correctly written as,

$$\left(\frac{\partial F}{\partial x}\right)_t = -\sum_j \rho_j.$$

So that if one assumes $F = F(t, v_i(t,x))$ and defines σ_i as above, then using equation [A1.1.3] this becomes

$$\left(\frac{\partial F}{\partial v_i}\right)_t \left(\frac{\partial v_i}{\partial x}\right)_t = -\sum_j \rho_j$$

$$\begin{aligned} \Longrightarrow \left(\frac{\partial F}{\partial v_i}\right)_t &= -\sum_j \rho_j \left(\frac{\partial x}{\partial v_i}\right)_t \\ &= (-1)^{i+1}\sigma_i - \sum_{j\neq i} \rho_j \left(\frac{\partial x}{\partial v_i}\right)_t \\ &= (-1)^{i+1}\sigma_i - \sum_{j\neq i} \rho_j \frac{(\partial x/\partial v_j)_t\,(\partial x/\partial v_i)_t}{(\partial x/\partial v_j)_t} \\ \Longrightarrow \left(\frac{\partial F}{\partial v_i}\right)_t &= (-1)^{i+1}\sigma_i + \sum_{j\neq i}(-1)^{j+1}\sigma_j \left(\frac{\partial v_j}{\partial v_i}\right)_t, \end{aligned}$$

upon using [A1.1.1] and again [A1.1.3]. For [4.1.2] we have,

$$v_i \left(\frac{\partial v_i}{\partial x}\right)_t + \left(\frac{\partial v_i}{\partial t}\right)_x = F$$

$$\Longrightarrow v_i + \left(\frac{\partial v_i}{\partial t}\right)_x \left(\frac{\partial x}{\partial v_i}\right)_t = F\left(\frac{\partial x}{\partial v_i}\right)_t$$

$$\Longrightarrow v_i - \left(\frac{\partial x}{\partial t}\right)_{v_i} = F\left(\frac{\partial x}{\partial v_i}\right)_t,$$

using $[A1.1.1]$ and $[A1.1.2]$. Finally for [4.1.3] we have,

$$\rho_i\left(\frac{\partial v_i}{\partial x}\right)_t + v_i\left(\frac{\partial \rho_i}{\partial x}\right)_t = -\left(\frac{\partial \rho_i}{\partial t}\right)_x$$

$$\Longrightarrow \rho_i\left(\frac{\partial}{\partial x}\left\{F\left(\frac{\partial x}{\partial v_i}\right)_t + \left(\frac{\partial x}{\partial t}\right)_{v_i}\right\}\right)_t + \left\{F\left(\frac{\partial x}{\partial v_i}\right)_t + \left(\frac{\partial x}{\partial t}\right)_{v_i}\right\}\left(\frac{\partial \rho_i}{\partial x}\right)_t$$
$$= -\left(\frac{\partial \rho_i}{\partial t}\right)_x$$

$$\Longrightarrow \rho_i\left(\frac{\partial F}{\partial x}\right)_t\left(\frac{\partial x}{\partial v_i}\right)_t + \rho_i F\left(\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial v_i}\right)_t\right)_t + \rho_i\left(\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial t}\right)_{v_i}\right)_t$$
$$+ F\left(\frac{\partial x}{\partial v_i}\right)_t\left(\frac{\partial \rho_i}{\partial x}\right)_t + \left(\frac{\partial x}{\partial t}\right)_{v_i}\left(\frac{\partial \rho_i}{\partial x}\right)_t = -\left(\frac{\partial \rho_i}{\partial t}\right)_x$$

$$\Longrightarrow (-1)^i\left(\frac{\partial \sigma_i F}{\partial x}\right)_t + \rho_i\left(\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial t}\right)_{v_i}\right)_t + \left(\frac{\partial x}{\partial t}\right)_{v_i}\left(\frac{\partial \rho_i}{\partial x}\right)_t = -\left(\frac{\partial \rho_i}{\partial t}\right)_x$$

$$\Longrightarrow (-1)^i\left(\frac{\partial \sigma_i F}{\partial x}\right)_t + \left(\frac{\partial}{\partial x}\left\{\rho_i\left(\frac{\partial x}{\partial t}\right)_{v_i}\right\}\right)_t = -\left(\frac{\partial \rho_i}{\partial t}\right)_x$$

$$\Longrightarrow (-1)^i\left(\frac{\partial \sigma_i F}{\partial x}\right)_t - (-1)^i\left(\frac{\partial}{\partial x}\left\{\sigma_i\left(\frac{\partial v_i}{\partial t}\right)_x\right\}\right)_t = -\left(\frac{\partial \rho_i}{\partial t}\right)_x$$

$$\Longrightarrow (-1)^i\left(\frac{\partial \sigma_i F}{\partial x}\right)_t - (-1)^i\left(\frac{\partial \sigma_i}{\partial x}\right)_t\left(\frac{\partial v_i}{\partial t}\right)_x - (-1)^i\sigma_i\left(\frac{\partial}{\partial x}\left(\frac{\partial v_i}{\partial t}\right)_x\right)_t$$
$$= -(-1)^i\left(\frac{\partial v_i}{\partial x}\right)_t\left(\frac{\partial \sigma_i}{\partial t}\right)_x + \left(\frac{\partial v_i}{\partial x}\right)_t\rho_i\left(\frac{\partial}{\partial t}\left(\frac{\partial x}{\partial v_i}\right)_t\right)_x$$

$$\Longrightarrow (-1)^i\left(\frac{\partial \sigma_i F}{\partial x}\right)_t - (-1)^i\left(\frac{\partial \sigma_i}{\partial x}\right)_t\left(\frac{\partial v_i}{\partial t}\right)_x - (-1)^i\sigma_i\left(\frac{\partial}{\partial x}\left(\frac{\partial v_i}{\partial t}\right)_x\right)_t$$
$$= -(-1)^i\left(\frac{\partial v_i}{\partial x}\right)_t\left(\frac{\partial \sigma_i}{\partial t}\right)_x - \left(\frac{\partial x}{\partial v_i}\right)_t\rho_i\left(\frac{\partial}{\partial t}\left(\frac{\partial v_i}{\partial x}\right)_t\right)_x$$

$$\Longrightarrow \left(\frac{\partial \sigma_i F}{\partial x}\right)_t - \left(\frac{\partial \sigma_i}{\partial x}\right)_t \left(\frac{\partial v_i}{\partial t}\right)_x - \sigma_i \left(\frac{\partial}{\partial x}\left(\frac{\partial v_i}{\partial t}\right)_x\right)_t$$

$$= - \left(\frac{\partial v_i}{\partial x}\right)_t \left(\frac{\partial \sigma_i}{\partial t}\right)_x - \sigma_i \left(\frac{\partial}{\partial t}\left(\frac{\partial v_i}{\partial x}\right)_t\right)_x$$

$$\Longrightarrow \left(\frac{\partial \sigma_i F}{\partial v_i}\right)_t - \left(\frac{\partial \sigma_i}{\partial t}\right)_x + \left(\frac{\partial \sigma_i}{\partial t}\right)_{v_i} = - \left(\frac{\partial \sigma_i}{\partial t}\right)_x$$

$$\Longrightarrow \left(\frac{\partial \sigma_i F}{\partial v_i}\right)_t + \left(\frac{\partial \sigma_i}{\partial t}\right)_{v_i} = 0,$$

where use has been made of equations [$A1.1.1$]–[$A1.1.4$]. To summarise, equations [4.1.1]–[4.1.3] transform to

$$\frac{\partial F}{\partial v_i} = (-1)^{i+1}\sigma_i + \sum_{j \neq i}(-1)^{j+1}\sigma_j \frac{\partial v_j}{\partial v_i}, \qquad [4.5.3]$$

$$v_i - \frac{\partial x}{\partial t} = F\frac{\partial x}{\partial v_i}, \qquad [4.5.4]$$

$$\frac{\partial \sigma_i F}{\partial v_i} + \frac{\partial \sigma_i}{\partial t} = 0, \qquad [4.5.5]$$

which represent our equations of motion in tv space.

We now analyse these equations using the same techniques discussed in the previous two sections. That is to say we shall look for similarity solutions which are invariant under simple scaling transformations in the t and v coordinates. We suppose that if $f(t, v_i)$ is any dependent variable then

$$\alpha^{k_f} f(t, v_i) = f(\alpha t, \alpha^{\beta} v_i), \qquad [4.5.6]$$

or equivalently,

$$f(t, v_i) = t^{k_f} \bar{f}(\xi_i). \qquad [4.5.7]$$

We note that the k_f and β represent different constants to those seen in the tx space analysis. If we insert [4.5.7] into equations [4.5.3]–[4.5.5] then

$$t^{k_F - \beta}\frac{\partial \bar{F}}{\partial \xi_i} = (-1)^{i+1} t^{k_{\sigma_i}}\bar{\sigma}_i + \sum_{j \neq i}(-1)^{j+1}\left(t^{k_{\sigma_j}}\bar{\sigma}_j\right)\left(t^{k_{v_j} - \beta}\frac{d\bar{v}_j}{d\xi_i}\right),$$

$$\xi_i t^\beta - k_x t^{k_x-1}\bar{x} + \beta t^{k_x-1}\xi_i \frac{d\bar{x}}{d\xi_i} = \left(t^{k_F}\bar{F}\right)\left(t^{k_x-\beta}\frac{d\bar{x}}{d\xi_i}\right),$$

$$\left(t^{k_{\sigma_i}-\beta}\frac{d\bar{\sigma}_i}{d\xi_i}\right)\left(t^{k_F}\bar{F}\right) + \left(t^{k_{\sigma_i}}\bar{\sigma}_i\right)\left(t^{k_F-\beta}\frac{d\bar{F}}{d\xi_i}\right) + k_{\sigma_i}t^{k_{\sigma_i}-1}\bar{\sigma}_i \quad - \beta t^{k_{\sigma_i}-1}\xi_i\frac{d\bar{\sigma}_i}{d\xi_i} = 0.$$

These equations simplify to,

$$\frac{d\bar{F}}{d\xi_i} = (-1)^{i+1}t^{k_{\sigma_i}-k_F+\beta}\bar{\sigma}_i + \sum_{j\neq i}(-1)^{j+1}t^{k_{\sigma_j}+k_{v_j}-k_F}\bar{\sigma}_j\frac{d\bar{v}_j}{d\xi_i},$$

$$\xi_i - k_x t^{k_x-1-\beta}\bar{x} + \beta t^{k_x-1-\beta}\xi_i\frac{d\bar{x}}{d\xi_i} = t^{k_F+k_x-2\beta}\bar{F}\frac{d\bar{x}}{d\xi_i},$$

$$t^{k_F-\beta+1}\bar{F}\frac{d\bar{\sigma}_i}{d\xi_i} + t^{k_F-\beta+1}\bar{\sigma}_i\frac{d\bar{F}}{d\xi_i} + k_{\sigma_i}\bar{\sigma}_i - \beta\xi_i\frac{d\bar{\sigma}_i}{d\xi_i} = 0.$$

Again, if we require these equations to be invariant under our scaling transformation then each t exponent must be zero. If this was not the case then for each value of t, the coefficients of the differential equations would change resulting in a different differential equation and thus different solution. It follows that the similarity degrees must satisfy,

$$k_x = \beta + 1, \quad k_F = \beta - 1, \quad k_{\sigma_i} = -1, \quad k_{v_i} = \beta,$$

and equations [4.5.3]–[4.5.5] become the following set of linear differential equations:

$$\frac{d\bar{F}}{d\xi_i} = (-1)^{i+1}\bar{\sigma}_i + \sum_{j\neq i}(-1)^{j+1}\bar{\sigma}_j\frac{d\bar{v}_j}{d\xi_i},$$

$$\left(\beta\xi_i - \bar{F}\right)\frac{d\bar{x}}{d\xi_i} - (\beta+1)\bar{x} + \xi_i = 0 \qquad [4.5.8]$$

and

$$\left(\beta\xi_i - \bar{F}\right)\frac{d\bar{\sigma}_i}{d\xi_i} + \left(1 - \frac{d\bar{F}}{d\xi_i}\right)\bar{\sigma}_i = 0. \qquad [4.5.9]$$

To obtain the boundary conditions for this problem, we again specialise by requiring that caustics bound the 3 dust region at $\xi_i = \pm 1$ for $\beta > 0$ (Note that the condition on β has now changed since, with an abuse of notation, β represents a different scaling parameter to that in §4.3.). Such a choice is really motivated by the equation for the caustic derived in §3.4 for the zero gravity case, i.e. equation [3.4.1].

We now have the domains for each ξ_i; they are $-\infty < \xi_1 \leq -1$, $-1 \leq \xi_2 \leq 1$ and $1 \leq \xi_3 < \infty$. Since this effectively defines the boundaries between each dust, we need to specify how dust particles are transferred. Now, for $\xi_i = \pm 1$, $v_i = \pm t^\beta$ for $i = 2, 3$ and $i = 1, 2$ respectively. Hence,

$$\frac{dx}{dt}(t, \pm t^\beta) = \pm t^\beta$$
$$\Longrightarrow x(t, \pm t^\beta) = \pm \int_c^t \tau^\beta \, d\tau.$$

If we impose the obvious symmetry requirements, i.e. if $v_i = -v_j$ then $x(t, v_i) = -x(t, v_j)$, where $j = 2, 3$ if $i = 2, 1$ respectively, then $c = 0$ and

$$x(t, \pm t^\beta) = \pm \int_0^t \tau^\beta \, d\tau$$
$$\Longrightarrow x(t, \pm t^\beta) = \pm \frac{t^{\beta+1}}{\beta + 1}$$
$$\Longrightarrow \bar{x}(\pm 1) = \pm \frac{t^{\beta+1-k_x}}{(\beta + 1)}$$
$$\Longrightarrow \bar{x}(\pm 1) = \frac{\pm 1}{(\beta + 1)}.$$

Hence the boundary conditions on $\bar{x}$ are,

$$\bar{x}(\xi_i) = -\bar{x}(\xi_j) \quad \forall \xi_i = -\xi_j, \quad \bar{x}(0) = 0, \qquad [4.5.10]$$

$$\bar{x}(1) = \frac{1}{(\beta + 1)}, \qquad [4.5.11]$$

$$\bar{x}(-1) = \frac{-1}{(\beta + 1)}, \qquad [4.5.12]$$

where again $j = 2, 3$ if $i = 2, 1$. Boundary conditions for each $\bar{\sigma}_i$ are needed if we wish to complete the specification of our problem. However, we shall see that this analysis suffers from the same problem of infinite density on the axis of symmetry as that of the tx space solution. For this reason we shall go no further than to stipulate $\bar{\sigma}_i$ to be finite everywhere (Note the more stringent restriction that the mass flux be bounded for all t and v, ref. §4.1.).

§4.6. Zero gravity case revisited.

For the case $\bar{F} = 0$, equation [4.5.8] becomes

$$\beta \xi_i \frac{d\bar{x}}{d\xi_i} - (\beta + 1)\bar{x} + \xi_i = 0, \qquad [4.6.1]$$

and this has solution

$$\bar{x}(\xi_i) = \xi_i + A_i \xi_i^{(\beta+1)/\beta} \qquad [4.6.2]$$

where the $A_i \in \mathbb{C}$ are constants of integration. We can immediately see the advantages of working in tv space for equation [4.6.1] is only singular when $\xi_2 = 0$. This should be compared with [4.3.10] which is singular in three regions and is essentially the equivalent equation in tx space. The resulting solutions ([4.6.2]) are therefore differentiable at the caustic. Furthermore, $\bar{x}(\xi)$ can be constructed by 'gluing' together the two solutions which are valid on $(-\infty, 0)$ and $(0, \infty)$. Again, this is an improvement on the equivalent solution in tx space, namely [4.4.2], where $w_i(\xi)$ is a function composed of four parts.

The constants, $A_i \in \mathbb{C}$, are fixed by [4.5.11] and [4.5.12] so that

$$A_i = \frac{\beta(-1)^{\beta/(\beta+1)}}{\beta + 1} \quad i = 1 \text{ and } i = 2 \text{ with } \xi < 0,$$

and

$$A_i = \frac{-\beta}{\beta + 1} \quad i = 2, \text{ with } \xi > 0 \text{ and } i = 3.$$

This specifies a solution subject to the boundary conditions given by [4.5.11] and [4.5.12]. However, we need to check that this is consistent with [4.5.10]. Suppose that $\xi_i > 0$ for $i = 2, 3$ and $\xi_j < 0$ for $j = 1, 2$ such that $\xi_i = -\xi_j$. Then

$$\begin{aligned}
\bar{x}(\xi_i) &= \xi_i - \frac{\beta}{\beta + 1}\xi_i^{(\beta+1)/\beta} \\
&= -\xi_j - \frac{\beta}{\beta + 1}(-\xi_j)^{(\beta+1)/\beta} \\
&= -\xi_j - \frac{(-1)^{(\beta+1)/\beta}\beta}{\beta + 1}(\xi_j)^{(\beta+1)/\beta} \\
&= -\bar{x}(\xi_j)
\end{aligned}$$

as required.

Having shown that the solution for $\bar{x}(\xi_i)$ satisfies all the constraints imposed by the boundary conditions, we can again demonstrate how this contains the cubic solution of §3.2 under an appropriate choice of β. Now the boundary condition, $\xi = \pm 1$, is effectively the equation of the caustic in TM and so is equivalent to [3.4.1]. If we again match the exponents of t and v in these equations then $\beta = 1/2$. Inserting this into our solution gives for $\xi_i = v_i/t^\beta$ and $i = 2, 3$,

$$\begin{aligned} \bar{x} &= \frac{v_i}{t^{1/2}} - \frac{1}{3} \cdot \frac{v_i^3}{t^{3/2}}, \\ \Longrightarrow \frac{x}{t^{3/2}} &= \frac{v_i}{t^{1/2}} - \frac{1}{3} \cdot \frac{v_i^3}{t^{3/2}}, \\ \Longrightarrow x &= v_i t - \frac{1}{3} v_i^3. \end{aligned}$$

Then, if we suppose that $t \longrightarrow t/3$ and $x \longrightarrow x/3$, our solution reduces to $x = q(t, v)$ as required. The case for $\xi_j = -v_j/t^\beta$ gives the same result by the symmetry argument, $\bar{x}(\xi_i) = -\bar{x}(\xi_j)$ where $j = 1, 2$ if $i = 3, 2$ respectively.

Let us now consider the differential equation for $\bar{\sigma}_i$, the transformed density, in the case when $\bar{F} = 0$. Solving [4.5.9] gives

$$\bar{\sigma}_i = B_i \xi_i^{-1/\beta} \quad \xi_i \neq 0 \qquad [4.6.3]$$

where the $B_i \in \mathbb{C}$ are constants. Now although we do not have any boundary conditions for this quantity, we are still able to analyse our solution and quite quickly see that it is giving us spurious results. In §4.1 we stated that the mass flux is finite everywhere in TM. Clearly, from equation [4.6.3], we have that as the velocity tends towards zero, σ_2 becomes unbounded contradicting our assertion. This means that our tv space solution suffers from the same problems that our tx space solution had. In terms of the density function we see that [4.6.3] implies, for $\xi_i > 0$,

$$\begin{aligned} (-1)^i \rho_i \frac{\partial x}{\partial v_i} &= \frac{B_i}{t} \xi_i^{-1/\beta} \\ \Longrightarrow (-1)^i \rho_i \left(\frac{\partial \xi_i}{\partial v_i} \frac{d}{d\xi_i} \left(t^{\beta+1} \bar{x} \right) \right) &= \frac{B_i}{t} \xi_i^{-1/\beta} \\ \Longrightarrow (-1)^i \rho_i \left(t \left(1 - \xi_i^{1/\beta} \right) \right) &= \frac{B_i}{t} \xi_i^{-1/\beta}, \end{aligned}$$

using [4.6.2] and assuming $\xi > 0$,

$$\Longrightarrow \rho_i = \frac{(-1)^i B_i}{t\left(1 - \xi_i^{1/\beta}\right)\xi_i^{1/\beta}},$$

and so again we have that $\rho_2 \longrightarrow \infty$ as $\xi_2 \longrightarrow 0$ (Recall that the axis of symmetry in tx space corresponds the $v = 0$ geodesic in the Newtonian case.).

This unwanted infinity is obviously produced by the $\xi_i^{-1/\beta}$ term in [4.6.3]. However, for the $\bar{F} = 0$ case, k_{σ_i} is effectively free to be chosen by a similar argument for k_{ρ_i} before. This means that by choosing $k_{\sigma_i} = 0$, [4.5.9] transforms to

$$\frac{d\bar{\sigma}_i}{d\xi_i} = 0,$$

which in turn forces the mass flux, σ_2, to be non-singular on $\xi_2 = 0$. We conclude this section therefore, by noting two things. The first is that although the similarity solutions obtained in tv space are perhaps more elegant than those obtained in tx space, the problems of the latter are still evident. The second point is that in the cases where $F = 0$, the similarity degrees for the mass fluxes could be arbitrarily chosen so that their corresponding solution was no longer singular on the axis of symmetry. These points are clearly analogous to those made in the concluding paragraph of §4.4. It seems that 'turning off the force', is the incorrect way of reducing our general Newtonian equations of motion to those of the gravity free case. We shall use this by way of a lead into our next chapter.

§4.7. A note on the generic condition for gravity free caustics.

During the discussion on the cubic surface in §3.2, it was mentioned that the existence of caustics depended on the existence of a point in M such that $\partial^2 x/\partial v^2 = 0$. It was also explained that the satisfaction of this condition was stable against small perturbations provided $\partial^3 x/\partial v^3 \neq 0$ and finite at this point. We can use this information to prove that β in the tv space formalism for the zero gravity problem is equal to 1/2. The procedure is to simply differentiate the solution for $x(t, v_i)$ given by [4.6.2]. We have,

$$\bar{x} = \xi_i + A_i \xi_i^{(\beta+1)/\beta}$$

$$\Longrightarrow \frac{x}{t^{\beta+1}} = \frac{v_i}{t^\beta} + A_i \frac{v_i^{(\beta+1)/\beta}}{t^{\beta+1}}$$

$$\Longrightarrow x = v_i t + A_i v_i^{(\beta+1)/\beta}$$

$$\Longrightarrow \frac{\partial^2 x}{\partial v_i^2} = A_i \frac{(\beta+1)}{\beta} \frac{1}{\beta} v_i^{(1-\beta)/\beta}$$

$$\Longrightarrow \frac{\partial^3 x}{\partial v_i^3} = A_i \frac{(\beta+1)}{\beta^2} \frac{(1-\beta)}{\beta} v_i^{(1-2\beta)/\beta},$$

and thus the second derivative is zero and the third derivative non-zero and finite at $v_2 = 0$ if and only if $\beta = 1/2$. Finally since,

$$\frac{x}{t^{\beta+1}} = \bar{x}(\xi_i),$$

is an invariant under the transformations described in equation [4.5.6], it follows that the value of $\beta = 1/2$ corresponds to a value of $\beta = 1/2 + 1 = 3/2$ if we are considering similarity solutions in tx space. This value of β is consistent with the smooth case [A].

CHAPTER 5. NEWTONIAN CAUSTICS II.

§5.1. Introduction.

In the previous chapter a model describing the physics of caustic formation was set up and analysed. The resulting differential equations were assumed to be invariant under a given set of scaling transformations and using this symmetry, similarity solutions were found for the case of zero gravity. This was done in both the tx and tv space formalism however, although the latter was more elegant, they both suffered from a single and fundamental flaw. That is to say, they both predicted unbounded densities on the axis of symmetry. We concluded the analysis in each case by noticing that it was the force equation which determined the similarity degrees of ρ_i and σ_i prior to setting $F = 0$. We also concluded that had we been allowed to freely choose k_{ρ_i} or k_{σ_i}, then a term in the corresponding differential equation could be made to vanish so that when we integrate, the unwanted infinity does not appear. This last statement suggests a way forward for us obtaining sensible solutions for the density function. Indeed, this chapter concentrates on formulating a new process that 'switches off' the gravitational interaction whilst allowing bounded densities on the axis of symmetry. This results in a greater understanding of how the gravity-free scenario fits within the more general picture.

This problem of infinite density can be split into two parts. The first is to find the correct description of how our general Newtonian equations reduce down to the zero gravity case. The second is how do we change the symmetry that we impose on our system so that the problem on the axis is resolved. We find that the first part arises quite naturally if we look at the asymptotics of our equations. This involves the introduction of what we call *asymptotic solutions.* These will be defined in the next section but their concept is strongly based on similarity solutions which were developed in §4.4. The idea is to transform our general Newtonian differential equations using the transformation group, g, developed in the previous chapter, in such a way so that as ε increases a point in the three-dust region moves along predetermined curves towards the origin. In the limit as ε tends to infinity, this point coincides with the cusp. The resulting differential equations describe the physics in a neighbourhood containing the origin. This can be put more boldly by saying that they describe the physics of cusp formation. This of course, is extremely important. From the point of view of understanding this particular type of singularity, it may tell us the extent of the role that gravity plays in determining whether or not a cusp (and then caustic) is formed.

If we return now to this problem of infinite densities on the axis, we see that the zero gravity differential equations can be obtained quite easily. Since in the limit we have our aforementioned reference point lying on the symmetry axis, we must require the force term to asymptotically tend towards zero. In other words, the 'switching off' of the force becomes a continuous process. If we assume that g represents the scaling transformations of §4.3 (as we shall), then this process is governed by the similarity degree, k_F, via the condition that $k_F < 0$.

At this point it is apparent why this asymptotic process does not solve the problem on the axis; we find that the asymptotic equations are identical to those of [4.1.2] and [4.1.3] with $F = 0$. We are therefore forced to reconstruct our transformation group and this is the second part mentioned above. We shall re-introduce the gravitational constant and assume that this can scale in much the same way as F or v_i for example. The aforementioned technique of retrieving the zero gravity case is then transferred from a restriction on k_F to a restriction on k_G. This procedure solves the problem on the axis. The way it does this, as we shall see, is to provide an extra degree of freedom which manifests itself in providing us with four similarity degrees rather than three. Since we have only three equations governing k_F, k_G, k_{v_i} and k_{ρ_i} (if we work in tx space), we find that in order to specify all the similarity degrees we are forced to choose either k_{ρ_i} or k_G. By choosing wisely, the term in the mass flux equation that when integrated produces the infinity can be removed. This work will be covered in the third section.

At this stage we can confidently say that the zero gravity case has been fully understood and that the solutions obtained are compatible with the full Newtonian picture if we consider asymptotic solutions. But, if we look again at the general equations of motion ([4.1.1]–[4.1.3]), we have to admit that no real progress has been made by way of finding a solution. This of course does not go for our understanding of the problem for we now realise the complexity and highly non-linear behaviour that we are dealing with. As a last ditch attempt at obtaining a full analytical solution, a more general approach to asymptotic/similarity solutions was considered. This involves generating the most general transformation group, g, that equations

[4.1.1]–[4.1.3] admit. The hope is that it might be possible to choose a particular transformation that is compatible with our boundary conditions and which provides us with a way of removing some of the non-linearities. Specifically, we would hope that the contributions from ρ_i and ρ_j to the acceleration at a point where a particle with velocity v_k ($i \neq j$, $i \neq k$ and $j \neq k$) exists, can be removed or simplified or perhaps represented by some symmetry transformation. This seems unlikely however, we will most definitely learn from this process and so this work is presented in §5.4.

To complete this chapter, we shall briefly discuss the asymptotic behaviour of the Newtonian equations of motion in a tr space formalism. The idea is to show that these differential equations have the same asymptotics as their equivalent partners in tx space. More importantly, however, this section, although small, will act as a prologue to the next two chapters where we consider caustics in the context of General Relativity. The significance being that we shall try to use the techniques of asymptotic solutions that we are about to develop to analyse the most general case.

§5.2. Asymptotic solutions.

The main conclusion from the previous chapter is that similarity solutions, in the context that we have used them, are incompatible with the process of discontinuously 'switching off' the force. This is not to say that similarity solutions cannot be used to obtain sensible results. Indeed, had we considered the zero gravity equations of motion in the first place, then there is no doubt that these techniques could be used successfully. The point, however, is that the procedure: look for similarity solutions, set $F = 0$, cannot be used because it passes an incomplete set of information concerning the gravitational picture through to the zero gravity picture via the similarity degrees.

We conclude that it is necessary to do two things to correct this. The first is to alter the method by which we obtain the zero gravity equations of motion from the full Newtonian picture. The second, which we leave to the next section, is to ensure that the right information is passed between the two cases. In fact ideas concerning the first point come as a result of answering the different question, does gravity play a significant role in the physics of cusp formation? Of course the answer to this has huge implications for the general case and therefore is an important question in its own right. It could be, for example, that gravity plays a key role and that given any initial situation, caustics will always be formed. Alternatively, it may have no significance whatsoever meaning that it is purely the boundary conditions that determine whether or not a cusp can form.

To answer these questions, we consider asymptotic solutions. These constructs are functions defined in the following manner. Let (t, x) and $(\tilde{t}, \tilde{x})$ represent two coordinate systems whose origins, O and $\tilde{O}$, coincide at the cusp. We assume that our tx coordinate system is fixed and 'pinned' onto our spacetime, the $\tilde{t}\tilde{x}$ coordinate system on the other hand is free to be defined. Next we define a transformation, g, that relates our two coordinate systems so that points, $\mathbf{x}$ and $\tilde{\mathbf{x}}$, are related by $\tilde{\mathbf{x}} = g(\varepsilon; \mathbf{x})$. It follows by the definition of g that our two coordinate systems are identical when $\varepsilon = 1$. We now define curves along which points can move towards the cusp in the following way. We consider $\mathbf{x}$ and $\tilde{\mathbf{x}}$ to be to be points that are fixed relative to O and $\tilde{O}$ respectively. Next we define g so that relative to tx space, the length and time scales of our $\tilde{t}\tilde{x}$ coordinate system increases as ε increases. This implies that the locus of points such as $\mathbf{x}$ describe curves in our $\tilde{t}\tilde{x}$ coordinate system as we 'sweep' this spacetime over them. Since in the limit as ε tends to infinity the length and time scales of our $\tilde{t}\tilde{x}$ spacetime become infinitely large, it follows that the distance between $\mathbf{x}$ and $\tilde{O}$ becomes infinitely small implying that $\mathbf{x}$ moves along curves in $\tilde{t}\tilde{x}$ space towards the origin. All of this is defining a magnification type of process: we are essentially taking a small and fixed region containing the origin and enlarging this according to the rules specified by g. As this process occurs, the Newtonian differential equations written in terms of (t, x) coordinates must change since $\mathbf{x}$ approaches $\tilde{O}$. This implies that the *solution* to these transformed differential equations must also change as certain terms become less significant. We call these limiting solutions asymptotic solutions. Since we have defined g so that small regions containing the cusp are magnified, these solutions must be describing how our dependent variables behave in an infinitesimally small region containing the cusp at the instant of cusp formation. Their behaviour gives us the physics that determine this process.

The above ideas can be formulated in terms of a mathematical definition: we define the asymptotic solution, f_a, of any variable to be $f_a(\mathbf{x}) = \lim_{\varepsilon \to \infty} g_f(\varepsilon; \tilde{\mathbf{x}}, \tilde{f}(\tilde{\mathbf{x}}))$. The existence of these functions, or equivalently the fact that the differential equations have a limiting form, is an assumption that we have to make. Now the above definition implies that asymptotic solutions are also similarity solutions. This is an artifact of the group structure of the transformation, g, which appears in both definitions. To prove that this is so, we need to show that $f_a(g_{\mathbf{x}}(\alpha;\mathbf{x})) = g_f(\alpha;\mathbf{x},f_a(\mathbf{x}))$ which is equation [4.2.3]. We begin by rewriting the definition of our asymptotic solutions so that as functions, f_a and its definition have the same arguments, i.e.,

$$f_a(\mathbf{x}) = \lim_{\varepsilon\to\infty} g_f(\varepsilon; g_x(\varepsilon^{-1};\mathbf{x}), f(g_x(\varepsilon^{-1};\mathbf{x}))). \qquad [5.2.1]$$

Note that the tilde has been removed from f since this is purely a label. Then,

$$\begin{aligned} g_f(\alpha;\mathbf{x},f_a(\mathbf{x})) &= g_f\left(\alpha;\mathbf{x},\lim_{\varepsilon\to\infty} g_f\left(\varepsilon; g_{\mathbf{x}}\left(\varepsilon^{-1};\mathbf{x}\right), f\left(g_{\mathbf{x}}\left(\varepsilon^{-1};\mathbf{x}\right)\right)\right)\right) \\ &= \lim_{\varepsilon\to\infty} g_f\left(\alpha;\mathbf{x}, g_f\left(\varepsilon; g_{\mathbf{x}}\left(\varepsilon^{-1};\mathbf{x}\right), f\left(g_{\mathbf{x}}\left(\varepsilon^{-1};\mathbf{x}\right)\right)\right)\right) \\ &= \lim_{\varepsilon\to\infty} g_f\left(\alpha\varepsilon; g_{\mathbf{x}}\left(\varepsilon^{-1};\mathbf{x}\right), f\left(g_{\mathbf{x}}\left(\varepsilon^{-1};\mathbf{x}\right)\right)\right) \\ &= \lim_{\beta\to\infty} g_f\left(\beta; g_{\mathbf{x}}\left(\alpha\beta^{-1};\mathbf{x}\right), f\left(g_{\mathbf{x}}\left(\alpha\beta^{-1};\mathbf{x}\right)\right)\right) \\ &= \lim_{\beta\to\infty} g_f\left(\beta; g_{\mathbf{x}}\left(\beta^{-1}; g_{\mathbf{x}}(\alpha;\mathbf{x})\right), f\left(g_{\mathbf{x}}\left(\beta^{-1}; g_{\mathbf{x}}(\alpha;\mathbf{x})\right)\right)\right) \\ &= f_a(g_{\mathbf{x}}(\alpha;\mathbf{x})) \end{aligned}$$

as required.

We are now in a position to analyse equations [4.1.2] and [4.1.3] using these new techniques. To do this we introduce an intermediate, dependant variable, f_ε, such that

$$f_a(\mathbf{x}) = \lim_{\varepsilon\to\infty} f_\varepsilon(\mathbf{x}) = \lim_{\varepsilon\to\infty} g_f\left(\varepsilon; g_x\left(\varepsilon^{-1};\mathbf{x}\right), f\left(g_x\left(\varepsilon^{-1};\mathbf{x}\right)\right)\right). \qquad [5.2.2]$$

This simply allows us to separate the two operations: assume similarity solutions and take the limit. The symmetry group we shall use is that of equation [4.3.1] and we shall begin by assuming that all dependant variables are invariant under these transformations. From [5.2.2] we have,

$$g_f\left(\varepsilon^{-1};\mathbf{x},f_\varepsilon(\mathbf{x})\right) = g_f\left(1, g_{\mathbf{x}}\left(\varepsilon^{-1};\mathbf{x}\right), f\left(g_{\mathbf{x}}\left(\varepsilon^{-1};\mathbf{x}\right)\right)\right) = f\left(g_{\mathbf{x}}\left(\varepsilon^{-1};\mathbf{x}\right)\right)$$

$$\Longrightarrow g_f\left(\varepsilon^{-1}; g_{\mathbf{x}}(\varepsilon;\mathbf{x}), f_\varepsilon(g_{\mathbf{x}}(\varepsilon;\mathbf{x}))\right) = f(\mathbf{x})$$

$$\Longrightarrow \varepsilon^{-k_f} f_\varepsilon(g_{\mathbf{x}}(\varepsilon;\mathbf{x})) = f(\mathbf{x}),$$

and this is the basis by which we transform our equations from $\mathbf{x}f$ space into $\mathbf{x}_\varepsilon f_\varepsilon$ space (Note that in the above calculation the coordinates, $\mathbf{x}_\varepsilon$ have been written as

$g_{\mathbf{x}}(\varepsilon;\mathbf{x}).$). Of course we need to consider how the derivatives of f transform. Thus, if our coordinates are (t,x) then using [4.3.1] we find that for any function,

$$\begin{aligned}\frac{\partial f}{\partial x}(t,x) &= \frac{\partial}{\partial x}\left\{g_f\left(\varepsilon^{-1}; g_{\mathbf{x}}(\varepsilon;\mathbf{x})\, f_\varepsilon\left(g_{\mathbf{x}}(\varepsilon;\mathbf{x})\right)\right)\right\}(t,x)\\ &= \frac{\partial}{\partial x}\left\{\varepsilon^{-k_f} f_\varepsilon\left(\varepsilon^{\beta} x\right)\right\}(t,x)\\ &= \varepsilon^{\beta-k_f}\frac{\partial f_\varepsilon}{\partial x_\varepsilon}(t_\varepsilon, x_\varepsilon).\end{aligned}$$

Similarly, we obtain

$$\frac{\partial f}{\partial t}(t,x) = \varepsilon^{1-k_f}\frac{\partial f_\varepsilon}{\partial t_\varepsilon}(t_\varepsilon, x_\varepsilon)$$

by simply replacing the similarity degree for x by that for t, i.e. by changing β for 1. We should note that since it is clear that we are now working in $\mathbf{x}_\varepsilon f_\varepsilon$ space (note the subscript on the f_ε), the subscript ε will be dropped from the coordinate variables. Using these results, equations [4.1.2] and [4.1.3] become,

$$v_{i\varepsilon}\frac{\partial v_{i\varepsilon}}{\partial x} + \varepsilon^{k_{v_i}-\beta+1}\frac{\partial v_{i\varepsilon}}{\partial t} = \varepsilon^{2k_{v_i}-\beta-k_F} F_\varepsilon$$

and

$$\rho_{i\varepsilon}\frac{\partial v_{i\varepsilon}}{\partial x} + v_{i\varepsilon}\frac{\partial \rho_{i\varepsilon}}{\partial x} = -\varepsilon^{k_{v_i}-\beta+1}\frac{\partial \rho_{i\varepsilon}}{\partial t}.$$

We need not concern ourselves with the force equation ([4.1.1]) for this becomes irrelevant in the limit.

Again we require that these equations are invariant; the functions f_ε are, by definition, similarity solutions. This means that the similarity degrees must satisfy $k_{v_i} - \beta + 1 = 0$ and $2k_{v_i} - \beta - k_F = 0$. We now take the limit as $\varepsilon \longrightarrow \infty$. This is a continuous process and it replaces the discontinuous operation of 'switching off' the force. The quantity that governs this limiting process is the similarity degree, k_F. We must have, in order for any force terms to tend towards zero, $k_F < 0$. Thus $F_a = \lim_{\varepsilon\to\infty} F_\varepsilon = \lim_{\varepsilon\to\infty} \varepsilon^{k_F} F = 0$ and we therefore finally obtain

$$v_{ia}\frac{\partial v_{ia}}{\partial x} + \frac{\partial v_{ia}}{\partial t} = 0 \qquad [5.2.3]$$

and

$$\rho_{ia}\frac{\partial v_{ia}}{\partial x} + v_{ia}\frac{\partial \rho_{ia}}{\partial x} = -\frac{\partial \rho_{ia}}{\partial t}. \qquad [5.2.4]$$

Again we draw the attention of the reader to the fact that $\mathbf{x}$ now represents the coordinate system, $\mathbf{x}_\varepsilon$, in the limit as $\varepsilon \longrightarrow \infty$.

On the face of it, this looks like an extremely important result. It seems we have shown that by considering asymptotic solutions of the form [5.2.1], the equations of motion for the general Newtonian picture reduce to those of the gravity free case. We *could* therefore conclude that gravity does not contribute to cusp formation. The problem with this however, is that v_{i_a} and ρ_{i_a} satisfy the same equations as v_i and ρ_i of §4.4 with the same similarity degrees. In other words the solutions are identical! The above conclusion concerning the role that gravity plays must therefore be treated with scepticism since clearly it is based on a result that predicts unbounded densities on the axis. The only achievement that this asymptotic approach has above the original similarity solution approach is to formulate the turning off of the gravitational force in a mathematically elegant manner.

§5.3. Asymptotic solutions with a scaled Newtonian constant.

The above section describes a new procedure by which we can obtain the zero gravity equations of motion from the general case. The problem of unbounded densities on the axis however, has not been solved and we now propose to do this. As mentioned in both the introduction and in the title of this section, we will achieve this by scaling the gravitational constant.

We shall consider the following equations:

$$\frac{\partial F}{\partial x} = -G\sum_i \rho_i,$$

$$v_i\frac{\partial v_i}{\partial x} + \frac{\partial v_i}{\partial t} = F$$

and

$$\rho_i\frac{\partial v_i}{\partial x} + v_i\frac{\partial \rho_i}{\partial x} = -\frac{\partial \rho_i}{\partial t},$$

alongside the transformation given by,

$$g(\varepsilon; t, x, F, v_i, \rho_i, G) = (\varepsilon^{k_t}t, \varepsilon^{k_x}x, \varepsilon^{k_F}F, \varepsilon^{k_{v_i}}v_i, \varepsilon^{k_{\rho_i}}\rho_i, \varepsilon^{k_G}G). \qquad [5.3.1]$$

We again look for asymptotic solutions but instead of having the limiting process governed by F or rather by k_F, we insure that the force tends to zero by imposing the boundary conditions, $k_G < 0$ and $F(t,0) = 0$. The combination of all these ideas allow us to obtain sensible results for the density as we shall now demonstrate.

We introduce four intermediate dependant functions, F_ε, $v_{i\varepsilon}$, $\rho_{i\varepsilon}$ and G_ε. They are defined by equation [5.2.2]. Remember that they are similarity solutions which means that we now have

$$\frac{\partial F_\varepsilon}{\partial x} = -G_\varepsilon \sum_i \rho_{i\varepsilon},$$

$$v_{i\varepsilon}\frac{\partial v_{i\varepsilon}}{\partial x} + \frac{\partial v_{i\varepsilon}}{\partial t} = F_\varepsilon$$

and

$$\rho_i \frac{\partial v_{i\varepsilon}}{\partial x} + v_{i\varepsilon}\frac{\partial \rho_{i\varepsilon}}{\partial x} = -\frac{\partial \rho_{i\varepsilon}}{\partial t},$$

where the similarity degrees, $k_x = \beta k_t$, $k_F = (\beta-2)k_t$, $k_{v_i} = (\beta-1)k_t$, $k_{\rho_i} = -(2+\gamma)k_t$ and $k_G = \gamma k_t$, have been chosen for invariance of the above equations.

The force equation must be included this time because at the moment we do not specifically have $F_a = 0$. To achieve this we need to consider our boundary conditions. The first boundary condition is that $k_G < 0$ meaning that $G_a = \lim_{\varepsilon\to\infty} G_\varepsilon = \lim_{\varepsilon\to\infty} \varepsilon^{k_G} G = 0$. If we apply this result to the force equation then it follows that in the limit we have,

$$\frac{\partial F_a}{\partial x} = 0.$$

This gives F_a independent of x. The boundary condition $F(t,0) = 0$ implies that $F_a = 0$. We therefore obtain

$$v_{ia}\frac{\partial v_{ia}}{\partial x} + \frac{\partial v_{ia}}{\partial t} = 0$$

and

$$\rho_i \frac{\partial v_{ia}}{\partial x} + v_{ia}\frac{\partial \rho_{ia}}{\partial x} = -\frac{\partial \rho_{ia}}{\partial t}$$

as expected.

Now comes the new bit. As mentioned in §5.2, asymptotic solutions are similarity solutions where for any variable, f_a, the k_f are as defined above. If we transform these equations using the notation that $f_a(t,x) = \varepsilon^{k_f}\tilde{f}(\xi)$ then we obtain

$$\tilde{v}_i\frac{d\tilde{v}_i}{d\xi} + (\beta-1)\tilde{v}_i - \beta\xi\frac{d\tilde{v}_i}{d\xi} = 0 \qquad [5.3.2]$$

and

$$\tilde{v}_i\frac{d\tilde{\rho}_i}{d\xi} + \tilde{\rho}_i\frac{d\tilde{v}_i}{d\xi} = (2+\gamma)\rho_i + \beta\xi\frac{d\tilde{\rho}_i}{d\xi}. \qquad [5.3.3]$$

The difference of this approach compared to that of §4.4 can now be seen Although the velocity equation ([5.3.2]) is identical in both cases, the density equation is not. We notice that the coefficient of the $\tilde{\rho}_i$ term has changed from 2 to $2+\gamma$. We will see the relevance of this in a moment.

The boundary conditions have not yet been mentioned. For the velocity equation, we choose the same boundary conditions as in §4.4 and obtain [4.4.2] as a solution. For the density we shall simply show that it now becomes bounded on the axis. To fully specify this function we need to prescribe the density function along some curve $\xi = \text{const}$. We shall not bother to do this as this corresponds to an arbitrary choice. The most appealing candidate however, is the curve corresponding to the caustic. This amounts to prescribing the density of dusts 1 and 3 as their particles traverse this curve and prescribing the density of dust 2 as it receives particles from the other two dusts at the caustic.

So, in order to check that the density function can now be made finite everywhere, we write the velocity in the same manner as before to illustrate the singular nature of our differential equations, i.e., $\tilde{v}_i = w_i + \beta\xi$. The solution to equation [5.3.3] is then

$$\tilde{\rho}_i = \frac{1}{|w_i|}\exp\left\{\int_{a_i}^{\xi}\frac{(2+\gamma-\beta)}{w_i}\,d\tilde{\xi}\right\},$$

where $a_i \in \mathbb{R}$ again represents the constants of integration. Since w_i has the same solution as before, the values of ξ/w_2 and w_2' as ξ tends to zero are the same. By following the analysis as in §4.4 we have that

$$\lim_{\xi\to 0}\tilde{\rho}_2 = \lim_{\xi\to 0}\left\{\pm\frac{(2+\gamma-\beta)}{(1-\beta)^2}\left(\frac{\xi}{a_i}\right)^{(1+\gamma)/(1-\beta)}\frac{1}{a_i}\right\}.$$

Thus $\tilde{\rho}_2$ is finite and non-zero at $\xi = 0$ if $\gamma = -1$.

We are now in a position to make several conclusions. Firstly, we have shown that by considering asymptotic solutions together with the new symmetry group given by [5.3.1], the equations of motion for the general picture again reduce to the gravity free case. Secondly, we have shown that the corresponding asymptotic solution for the density now predicts finite behaviour on the symmetry axis. These two results are extremely important for it means that the physics governing cusp formation is relatively simple. In other words, we can finally conclude that gravity does not contribute except to shape the caustic to the future of the cusp. Also, by using the idea that G should scale and determine the asymptotics of our system rather than F, we have understood how information is passed from the full gravitational picture to the zero gravity picture in our reduction process. This is important for we shall again use these ideas when we consider the asymptotics of the General Relativistic equations of motion.

§5.4. The general symmetry group.

Although the gravity free case has been analysed and how it can be realised as a limit of the full Newtonian equations of motion understood, no real headway into solving the general case has been made. Therefore, in this section, we shall try to determine the most general symmetry group that equations [4.1.1]–[4.1.3] admit. The idea is that possibly, by choosing carefully the symmetry group, g, we may find similarity solutions which greatly reduce the degree of non-linearity that this problem possesses. An excellent account of how to find these general transformations for a system of first order partial differential equations is given in [L]. Hence we shall simply give an overview of the techniques used and state the results.

Before we launch into the mathematics, it might be instructive to briefly describe the process that we will use to determine our symmetry transformations. This technique hinges on the fact that we assume our one-parameter group of transformations, g, to be analytic functions of ε. This means that for ε close to the identity, we can construct an approximation to g by writing each transformation as a first order Taylor series. These *infinitesimal transformations* define quantities which we call the *generators* of g. We shall denote this set of functions by $(\mathbf{X}, \mathbf{U})$, the dimensions of these vectors being m (the number of independent variables) and n (the number of dependant variables) respectively. Now g does not tell us any information about how the derivatives of our functions transform. Since we are dealing with first order partial differential equations, this is important and we need to know this. We shall therefore extend g to include transformations of all possible derivatives. This new group we shall call g_E. If we suppose that g_E is also analytic, then by a similar argument we obtain $m + n + m \times n$ generators which we denote by $(\mathbf{X}, \mathbf{U}, \mathbf{P})$.

The generators turn out to be important quantities because we can develop a technique that presents the most general symmetry transformation as the solution to a set of simultaneous equations in $\mathbf{X}$, $\mathbf{U}$ and $\mathbf{P}$. However, since we are only interested in obtaining similarity solutions for $\mathbf{u}$, we need only determine g, and we can therefore simplify things if we write each component of $\mathbf{P}$ in terms of the other $m+n$ generators. This process gives us a set of linear equations in p_i^j, the coefficients of which are functions of $\mathbf{X}$, $\mathbf{U}$ and their derivatives. Since each p_i^j are independent, we can equate coefficients to give (usually) a huge set of simultaneous, differential equations. Their solution gives us $\mathbf{X}$ and $\mathbf{U}$. We shall show that these generators are related to $g_{\mathbf{x}}$ and $g_{\mathbf{u}}$ by a simple differential equation. The problem with this, however, is that it is not always soluble. In fact the same problem occurs when finding the most general similarity solution, for any candidate satisfies a similar differential equation.

We begin by introducing a new quantity that parameterises our transformation group, g. In fact if ε is the parameter of §4.2, then we suppose that $\varepsilon \longrightarrow e^{\varepsilon}$ so that now $\varepsilon \in (-\infty, \infty)$. This simply changes the identity from 1 to 0, i.e. $g(0; \mathbf{x}, f(\mathbf{x})) = (\mathbf{x}, f(\mathbf{x}))$. Let us also define $\mathbf{u} = \left(u^1, \ldots, u^m\right) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$. Then the transformation, g, can be written as

$$\tilde{x}^i = g_{x^i}(\varepsilon; \mathbf{x}), \quad i = 1, 2,$$

if we index our coordinates by a superscript rather than a subscript, and

$$\tilde{u}^j = g_{u^j}(\varepsilon; \mathbf{x}, \mathbf{u}), \quad j = 1, \ldots, m.$$

We now assume that g is analytic. In other words, for $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$ close to the identity,

$$\tilde{x}^i = g_{x^i}(0;\mathbf{x}) + \varepsilon \frac{\partial g_{x^i}}{\partial \varepsilon}(0;\mathbf{x}) + o(\varepsilon)$$
$$= x^i + \varepsilon X^i(\mathbf{x}) + o(\varepsilon).$$

Similarly,

$$\tilde{u}^j = u^j + \varepsilon U^j(\mathbf{x},\mathbf{u}) + o(\varepsilon).$$

The quantities X^i and U^j given by

$$X^i(\mathbf{x}) = \frac{\partial g_{x^i}}{\partial \varepsilon}(0;\mathbf{x}) \quad \text{and} \quad U^j(\mathbf{x},\mathbf{u}) = \frac{\partial g_{u^j}}{\partial \varepsilon}(0;\mathbf{x},\mathbf{u})$$

are called the generators of the transformation group and are uniquely determined by g. Conversely, the generators, X^i and U^j, determine g itself. To see this, let us suppose that $(\mathbf{x},\mathbf{u})$ represents a point fixed in $\mathbb{R}^{m+2}$. Then $g(\varepsilon;\mathbf{x},\mathbf{u})$ corresponds to a curve, parameterized by ε, that passes through $(\mathbf{x},\mathbf{u})$ when $\varepsilon = 0$, with tangent vector, $\partial g/\partial\varepsilon(0;\mathbf{x},\mathbf{u}) = \left(X^1(\mathbf{x},\mathbf{u}), X^2(\mathbf{x},\mathbf{u}), U^1(\mathbf{x},\mathbf{u}), \ldots, U^m(\mathbf{x},\mathbf{u})\right)$. Of course we can repeat this argument for any point such as $(\tilde{\mathbf{x}},\tilde{\mathbf{u}}) = g(\tilde{\varepsilon};\mathbf{x},\mathbf{u})$ for example. This implies that the generators form a vector field on $\mathbb{R}^{m+2}$ which has integral curves given by g. It follows then that to determine the group from the generators we simply solve

$$\frac{dg_{x^i}}{d\varepsilon}(0;\tilde{\mathbf{x}}) = X^i(\tilde{\mathbf{x}}) \quad \text{and} \quad \frac{dg_{u^j}}{d\varepsilon}(0;\tilde{\mathbf{x}},\tilde{\mathbf{u}}) = U^j(\tilde{\mathbf{x}},\tilde{\mathbf{u}}). \qquad [5.4.1]$$

Uniqueness is given by the boundary conditions, $\tilde{\mathbf{x}} = \mathbf{x}$ and $\tilde{\mathbf{u}} = \mathbf{u}$ at $\varepsilon = 0$.

Let us now introduce a further set of dependant variables given by $\mathbf{p} = \left(p_i^j\right) = \left(\partial u^j/\partial x^i\right)$. These quantities can be used to extend g so that it now includes transformations for each derivative. We shall denote this extension by g_E. Infinitesimally, this new group has the form,

$$\tilde{x}^i = x^i + \varepsilon X^i(\mathbf{x}) + o(\varepsilon),$$

$$\tilde{u}^j = u^j + \varepsilon U^j(\mathbf{x},\mathbf{u}) + o(\varepsilon)$$

and

$$\tilde{p}_i^j = p_i^j + \varepsilon P_i^j(\mathbf{x},\mathbf{u},\mathbf{p}) + o(\varepsilon),$$

where P_i^j are the generators of the derivative transformations in g_E. Now it is possible, although too lengthy to be given here in this document, to write the function that transforms each p_i^j in terms of the functions $g_{x^i}(\varepsilon; \mathbf{x})$ and $g_{u^j}(\varepsilon; \mathbf{x}, \mathbf{u})$ (a simple account for the case when $m = 2$ and $n = 1$ is given in [L]). Hence it follows that the generators, P_i^j, can be written down in terms X^i and U^j. We shall simply quote the result here,

$$P_i^j = \frac{\partial U^j}{\partial x^i} + \sum_{k=1}^{n} \frac{\partial U^j}{\partial u^k} p_i^k - \sum_{k=1}^{2} \frac{\partial X^k}{\partial x^i} p_k^j. \qquad [5.4.2]$$

The above ideas are the building blocks with which we develop the theory to determine the most general group of transformations (g in other words) that a system of first order partial differential equations admit. To do this, we need to modify our definition of invariance ([4.2.1]) to include a system of partial differential equations.

Definition. *Suppose that*

$$F^{(r)}(\mathbf{x}, \mathbf{u}, \mathbf{p}) = 0 \qquad [5.4.3]$$

represents a system of R first order partial differential equations. Then [5.4.3] *is said to be constantly conformally invariant under g_E if, and only if,*

$$F^{(r)}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}, \tilde{\mathbf{p}}) = \sum_{s=1}^{R} A_{rs}(\varepsilon) F^{(s)}(\mathbf{x}, \mathbf{u}, \mathbf{p}), \qquad [5.4.4]$$

where $A_{rs}(0) = \delta_{rs}$.

Unfortunately, although this is useful if one is considering stretching transformations as in §5.3, for example, this definition is not of the form where it can be effectively applied to the general case. We therefore reformulate this using the following

Proposition. *The system of R partial differential equations, $F^{(r)}(\mathbf{x}, \mathbf{u}, \mathbf{p}) = 0$, is constant conformally invariant under the group g_E if, and only if,*

$$\frac{\partial}{\partial \varepsilon} F^{(r)}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}, \tilde{\mathbf{p}}) \Big|_{\varepsilon=0} = \sum_{s=1}^{R} k_{rs} F^{(s)}(\mathbf{x}, \mathbf{u}, \mathbf{p}). \qquad [5.4.5]$$

Proof. Suppose the R differential equations are constant conformally invariant as in the above definition. Then if we differentiate equation [5.4.4] we have,

$$\frac{\partial}{\partial \varepsilon} F^{(r)}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}, \tilde{\mathbf{p}})\bigg|_{\varepsilon=0} = \sum_{s=1}^{R} \frac{\partial A_{rs}}{\partial \varepsilon}\bigg|_{\varepsilon=0} F^{(s)}(\mathbf{x}, \mathbf{u}, \mathbf{p})$$
$$= \sum_{s=1}^{r_n} k_{rs} F^{(s)}(\mathbf{x}, \mathbf{u}, \mathbf{p})$$

Running the proof in reverse proves the converse.

Corollary. *The system of partial differential equations, $F^{(r)}(\mathbf{x}, \mathbf{u}, \mathbf{p}) = 0$ is invariant under g_E if, and only if,*

$$\mathcal{D} F^{(r)} = \sum_{s=1}^{R} k_{rs} F^{(s)}, \qquad [5.4.6]$$

where,

$$\mathcal{D} = X^1 \frac{\partial}{\partial x^1} + X^2 \frac{\partial}{\partial x^2} + U^1 \frac{\partial}{\partial u^1} + \ldots + U^n \frac{\partial}{\partial u^n} + P_1^1 \frac{\partial}{\partial p_1^1} + \ldots + P_1^n \frac{\partial}{\partial p_1^n}$$
$$+ P_2^1 \frac{\partial}{\partial p_2^1} + \ldots + P_2^n \frac{\partial}{\partial p_2^n}. \qquad [5.4.7]$$

Proof. One simply expands the left hand side of [5.4.5].

This gives us a way of determining the generators of our symmetry group. We can see that equation [5.4.6] is simply a set of R simultaneous equations in the components of $(\mathbf{X}, \mathbf{U}, \mathbf{P})$. If, however, we are looking for similarity solutions, then we are not really interested in determining each P_i^j. Indeed, with the equations in their current form, we do not have enough information to solve them. The usual thing therefore, is to substitute each P_i^j by the sum of derivatives of $\mathbf{X}$ and $\mathbf{U}$ as given by equation [5.4.2]. This yields a first order, homogeneous polynomial in $\mathbf{p}$. Since the components of $\mathbf{p}$ can be considered as independent variables, we set the coefficients of each p_i^j to be zero. The result is a rather large set of simultaneous differential equations in the components of $\mathbf{X}$ and $\mathbf{U}$. In theory this set of equations can be solved to obtain X^i and U^j. Equation [5.4.1] then gives us g. This last step, however, is usually the deciding factor for whether or not we can obtain g depends on the complexity of $\mathbf{X}$ and $\mathbf{U}$.

To illustrate this process, we shall now apply these techniques to our general Newtonian equations of motion. We first of all rewrite equations [4.1.1]–[4.1.3] so that they fit the formalism of [5.4.3]. We therefore have $R = 3$, $i = 2$ and $j = 4$. If $\mathbf{x} = (x^1, x^2) = (t, x)$ and $\mathbf{u} = (u^1, u^2, u^3, u^4) = (F, v, \rho, G)$ then we have

$$F^{(1)}(\mathbf{x}, \mathbf{u}, \mathbf{p}) = p_1^1 + u^3 u^4,$$

$$F^{(2)}(\mathbf{x}, \mathbf{u}, \mathbf{p}) = u^2 p_1^2 + p_2^2 - u^1$$

and

$$F^{(3)}(\mathbf{x}, \mathbf{u}, \mathbf{p}) = u^3 p_1^2 + u^2 p_1^3 + p_2^3.$$

The Newtonian constant is special in the sense that we do not require it to be a function of t or x; we would expect a spacetime that contains matter which is uniformly distributed, to have the same strength gravitational field throughout. The upshot of this statement is that

$$P_i^4 = \sum_{k=1}^{4} \frac{\partial U^4}{\partial u^k} p_i^k - \sum_{k=1}^{2} \frac{\partial X^k}{\partial x^i} p_k^4. \qquad [5.4.8]$$

We also, for simplicity, consider a single dust and so drop the subscript i. Since we are really only duplicating equations by having the extra variables, we do not expect any new symmetries. We therefore suppose that the i in equations [4.1.1]–[4.1.3] equals 1 and assume that for the case where $i \neq 1$, v_i and ρ_i posses the same symmetries as v and ρ.

The condition for invariance (equation [5.4.6]) yields the following set of simultaneous equations:

$$P_1^1 + U^3 u^4 + u^3 U^4 = k_{11}(p_1^1 + u^3 u^4) + k_{12}(u^2 p_1^2 + p_2^2 - u^1) + k_{13}(u^3 p_1^2 + u^2 p_1^3 + p_2^3),$$

$$u^2 P_1^2 + U^2 p_1^2 + P_2^2 - U^1 = k_{21}(p_1^1 + u^3 u^4) + k_{22}(u^2 p_1^2 + p_2^2 - u^1) + k_{23}(u^3 p_1^2 + u^2 p_1^3 + p_2^3)$$

and

$$u^3 P_1^2 + U^3 p_1^2 + u^2 P_1^3 + U^2 p_1^3 + P_2^3 = k_{31}(p_1^1 + u^3 u^4) + k_{32}(u^2 p_1^2 + p_2^2 - u^1) + k_{33}(u^3 p_1^2 + u^2 p_1^3 + p_2^3).$$

If we substitute for each P_i^j using equations [5.4.2] and [5.4.8] and equate all coefficients of p_i^j to zero as suggested above, we obtain the following set of simultaneous differential equations:

$$\frac{\partial U^1}{\partial u^1} - \frac{\partial X^1}{\partial x} - k_{11} = 0,$$

$$\frac{\partial U^1}{\partial u^2} = \frac{\partial U^1}{\partial u^3} = \frac{\partial U^1}{\partial u^4} = 0,$$

$$\frac{\partial X^2}{\partial x} = 0,$$

$$\frac{\partial U^1}{\partial x} + u^4 U^3 + u^3 U^4 - k_{11} u^3 u^4 = 0,$$

$$\frac{\partial U^2}{\partial u^1} = \frac{\partial U^2}{\partial u^4} = 0,$$

$$u^2 \frac{\partial U^2}{\partial u^2} - u^2 \frac{\partial X^1}{\partial x} + U^2 - \frac{\partial X^1}{\partial t} - k_{22} u^2 - k_{23} u^3 = 0,$$

$$\frac{\partial U^2}{\partial u^3} - k_{23} = 0,$$

$$\frac{\partial U^2}{\partial u^2} - \frac{\partial X^2}{\partial t} - k_{22} = 0,$$

$$u^2 \frac{\partial U^2}{\partial x} + \frac{\partial U^2}{\partial t} - U^1 + k_{22} u^1 = 0,$$

$$\frac{\partial U^3}{\partial u^1} = \frac{\partial U^3}{\partial u^4} = 0,$$

$$u^3 \frac{\partial U^2}{\partial u^2} - u^3 \frac{\partial X^1}{\partial x} + U^3 + u^2 \frac{\partial U^3}{\partial u^2} - k_{32} u^2 - k_{33} u^3 = 0,$$

$$u^3 \frac{\partial U^2}{\partial u^3} + u^2 \frac{\partial U^3}{\partial u^3} - u^2 \frac{\partial X^1}{\partial x} + U^2 - \frac{\partial X^1}{\partial t} - k_{33} u^2 = 0,$$

$$\frac{\partial U^3}{\partial u^2} - k_{32} = 0,$$

$$\frac{\partial U^3}{\partial u^3} - \frac{\partial X^2}{\partial t} - k_{33} = 0$$

and

$$u^3\frac{\partial U^2}{\partial x}+u^2\frac{\partial U^3}{\partial x}+\frac{\partial U^3}{\partial t}+k_{32}u^1=0.$$

The solution to the above set of equations is a long and laborious task. We shall therefore simply state the results:

$$X^1=(k_{22}-k_{11})\,x+h(t),$$

$$X^2=-\frac{k_{11}}{2}t+c_1,$$

$$U^1=k_{22}u^1+h_{tt}(t),$$

$$U^2=\left(k_{22}-\frac{k_{11}}{2}\right)u^2+h_t(t),$$

$$U^3=\left(k_{33}-\frac{k_{11}}{2}\right)u^3$$

and

$$U^4=\left(\frac{3k_{11}}{2}-k_{33}\right)u^4.$$

In the above, c_1 is a constant and $h(t)$ is an arbitrary function with first and second order derivatives, $h_t(t)$ and $h_{tt}(t)$. Using equation [5.4.1], we can now determine our general symmetry group. Again the integration is complicated so we list the results. We have:

$$\begin{aligned}
\tilde{x}&=(t-d)^{\beta}\,e^{\beta\zeta}\left\{H(\tilde{t})-H(t)\right\}+xe^{\beta\zeta},\\
\tilde{t}&=te^{\zeta}+d\left(1-e^{\zeta}\right),\\
\tilde{u}^1&=(t-d)^{\beta}\,e^{\beta\zeta}\left\{\frac{d^2H}{d\tilde{t}^2}(\tilde{t})-e^{-2\zeta}\frac{d^2H}{dt^2}(t)\right\}\\
&\quad+2\beta\,(t-d)^{\beta-1}\,e^{(\beta-1)\zeta}\left\{\frac{dH}{d\tilde{t}}(\tilde{t})-e^{-\zeta}\frac{dH}{dt}(t)\right\}\\
&\quad+\beta\,(\beta-1)\,(t-d)^{\beta-2}\,e^{(\beta-2)\zeta}\left\{H(\tilde{t})-H(t)\right\}+u^1e^{(\beta-2)\zeta},\\
\tilde{u}^2&=(t-d)^{\beta}\,e^{\beta\zeta}\left\{\frac{dH}{d\tilde{t}}(\tilde{t})-e^{-\zeta}\frac{dH}{dt}(t)\right\}\\
&\quad+\beta\,(t-d)^{\beta-1}\,e^{(\beta-1)\zeta}\left\{H(\tilde{t})-H(t)\right\}+u^2e^{(\beta-1)\zeta},\\
\tilde{u}^3&=u^3e^{-(2+\gamma)\zeta},\\
\tilde{u}^4&=u^4e^{\gamma\zeta}
\end{aligned}\qquad [5.4.9]$$

where $d = 2c_1/k_{11}$, $\beta = 2\,(k_{11} - k_{22})\,/k_{11}$, $\zeta = -k_{11}\varepsilon/2$, $\gamma = 3k_{33}/k_{11} - 3$ and,

$$H(t) = -2k_{11}^{\beta} \int_{c_2}^{t} h(s)\,(s-d)^{-(\beta+1)}\,ds.$$

The right hand side of these equations should technically be functions of t and x only. However, in an attempt to obtain some brevity, we write some terms as functions of $\tilde{t}$.

This set of functions represents the components of the most general transformation that leaves equations [4.1.1]–[4.1.3] invariant given the assumption that $\tilde{\mathbf{x}} = g_{\mathbf{x}}(\varepsilon;\mathbf{x})$ and $\tilde{\mathbf{u}} = g_{\mathbf{u}}(\varepsilon;\mathbf{x},\mathbf{u})$. Of course to complete the argument we need to prove that this is the case, and we do so by brute force. We firstly consider equation [4.1.2]. The derivatives of the velocity transform as,

$$\begin{aligned}
\frac{\partial v_i}{\partial x}(t,x) &= e^{-(\beta-1)\zeta}\left[\frac{\partial \tilde{v}_i}{\partial x}(t,x)\right.\\
&\quad\left.-\frac{\partial}{\partial x}\left[(t-d)^{\beta}\,e^{\beta\zeta}\left\{\frac{dH}{d\tilde{t}}(\tilde{t}) - e^{-\zeta}\frac{dH}{dt}(t)\right\} + \beta\,(t-d)^{\beta-1}\,e^{(\beta-1)\zeta}\left\{H(\tilde{t}) - H(t)\right\}\right]\right]\\
&= e^{-(\beta-1)\zeta}\left[\frac{\partial \tilde{v}_i}{\partial \tilde{x}}\left(\tilde{t},\tilde{x}\right)\frac{\partial \tilde{x}}{\partial x}(t,x) + \frac{\partial \tilde{v}_i}{\partial \tilde{t}}\left(\tilde{t},\tilde{x}\right)\frac{\partial \tilde{t}}{\partial x}(t,x)\right]\\
&= e^{-(\beta-1)\zeta}\frac{\partial \tilde{v}_i}{\partial \tilde{x}}\left(\tilde{t},\tilde{x}\right)e^{\beta\zeta}\\
&= e^{\zeta}\frac{\partial \tilde{v}_i}{\partial \tilde{x}}\left(\tilde{t},\tilde{x}\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial v_i}{\partial t}(t,x) &= e^{-(\beta-1)\zeta}\left[\frac{\partial \tilde{v}_i}{\partial t}(t,x)\right.\\
&\quad\left.-\frac{\partial}{\partial t}\left[(t-d)^{\beta}\,e^{\beta\zeta}\left\{\frac{dH}{d\tilde{t}}(\tilde{t}) - e^{-\zeta}\frac{dH}{dt}(t)\right\} + \beta\,(t-d)^{\beta-1}\,e^{(\beta-1)\zeta}\left\{H(\tilde{t}) - H(t)\right\}\right]\right]\\
&= e^{-(\beta-1)\zeta}\left[\frac{\partial \tilde{v}_i}{\partial \tilde{t}}\left(\tilde{t},\tilde{x}\right)\frac{\partial \tilde{t}}{\partial t}(t,x) + \frac{\partial \tilde{v}_i}{\partial \tilde{x}}\left(\tilde{t},\tilde{x}\right)\frac{\partial \tilde{x}}{\partial t}(t,x)\right.\\
&\qquad -\beta\,(t-d)^{\beta-1}\,e^{\beta\zeta}\left\{\frac{dH}{d\tilde{t}}(\tilde{t}) - e^{-\zeta}\frac{dH}{dt}(t)\right\}\\
&\qquad -(t-d)^{\beta}\,e^{\beta\zeta}\left\{\frac{d^2H}{d\tilde{t}^2}(\tilde{t})\frac{d\tilde{t}}{dt}(t) - e^{-\zeta}\frac{d^2H}{dt^2}(t)\right\}\\
&\qquad -\beta\,(\beta-1)\,(t-d)^{\beta-2}\,e^{(\beta-1)\zeta}\left\{H(\tilde{t}) - H(t)\right\}
\end{aligned}$$

$$
\begin{aligned}
&-\beta\,(t-d)^{\beta-1}\,e^{(\beta-1)\zeta}\left\{\frac{dH}{d\tilde{t}}(\tilde{t})\frac{d\tilde{t}}{dt}(t)-\frac{dH}{dt}(t)\right\}\Bigg]\\
&= e^{-(\beta-2)\zeta}\frac{\partial\tilde{v}_i}{\partial\tilde{t}}\left(\tilde{t},\tilde{x}\right)+\beta\,(t-d)^{\beta-1}\,e^{\zeta}\left\{H(\tilde{t})-H(t)\right\}\frac{\partial\tilde{v}_i}{\partial\tilde{x}}\left(\tilde{t},\tilde{x}\right)\\
&\quad+(t-d)^{\beta}\,e^{\zeta}\left\{e^{\zeta}\frac{dH}{d\tilde{t}}(\tilde{t})-\frac{dH}{dt}(t)\right\}\frac{\partial\tilde{v}_i}{\partial\tilde{x}}\left(\tilde{t},\tilde{x}\right)\\
&\quad-(t-d)^{\beta}\left\{e^{2\zeta}\frac{d^2H}{d\tilde{t}^2}(\tilde{t})-\frac{d^2H}{dt^2}(t)\right\}-2\beta\,(t-d)^{\beta-1}\left\{e^{\zeta}\frac{dH}{d\tilde{t}}(\tilde{t})-\frac{dH}{dt}(t)\right\}\\
&\quad-\beta\,(\beta-1)\,(t-d)^{\beta-2}\left\{H(\tilde{t})-H(t)\right\}.
\end{aligned}
$$

In **xu** space equation [4.1.2] is given by

$$
v_i\frac{\partial v_i}{\partial x}+\frac{\partial v_i}{\partial t}-F=0. \qquad [5.4.10]
$$

If we insert the above expressions for $\partial v_i/\partial x(t,x)$ and $\partial v_i/\partial t(t,x)$ as well as that for $F(t,x)$ given by equation [5.4.9] into [5.4.10], then it turns out that all terms containing H or any of its derivatives cancel leaving,

$$
e^{-(\beta-2)\zeta}\left[\tilde{v}_i\frac{\partial\tilde{v}_i}{\partial\tilde{x}}+\frac{\partial\tilde{v}_i}{\partial\tilde{t}}-\tilde{F}\right]=0.
$$

This of course is equivalent to [4.1.2]. By a similar process we find that [4.1.1] and [4.1.3] also look the same when written in $\tilde{\mathbf{x}}\tilde{\mathbf{u}}$ space. This proves invariance.

In previous sections we discussed how the zero gravity picture could be obtained from the general case using asymptotic solutions. Let us suppose that we require the above transformations to exhibit the same behaviour. This means that given a point $\tilde{\mathbf{x}}$ that is fixed, then we require that $\lim_{\zeta\to\infty}\mathbf{x}=\lim_{\varepsilon\to\infty}g_{\mathbf{x}}(\varepsilon^{-1};\tilde{\mathbf{x}})=0$ (This is equivalent to saying that as $\varepsilon\longrightarrow\infty$, $|g_x(\varepsilon;\mathbf{x})|\longrightarrow\infty$, which illustrates the magnification process spoken of earlier.). Let us consider the x component. Clearly, from equation [5.4.9], this is true provided $\beta>0$. The time component also satisfies this requirement provided $d=0$. This defines our family of curves with which we can asymptotically approach the cusp. The final requirement is that $\lim_{\zeta\to\infty}\tilde{G}=0$ and this is true provided $\gamma<0$. Then, with the equations of motion now written in $\tilde{\mathbf{x}}\tilde{\mathbf{u}}$ space, we obtain the zero gravity case in the limit as $\varepsilon\longrightarrow\infty$ provided we assume that F is zero along the time axis.

So what do we conclude? Well, we have found the most general transformation group and showed that if we restrict β and γ (the similarity degrees for x and G respectively) then the differential equations asymptotically reduce to those of the gravity free case. We recognise this process as being identical to that of §5.3 and so our original procedure is compatible with the more complicated symmetry transformation. The presence of H and its derivatives implies that equations [4.1.1]–[4.1.3] allow a more complicated symmetry transformation than that given by equations [5.3.1]. However, we notice that these additional terms can be thought of as representing an arbitrary set of Galilean transformations. This is because the x coordinate is simply being translated by a factor $t^{\beta}e^{\beta\zeta}\{H(\tilde{t}) - H(t)\}$. It was because of this we decided that it was not worth pursuing the problem any further using the more complicated similarity solution approach. We felt that if any further progress was to be made on an analytical solution, it would be to simply prove existence.

§5.5. Asymptotic behaviour of Newtonian equations of motion in *tr* space.

This section acts as a prologue to chapter seven which discusses the asymptotic limit of the spherically symmetric, General Relativistic equations of motion. We shall illustrate how the Newtonian equations of motion in tr space reduce to that of the gravity free case, written in terms of Cartesian coordinates, in the appropriate limit. As before the asymptotic solutions are similarity solutions and therefore are invariant under a certain symmetry group. The crucial point to realise concerning this analysis, however, is that g, only represents a symmetry group in the limit. It does not form a symmetry group for the general tr Newtonian equations.

One dimensional Newtonian gravity in the context of spherical coordinates means that the equations of motion for three superimposed dusts are:

$$\frac{\partial F}{\partial r} + \frac{2F}{r} = -G\sum_i \rho_i, \qquad [5.5.1]$$

$$v_i\frac{\partial v_i}{\partial r} + \frac{\partial v_i}{\partial t} = F \qquad [5.5.2]$$

and

$$v_i\frac{\partial \rho_i}{\partial r} + \rho_i\frac{\partial v_i}{\partial r} + \frac{2\rho_i v_i}{r} = -\frac{\partial \rho_i}{\partial t}. \qquad [5.5.3]$$

Let us assume that the cusp forms at $r = r_c$. If we make a transformation, $x = r - r_c$, then we have effectively moved our origin from $r = 0$ to $r = r_c$. x then becomes the position of some point as measured from r_c in much the same way as in previous sections. If r_c is constant then [5.5.1]-[5.5.3] become

$$\frac{\partial F}{\partial x} + \frac{2F}{x + r_c} = -G \sum_i \rho_i,$$

$$v_i \frac{\partial v_i}{\partial x} + \frac{\partial v_i}{\partial t} = F,$$

and

$$v_i \frac{\partial \rho_i}{\partial x} + \rho_i \frac{\partial v_i}{\partial x} + \frac{2\rho_i v_i}{x + r_c} = -\frac{\partial \rho_i}{\partial t}.$$

We now transform our equations according to the following group,

$$g(\varepsilon; t, x, F, v_i, \rho_i, G) = (\varepsilon^{k_t} t, \varepsilon^{k_x} x, \varepsilon^{k_F} F, \varepsilon^{k_{v_i}} v_i, \varepsilon^{k_{\rho_i}} \rho_i, \varepsilon^{k_G} G) \qquad [5.3.1]$$

and define a new coordinate system according to $\tilde{\mathbf{x}} = g_{\mathbf{x}}(\varepsilon; \mathbf{x})$. As mentioned above, we do not necessarily require equations [5.5.1]–[5.5.3] to be invariant under [5.3.1]. In fact, the presence of the $1/(x + r_c)$ factor forces these equations to be *not* invariant under these transformations. However, we wish to show that in the limit as $\varepsilon \longrightarrow \infty$, the above differential equations reduce to those that govern the gravity free case written in terms of Cartesian coordinates, (t, x). Following the process described in sections 5.2 and 5.3, we fix $\mathbf{x}$ and suppose that the length and time scales of our $\tilde{t}\tilde{x}$ coordinate system increase asymptotically. To achieve this we insist that $k_x > 0$ and $k_t > 0$. Now under the transformations given in [5.3.1], we have:

$$\frac{\partial F}{\partial x} - \frac{2F}{x + r_c} + G \sum_i \rho_i = \varepsilon^{k_x - k_F} \left\{ \frac{\partial \tilde{F}}{\partial \tilde{x}} - \frac{2\tilde{F}}{\tilde{x} + \varepsilon^{k_x} r_c} + \varepsilon^{k_F - k_x - k_G - k_{\rho_i}} \tilde{G} \sum_i \tilde{\rho}_i \right\},$$

$$v_i \frac{\partial v_i}{\partial x} + \frac{\partial v_i}{\partial t} - F = \varepsilon^{k_x - 2k_{v_i}} \left\{ \tilde{v}_i \frac{\partial \tilde{v}_i}{\partial \tilde{x}} + \varepsilon^{k_t + k_{v_i} - k_x} \frac{\partial \tilde{v}_i}{\partial \tilde{t}} - \varepsilon^{2k_{v_i} - k_F - k_x} \tilde{F} \right\}$$

and

$$v_i \frac{\partial \rho_i}{\partial x} + \rho_i \frac{\partial v_i}{\partial x} + \frac{2\rho_i v_i}{x + r_c} + \frac{\partial \rho_i}{\partial t} = \varepsilon^{k_x - k_{v_i} - k_{\rho_i}} \left\{ \tilde{v}_i \frac{\partial \tilde{\rho}_i}{\partial \tilde{x}} + \tilde{\rho}_i \frac{\partial \tilde{v}_i}{\partial \tilde{x}} + \frac{2\tilde{\rho}_i \tilde{v}_i}{\tilde{x} + \varepsilon^{k_x} r_c} + \varepsilon^{k_t + k_{v_i} - k_x} \frac{\partial \tilde{\rho}_i}{\partial \tilde{t}} \right\}.$$

Hence in the limit of $\varepsilon \longrightarrow \infty$, these equations reduce to the Cartesian differential equations provided

$$k_x + k_G + k_{\rho_i} - k_F = 0,$$

$$k_x - k_{v_i} - k_t = 0,$$

and

$$k_F + k_x - 2k_{v_i} = 0,$$

which are precisely the conditions for the Cartesian differential equations to be invariant under [5.3.1]. If we further stipulate that $\gamma < 0$ and that $F(t, 0) = 0$, then we obtain the zero gravity situation.

This comes expected since r_c is being scaled by a factor, ε^{k_x}, where $k_x > 0$, which increases as we increase ε. Physically, as we change ε and move towards the cusp, we move from different pictures of caustic formation in which the cusp position, $\varepsilon^{k_x} r_c$, increases from picture to picture. Hence, locally, along any radial axis, the spherical shells begin to look like planes. This of course corresponds to the Cartesian picture.

CHAPTER 6. NUMERICAL APPROACH TO CAUSTICS IN GENERAL RELATIVITY.

§6.1. Introduction.

The beginning of this chapter marks a change in tack for this thesis for we now start to look at caustic formation within the context of General Relativity. Now although General Relativity is a very different theory of gravity to that of the Newtonian description, with regards to our problem, there are some similarities. These are mostly associated with the concepts to do with the setting up of the problem rather than the mathematics. Having said that, the construction of the tangent bundle surface in the General Relativistic case is compatible with the Σ-formulation described in §2.2. If we recall, the construction of S as a congruence of integral curves of a vector field, $Z \in T_{(p,X_p)}(TM)$, was made specifically to be metric independent. It was only when we started to put restrictions on the form of Z did we obtain the Newtonian picture of caustic formation. Perhaps, to remind ourselves, we should be more specific. We found that our formulation reduces to the Newtonian case if the vertical part of Z is a function of time. This implies that the projection onto M of the integral curves of Z, i.e. $x^i(t)$, satisfies a Newtonian-like force law. That is to say, $d^2x^i/dt^2 = f^i(t)$. If $f^i(t) = (\nabla V)^i(\mathrm{x}(t))$, then the likeness becomes even more obvious. The General Relativistic picture arises from a different restriction on Z. Let us now suppose that t represents *proper time*. Then, if we make a different assertion as to the form of f^i, we obtain the geodesic equation, i.e., if $f^i(t) = -\Gamma^i_{jk}dx^j/dt(t)dx^k/dt(t)$ where each Γ^i_{jk} is an, as yet, undefined function of t.

Now in the Newtonian case, we described the problem as the solution to the equations conserving mass and momentum. When we consider caustic formation in General Relativity, we study the analogous equations. That is to say, in this chapter we shall analyse

$$G^{ij} = -\kappa T^{ij} = -\kappa \sum_{p=1}^{k} T^{ij}_{(p)} = -\kappa \sum_{p=1}^{k} \mu_{(p)} v^i_{(p)} v^j_{(p)}, \qquad [6.1.1]$$

$$v^i_{(p);j} v^j_{(p)} = 0 \qquad [6.1.2]$$

and

$$T^{ij}_{(p);j} = 0. \qquad [6.1.3]$$

by looking for a numerical solution to these equations. Here, G^{ij} and T^{ij} are the Einstein and energy-momentum tensors; the other quantities, $v^i_{(p)}$ and $\mu_{(p)}$, represent the components of the velocity field and density of dust p. Equation [6.1.1] is simply Einstein's field equation relating the geometry of our spacetime to the mass distribution in it. It illustrates the fact that we are again considering a multi-dust region (cf. equation [4.1.1], the Newtonian equivalent). [6.1.2] is, of course, the geodesic equation and is analogous to the conservation of momentum equation, [4.1.2] in other words. The final equation completes the specification of our problem. It represents the conservation of matter in our system and this should be compared with equation [4.1.3]. The fact that we have a dust index on the energy-momentum tensor is important. It shows that there is no interaction between the dusts and that each separately satisfies the equation of motion (the criteria for superimposed dusts). This equation, therefore, represents a specific input to the problem rather than simply a consequence of the Einstein equation.

As in the Newtonian case, we again have to address the problem of joining conditions. In this sense, the General Relativistic case is equivalent to the Newtonian example considered in chapter four. However, as we are developing a model with the view to solving the equations numerically, we are able to approach this from a different point of view. For the moment, reconsider the analytical case. There we were trying to solve our equations as functions of t and x. We knew that information is lost at the caustic because of the unbounded behaviour in some of our derivatives and density functions and we therefore prescribed information to replace this. In the numerical approach, however, we adopt the Euler method to integrate and solve for the geodesics. Things are now different for no matter how you associate information with these particles, the information is there when you get to the caustic and remains so whilst you pass through it. Thus, it is not a problem of loosing information, rather it is what do you do with the information when you get there.

With regards to the velocity, there is no problem. A particle has a velocity as it arrives at the caustic, and the equations of motion tell it how to move off. This is the situation whether it's particles entering the multi-dust region or if their trajectories are being re-labelled as they become tangent to the caustic. In order to make this statement, however, certain assumptions regarding the metric are required. Specifically, we assume that the metric is continuous (though not, at this stage, necessarily its derivatives) so that the connection does not contain an impulsive part. This is important for it means that the flow vector is continuous and remains so particularly as we parallelly transport it through the caustic. For the density functions we have a slight problem. If a conserved mass is associated with each 'particle' (which could mean either a particle in the normal sense or simply a reference point on a curve) then how this information is transferred depends on the picture we are using to describe mass. We shall therefore postpone this discussion to a later section when we consider representations of the dust continuum.

Before we move on to describe the structure for the rest of this chapter, let us conclude the general description of the problem by choosing our coordinate system. Ideally, we would like to formulate our problem using planar symmetry, as defined in §2.2, so that we have a direct comparison with the Newtonian work. This choice would also simplify matters for we could work with a two dimensional manifold. For the case of General Relativity, however, such a formalism is not necessarily the simplest. Instead, therefore, we choose spherical symmetry to formulate our problem because of its greater familiarity and physical relevance. Now there are many different coordinate systems that reflect spherical symmetry, each with its own particular attributes. Curvature coordinates, however, provide us with the greatest degree of simplification of the field equations for spherical symmetry. In fact we are able to obtain expressions for the metric components as integrals of $G^i{}_j$, the mixed Einstein tensor.

The other reason, and this is important, is that we expect the asymptotics of this system to be similar to that of the gravity free case formulated in Cartesian coordinates (ref. §5.5). This enables us to answer the question, does gravity *really* play no part in cusp formation. We explain; classically, General Relativity is considered to be a far more descriptive theory of gravity than Newtonian Theory. Thus it could be that the asymptotics of a General Relativistic formulation of caustic formation is different to that obtained in §5.5. This would mean that the previous conclusion, that gravity does not play a significant part in caustic formation, is an artifact of the simpler, Newtonian theory. On the other hand, we may obtain the same results. Both conclusions are equally exciting, however, if we obtain the former result then the significance of the work presented in chapter five is decreased. This gives the reader a sneak preview to the work we will present in the next chapter.

We shall continue this discussion on our adopted coordinate system in the next section. The ideas will follow those presented by Synge [S], however, because they represent the building blocks with which we construct our model, it is important that we redevelop them in this thesis. In any case, the aim will be to present the metric coefficients as functions of $G^i{}_j$. This then allows us to readily include matter via $G^i{}_j = -\kappa T^i{}_j$.

In §6.3, Synge's formulation is applied to the case of a spherically symmetric dust. That is to say we shall define an energy-momentum tensor that reflects the multi-dust model described in the Newtonian case, and try to relate our metric components directly to the quantities within this definition that describes the matter. The problem encountered here is that $T^i{}_j$ is in fact a function of g_{ij}. Thus we cannot write the metric components as a function of the mass descriptors alone. To solve this problem we shall introduce an orthonormal tetrad of vectors to act as a new basis for TM. Using this technique, we are able to write g_{ij} as functions of $u^i_{(p)}$ and $\mu_{(p)}$, the components of the fluid flow vectors relative to this tetrad and the densities respectively. This would be a rather nice formulation to invoke in a computer program designed to numerically solve our equations. It would mean discretising our spacetime into a series of 'particles' (a term which can mean many things and will be defined later), each holding $u^i_{(p)}$ and $\mu_{(p)}$ as part of the information defining the dust. In fact, §6.4 describes this process and presents a method of obtaining a solution to the geodesic equation *assuming* we can describe the matter using these variables.

In reality, however, we need to define what we mean by density. In §6.5 this will be discussed, and in doing so we shall see that the above formulation is again corrupted, i.e., g_{ij} once again becomes the solution to a first order, differential equation. We shall introduce the concept of a *conserved mass* between geodesics and formulate a functional definition for the density in terms of these quantities. Having done so, we then proceed to define the representation of our dust continuum that we will adopt in our computer program. In addition, some of the numerical techniques which are not specifically designed to solve the equations, but provide valuable information nonetheless, will be described. Examples of these are: determination of shell crossings, determination of dust number and the treatment of how mass is transferred between dusts at the caustic.

In §6.6, we consider ways to determine the metric coefficients. Now, even though we no longer have the neat formalism presented in §6.3, there is a feature of this problem that enables us to find the metric components, numerically at least. This is that g_{ij} at any point in our spacetime is a function of the integrated mass between the origin and that point. Thus given the metric at $r = 0$, we can construct an Euler scheme that starts at the origin and then proceeds to find an approximation to g_{ij} anywhere on a t = const timeslice in M. Of course the reader might ask why hadn't this procedure been adopted at the outset with the variables $v^i_{(p)}$, thus avoiding the need to introduce an orthonormal tetrad. The only answer to this is that the construction of the computer code was an evolutionary process and that this was realised perhaps later than it should. It possibly also provides us with equations that contain fewer terms. In any case, this is work discussed in §6.6.

Finally, in the last section, we shall present the initial conditions and a summary of the procedures used in the program. The expectation is that the reader will find a lot of the material presented in this chapter to be rather abstract, a consequence of the fact that we are, in essence, trying to describe a computer program that evolved over a number of months. To try and bring this information together into a coherent set of processes, we shall list, and then dry run so to speak, the procedures that constitute a single time step. Whilst doing this, it is hoped that the reader will gain an understanding of the ideas thus previously presented.

At some stage, we shall need to define the initial conditions. We choose to do so at the end. The reason for this is that to define our variables on an initial $t = 0$ time slice, we need to know what variables we are dealing with and of course we don't. Thus we shall try to construct the argument in general terms. This, in actual fact, is virtually impossible, for the program does make some assumptions on the type of caustic we wish to model. Therefore, so as to make the reader aware of these assumptions, we provide a global picture of the caustics we are considering. This can be described quite quickly. We suppose that there exists a point, $(r_c, 0)$ at which the cusp forms. This point is analogous to the axis of symmetry considered in the gravity free case. Moreover, we consider our velocity distribution about $r = r_c$, to look like that in the gravity free case about $x = 0$. Thus, $v^1_{(1)} < 0$, $v^1_{(2)} = 0$ and $v^1_{(3)} > 0$. In addition we suppose that if $r_{(p)} = x^1_{(p)}$ represent geodesics, then $r_{(1)} > r_c$, $r_{(2)} = r_c$ and $r_{(3)} < r_c$.

§6.2. Synge's formulation of spherically symmetric spacetimes.

In this section we shall be primarily concerned with the geometrical aspects of the problem we are considering. That is to say, given the components of the Einstein tensor, can we obtain the metric. Of course, if we can do this, then due to Einstein's equation, we automatically have the metric components as functions of the matter variables. This procedure, however, produces complications and because of this, we choose to defer our discussion on the inclusion of matter until the next section.

Now we said that for simplicity in the components of the Einstein tensor, curvature coordinates will be adopted. This means that the line element for this system becomes

$$ds^2 = e^\alpha dr^2 + r^2 d\sigma^2 - e^\gamma dt^2, \qquad [6.2.1]$$

where $d\sigma^2 = d\theta^2 + \sin^2\theta d\phi^2$. The quantities, $e^{\alpha(t,r)}$ and $e^{\gamma(t,r)}$, are as yet, undefined functions of t and r, however, we assume that they satisfy the following boundary conditions:

$$e^{\alpha(t,0)} = e^{\gamma(t,0)} = 1,$$

or equivalently,

$$\alpha(t,0) = \gamma(t,0) = 0. \qquad [6.2.2]$$

This is the condition for elementary flatness [S]. Its formal definition requires that the ratio of the circumference of a small circle to its radius is 2π. In more meaningful terms, however, this is equivalent to the assumption that spacetime becomes Minkowskian as $r \longrightarrow 0$. It is simply a different, but analogous assertion to that requiring the spacetime to be Minkowskian in the limit as $r \longrightarrow \infty$ for asymptotic flatness.

Using the above form for the metric, we can now calculate the components of the Christoffel symbol, the Riemann tensor, the Ricci tensor and finally the mixed

Einstein tensor. We find that the only remaining non-zero components of $G^i{}_j$ are

$$G^1{}_1 = r^{-2} - r^{-2}e^{-\alpha}\left(1 + r\gamma_1\right), \qquad [6.2.3]$$

$$G^2{}_2 = G^3{}_3 = e^{-\alpha}\left(-\frac{1}{2}\gamma_{11} - \frac{1}{4}\gamma_1^2 - \frac{1}{2}r^{-1}\gamma_1 + \frac{1}{2}r^{-1}\alpha_1 + \frac{1}{4}\alpha_1\gamma_1\right)$$
$$+ e^{-\gamma}\left(\frac{1}{2}\alpha_{44} + \frac{1}{4}\alpha_4^2 - \frac{1}{4}\alpha_4\gamma_4\right),$$

$$G^4{}_4 = r^{-2} - r^{-2}e^{-\alpha}\left(1 - r\alpha_1\right) \qquad [6.2.4]$$

and

$$e^{\alpha}G^1{}_4 = -e^{\gamma}G^4{}_1 = -r^{-1}\alpha_4, \qquad [6.2.5]$$

where the subscripts 1 and 4 represent partial derivatives with respect to r and t respectively. This is quite a remarkable result for we can see that equations, [6.2.3] and [6.2.4], can be integrated to give α and γ as functions of $G^i{}_j$. For e^{α} we have from [6.2.4] that

$$r^2 G^4{}_4 = 1 - e^{-\alpha}\left(1 - r\alpha_1\right)) = 1 - \frac{\partial}{\partial r}\left(re^{-\alpha}\right),$$

which implies that

$$e^{-\alpha} = 1 - \frac{1}{r}\int_{c(t)}^{r} r^2 G^4{}_4\, dr.$$

The function, $c(t)$, is arbitrary, however, if we use one of the conditions for elementary flatness, this can be eliminated. We obtain the result,

$$e^{-\alpha} = 1 - \frac{1}{r}\int_0^r r^2 G^4{}_4\, dr.$$

In a similar manner we can now integrate [6.2.3]. This differential equation can be rewritten as

$$\frac{\partial\gamma}{\partial r} = \frac{e^{\alpha} - 1}{r} - re^{\alpha}G^1{}_1$$

$$\Longrightarrow \gamma = \int_0^r \frac{e^{\alpha} - 1}{r} - re^{\alpha}G^1{}_1\, dr$$

where again, the elementary flatness condition has been used to determine the function of integration. Alternatively, we can obtain a different formulation for $\gamma(t,r)$ simply by subtracting [6.2.4] from [6.2.3]. This gives

$$-re^{-\alpha}\left(\alpha_1 + \gamma_1\right) = r^2\left(G^1{}_1 - G^4{}_4\right)$$

$$\Longrightarrow \alpha_1 + \gamma_1 = -re^{\alpha}\left(G^1{}_1 - G^4{}_4\right)$$

$$\Longrightarrow \gamma = -\alpha - \int_0^r re^{\alpha}\left(G^1{}_1 - G^4{}_4\right)\, dr.$$

The above ideas can be summarised by saying that we have now found $\alpha(t,r)$ and $\gamma(t,r)$ as functions of $G^4{}_4$ and $G^1{}_1$. This, given $G^i{}_j$, completely determines the metric. It can be shown [S], using [6.2.5] and the identity, $\nabla_i G^i{}_j = 0$, that the remaining unconsidered, non-zero components of G, i.e., $G^1{}_4$, $G^4{}_1$, $G^2{}_2$ and $G^3{}_3$, can be written in terms of $G^4{}_4$, $G^1{}_1$ and their derivatives. This means that we have two fundamental functions, either α and γ or $G^1{}_1$ and $G^4{}_4$, that describe the geometry of our system. If we do have some matter with energy-momentum tensor, T, then only two of its components need be used to determine the metric components. Moreover, no matter what form our energy-momentum tensor takes, the limits between which we integrate imply that at any point the metric components are determined by the mass that the shell passing through that point encloses. Although we promised not to include matter within this section, we shall finish by presenting these results in terms of $T^i{}_j$:

$$e^{-\alpha} = 1 + \frac{\kappa}{r}\int_0^r r^2 T^4{}_4\, dr \qquad [6.2.6]$$

and

$$\gamma = -\alpha + \kappa\int_0^r re^{\alpha}\left(T^1{}_1 - T^4{}_4\right)\, dr \qquad [6.2.7]$$

or

$$\gamma = \int_0^r \frac{e^{\alpha} - 1}{r} + \kappa re^{\alpha} T^1{}_1\, dr. \qquad [6.2.8]$$

§6.3. Synge's formulation applied to spherically symmetric dusts.

Equations [6.2.6]–[6.2.8] tell us how to relate the metric components to an arbitrary energy-momentum tensor. In this section we shall define $T^i{}_j$ and therefore the kind of matter we wish our spacetime to have. We should stress that the equations that constitute the model of caustic formation, which we are setting up, will not be presented here in this section. Instead, we merely plan to define our matter in such a way so as to be compatible with our model. In other words, the aim of this section is to simply take equations [6.2.6] and [6.2.8] and develop them further by specifying the right hand side in terms of the mass or matter variables that describe a dust.

Now, in the Newtonian discussion we modelled caustic formation by considering converging dust particles. In the General Relativistic case, we do the same thing. With the notion of a dust given in chapter one, we are able to immediately write down the energy-momentum tensor for this kind of matter to the *future* of the caustic set. That is to say, since a dust (collision-less fluid) implies zero pressure, we have

$$T^i{}_j = \sum_{p=1}^{k} \mu_{(p)} v^i_{(p)} v^k_{(p)} g_{kj}. \qquad [6.3.1]$$

This definition involves two new functions, $\mu_{(p)}$ and $v_{(p)}$, which represent the proper density and 4-velocity of a dust indexed by p. For the moment, we shall assume that the density is simply a given function. The velocity, however, is defined by $v^i_{(p)} = dx^i_{(p)}/d\tau$, where $x^i_{(p)}$ represents the components of a geodesic along which particles from dust p might travel and τ is the proper time. For convenience we shall normalise the velocity so that $v^i_{(p)} v_{(p)i} = -1$.

If we picture our spacetime as a single entity, then regardless of the number of dusts, we again consider our matter continuum to be constructed as a series of shells (Ref. §2.1.). Since we are dealing with spherical symmetry, if we project these onto the $r\theta\phi$ plane we obtain a set of concentric hollow spheres. It also means that fluctuations in the density and velocity are in the t or r direction only, so our 4-velocity becomes of the form,

$$v_{(p)} = \left(v^1_{(p)}, 0, 0, v^4_{(p)}\right).$$

If we use the condition that $v^i_{(p)} v_{(p)j} = -1$, then this becomes

$$v_{(p)} = \left(v^1_{(p)}, 0, 0, e^{-\gamma/2}\sqrt{1 - v^1_{(p)} v^1_{(p)} e^{\alpha}}\right),$$

which is useful since we now only need two variables, $\mu_{(p)}$ and $v^1_{(p)}$, to specify our matter.

Clearly, equation [6.3.1] represents the energy-momentum tensor for a multi-dust region in our spacetime. If we choose k equal to 3 then within the context of a spherically symmetric dust exhibiting shell crossing singularities, $\pi(S)$ becomes a simple fold catastrophe (ref. §4.1). Here, $\pi: TM \longrightarrow M$ represents the projection map and $S \subset TM$, the tangent bundle surface generated by $x^i_{(p)}$. For the region not enclosed by the caustic we have $k = 1$ and we can drop the dust index. The simplicity of [6.3.1] hides the fact that when inserted into [6.2.6]–[6.2.8], the metric components become complicated functions of the 6 mass variables: $\mu_{(1)}$, $v^1_{(1)}$, $\mu_{(2)}$, $v^1_{(2)}$, $\mu_{(3)}$ and $v^1_{(3)}$. This means that g_{ij} becomes a function of 3 velocity vectors and is therefore determined by quantities defined on 3 disjoint regions of S. Since the acceleration of any particle is a function of the metric components, this corroborates the description of the gravitational interaction given at the end of §2.1.

With the above form for $T^i{}_j$, however, we encounter the first problem with Synge's formulation—by inserting [6.3.1] into [6.2.6], we obtain an integral equation for $e^{-\alpha}$ that involves e^{α} as part of the integrand. This means that we no longer have a nice functional formulation for the metric components. To surmount this problem we introduce a tetrad, e_a, of orthonormal vectors to act as a new basis for TM at any point in the manifold. Using this formulation, a vector with components, v^i, relative to the coordinate basis, ∂_i, will have components, u^a, with respect to the basis, e_a, such that $u^a = e^a{}_i v^i$. Here $a = 1, \ldots, 4\,(\dim M)$ and the matrix, $(e^a{}_i)$, is invertible with $(e^a{}_i)^{-1} = \left(e_a{}^i\right)$. Now, for curvature coordinates the vectors, ∂_i, are in fact orthogonal; a direct consequence of having a diagonal metric tensor. This means that we can choose the e_a to be ∂_i but rescaled so that their inner product, with respect to the Minkowski metric, is unity. Thus we have

$$e_1 = e_1{}^1\partial/\partial x^1, \quad e_2 = e_2{}^2\partial/\partial x^2, \quad e_3 = e_3{}^3\partial/\partial x^3, \quad e_4 = e_4{}^4\partial/\partial x^4, \qquad [6.3.2]$$

implying that $(e_a{}^i)$ is also diagonal. Using the condition, $g(e_a, e_a) = 1$, the above scaling parameters, $e_i{}^i$ (no summation implied), can now be identified. We have

$$\begin{aligned} g(e_a, e_a) &= g\left(e_a{}^i\frac{\partial}{\partial x^i}, e_a{}^j\frac{\partial}{\partial x^j}\right) \\ &= e_a{}^i e_a{}^j g_{ij} \end{aligned}$$

and so obtain the result:

$$\begin{aligned} e_1 &= \left(e^{-\alpha/2}, 0, 0, 0\right), \\ e_2 &= \left(0, \frac{1}{r}, 0, 0\right), \\ e_3 &= \left(0, 0, \frac{1}{r\sin\theta}, 0\right), \\ e_4 &= \left(0, 0, 0, e^{-\gamma/2}\right). \end{aligned}$$

We can now determine the relationship between the components of v and u. We have that $u^a = e^a{}_i v^i$ and therefore,

$$u = \left(v^1 e^{\alpha/2}, 0, 0, v^4 e^{\gamma/2}\right).$$

We also have $e_a{}^j u^a = e_a{}^j e^a{}_i v^i = \delta^j_i v^i = v^j$ and thus,

$$v = \left(u^1 e^{-\alpha/2}, 0, 0, u^4 e^{-\gamma/2}\right).$$

The upshot of all this is that we now have a new function describing our matter, i.e., $u_{(p)}$. We shall see that by rewriting the energy-momentum tensor using this tetrad formalism, the critical components of $T^i{}_j$ become independent of g_{ij}. Now, from [6.3.1] we have

$$T^i{}_j = \sum_{p=1}^{3} \mu_{(p)} u^a_{(p)} e_a{}^i u^b_{(p)} e_b{}^k g_{kj},$$

and in particular,

$$\begin{aligned} T^1{}_1 &= \sum_{p=1}^{3} \mu_{(p)} u^a_{(p)} e_a{}^1 u^b_{(p)} e_b{}^k g_{k1} \\ &= \sum_{p=1}^{3} \mu_{(p)} u^1_{(p)} e^{-\alpha/2} u^1_{(p)} e^{-\alpha/2} e^{\alpha} \\ &= \sum_{p=1}^{3} \mu_{(p)} u^1_{(p)} u^1_{(p)} \end{aligned} \qquad [6.3.3]$$

and

$$\begin{aligned} T^4{}_4 &= \sum_{p=1}^{3} \mu_{(p)} u^4_{(p)} e^{-\gamma/2} u^4_{(p)} e^{-\gamma/2} \left(-e^{\gamma}\right) \\ &= \sum_{p=1}^{3} -\mu_{(p)} u^4_{(p)} u^4_{(p)}. \end{aligned} \qquad [6.3.4]$$

Although not relevant to the current discussion, we shall include the $T^1{}_4$ component,

$$\begin{aligned} T^1{}_4 &= \sum_{p=1}^{3} \mu_{(p)} u^1_{(p)} e^{-\alpha/2} u^4_{(p)} e^{-\gamma/2} \left(-e^{\gamma}\right) \\ &= \sum_{p=1}^{3} -\mu_{(p)} u^1_{(p)} u^4_{(p)} e^{(\gamma-\alpha)/2}. \end{aligned} \qquad [6.3.5]$$

This will be required in the next section.

Finally, if we insert [6.3.3] and [6.3.4] into the equations that were developed in the previous chapter giving the metric coefficients as functions of $T^i{}_j$ ([6.2.6] and [6.2.7]), we obtain

$$e^{-\alpha} = 1 + \frac{\kappa}{r}\int_0^r r^2 \sum_{p=1}^{3} \mu_{(p)} u^4_{(p)} u^4_{(p)}\, dr \qquad [6.3.6]$$

and

$$\gamma = -\alpha + \kappa \int_0^r r e^{\alpha} \sum_{p=1}^{3} \mu_{(p)} \left(u^1_{(p)} u^1_{(p)} + u^4_{(p)} u^4_{(p)} \right) dr. \qquad [6.3.7]$$

Of course, if we use the normalisation condition imposed on our velocity vectors, we find that

$$v^i_{(p)} v_{(p)i} = v^i_{(p)} v^j_{(p)} g_{ij} = e_a{}^i u^a_{(p)} e_b{}^j u^b_{(p)} g_{ij} = u^a_{(p)} u^b_{(p)} \eta_{ab} = u^1_{(p)} u^1_{(p)} - u^4_{(p)} u^4_{(p)} = -1$$

implying

$$u^4_{(p)} = \sqrt{1 + u^1_{(p)} u^1_{(p)}},$$

so that the left hand side of the above equations become functions of only six mass variables as expected.

Thus the problem with Synge's formulation applied to spherically symmetric dusts is solved, i.e., one can find α and γ at any point in the spacetime given $\mu_{(p)}(t,r)$ and $u^1_{(p)}(t,r)$. At first sight this seems like an extremely useful result, however, as we shall show in §6.5, when we consider writing the proper density in terms of a conserved mass, metric components reappear inside the integrals for α and γ. There is no cure for this new complication and to get around this we must choose an appropriate integration scheme.

§6.4. Numerical evolution of a spherically symmetric dust.

The earlier part of this chapter presented a method of obtaining the metric components, i.e. e^{α} and e^{γ}, as a function of the mass descriptors, $u^1_{(p)}$ and $\mu_{(p)}$, for a spherically symmetric dust. Whilst we concentrated on obtaining these expressions we did not discuss in any detail the model we are trying to set up. In other words, we defined our matter in a way immediately applicable to modelling caustic formation, but neglected to describe how this matter evolved. In this section, therefore, we present the model as a set of differential equations that determine how we must evolve our dusts. Since we are interested in determining the curves that our dust particles follow, it is clear that the geodesic equation will be the key equation and our problem really boils down to solving this. We shall opt for a numerical solution, primarily because of the highly non-linear differential equations that we expect to obtain. This, of course, is a characteristic of formulating physical problems within the context of General Relativity.

Now, the equations we wish to solve are [6.1.1]–[6.1.3]. For convenience, we shall restate them here:

$$G^{ij} = -\kappa T^{ij} = -\kappa \sum_{p=1}^{k} T^{ij}_{(p)} = -\kappa \sum_{p=1}^{k} \mu_{(p)} v^i_{(p)} v^j_{(p)}, \qquad [6.1.1]$$

$$v^i_{(p);j} v^j_{(p)} = 0 \qquad [6.1.2]$$

and

$$T^{ij}_{(p);j} = 0. \qquad [6.1.3]$$

The first equation, [6.1.1], defines the metric in terms of the mass parameters. This was discussed in great detail in §6.2 and §6.3. [6.1.2] is, of course, the geodesic equation and as mentioned above, this tells us how we must evolve our dusts. It is really this equation that we need to solve; the other two can be thought of as there simply to supply us with information so that all terms in the geodesic equation are known. The final equation is the General Relativistic equivalent of the conservation of matter equation, [4.1.3]. This provides us with a single piece of information which allows us to determine the density as a function of the metric components, each dust's 4-velocity and a *conserved mass* (more of this later).

We shall start by explaining the techniques used to numerically solve the geodesic equation. Of course the density function should be defined since clearly it will appear within the geodesic equation in one form or another. In order to simplify the discussion, however, we shall delay specifying the functional form of $\mu_{(p)}$, but assume that it is known. This enables us to quickly present the ideas for solving the geodesic equation. The reason why we adopt this segmented approach is because, in actual fact, it is the specification of $\mu_{(p)}$ and the rules by which we must treat this function on the caustic that provides the complexity of the problem. In other words, we could say that the real heart of the problem is understanding how the crossing of shells and therefore the transference of matter between dusts effect the geodesics. These phenomena alter our equations exclusively through the density function.

Now, the Euler method is perhaps the simplest that one can use to solve differential equations and therefore, as a numerical technique, it presents an adequate starting point to model the evolution of a relativistic dust. By way of revision, suppose we have the differential equation,

$$\frac{dr}{dt} = f(t, r), \qquad [6.4.1]$$

for some f. Then, if we know $r(t)$ at $t = t_n$ say, we can write down an approximation to the solution for $r(t)$ at $t = t_n + \delta t$ as follows. If $r_{(n)}$ represents $r(t_n)$ and $t_{n+1} = t_n + \delta t$, we have, from [6.4.1],

$$\frac{r_{(n+1)} - r_{(n)}}{t_{n+1} - t_n} \approx f\left(t_n, r_{(n)}\right)$$

$$\Longrightarrow r_{(n+1)} \approx r_{(n)} + (t_{n+1} - t_n) f\left(t_n, r_{(n)}\right).$$

In the case we wish to consider r represents a geodesic, which we now denote by $r_{(p)}$, along which a particle from the pth dust might travel. The function, f, must then represent the velocity with respect to our coordinate basis, i.e., $f(t, r_{(p)}) = dr_{(p)}/dt(t)$. Thus we have, for a geodesic of dust p,

$$\begin{aligned} r_{(p,n+1)} &\approx r_{(p,n)} + \left(t_{n+1} - t_n\right) \frac{dr_{(p)}}{dt}(t_n) \\ &= r_{(p,n)} + \left(t_{n+1} - t_n\right) \frac{v^1_{(p,n)}}{v^4_{(p,n)}} \\ &= r_{(p,n)} + \left(t_{n+1} - t_n\right) \frac{u^1_{(p,n)}}{u^4_{(p,n)}} e^{\left(\gamma_{(n)} - \alpha_{(n)}\right)/2} \\ &= r_{(p,n)} + \left(t_{n+1} - t_n\right) \frac{u^1_{(p,n)}}{\sqrt{1 + u^1_{(p,n)} u^1_{(p,n)}}} e^{\left(\gamma_{(n)} - \alpha_{(n)}\right)/2}, \end{aligned} \qquad [6.4.2]$$

where we have used the requirement that the $v_{(p)}$ be normalised. It can also be seen that we have developed the notation further. From now on we shall discriminate between coordinate and non-coordinate indices by grouping the latter in brackets.

Moreover, for clarity in the argument, only those indices relevant to the discussion will be grouped, but we will try to explain this at the appropriate juncture.

The numerical scheme is not complete for we have yet to find a way of estimating the $(t_{n+1}-t_n)$ coefficient in equation [6.4.2]. For this the geodesic equation is required, however, before we consider this, it might be worth listing all the procedures that are required in a single time step of a computer simulation. Although at this stage, it may seem rather abstract, the hope is to provide the reader with the chance to formulate a global picture of what is going on. Thus, the logical sequence of events for a *single* geodesic are as follows:

1. Given all variables evaluated at $t = t_n$, i.e., $r_{(p,n)}$, $u^1_{(p,n)}$ and $M_{(p)}$ (which represents a conserved mass for this geodesic and will be defined in the next section), calculate $e^{\alpha(n)}$ at $r = r_{(p,n)}$ (process not yet defined).

2. Using information from step 1, calculate $e^{\gamma(n)}$ at $r = r_{(p,n)}$ (process not yet defined).

3. Using information from step 1, calculate $\mu_{(p,n)}$ at $r = r_{(p,n)}$ (process not yet defined).

4. Using information from steps 1, 2 and 3, calculate $u^1_{(p,n+1)}$ (solution of geodesic equation and process not defined).

5. Using information from steps 1 and 2, calculate $r_{(p,n+1)}$ (equation [6.4.2]).

We can use this list of procedures as an aide-mémoire in the rest of this chapter to ensure that the program and all required analytical calculations are described. Once this is done, we shall revisit this list and fill in the blanks.

The first blank that we shall fill in is a description of how to obtain an approximation for $u^1_{(p,n+1)}$, given every other function evaluated at $t = t_n$. To do this, we again use the Euler technique with the result that

$$u^1_{(p,n+1)} \approx u^1_{(p,n)} + (t_{n+1} - t_n)\frac{du^1_{(p)}}{dt}(t_n). \qquad [6.4.3]$$

The $(t_{n+1} - t_n)$ coefficient is essentially the time derivative of the velocity. This information can be supplied by the geodesic equation provided we can write it in a form that we can use. In other words, we require equation [6.1.2] in terms of the tetrad velocity, $u^a_{(p)}$. Until now we have been formulating our ideas in terms of discrete time steps, if we return to the continuous picture then we can drop the subscript n and simplify the notation. Thus to obtain $du^1_{(p)}/dt$ we have,

$$\nabla_{v_{(p)}} v_{(p)} = 0 \qquad [6.1.2]$$

$$\Longrightarrow u^a_{(p)} \nabla_{e_a} \left(u^b_{(p)} e_b \right) = 0$$

$$\Longrightarrow u^a_{(p)} e_a \left(u^b_{(p)} \right) e_b + u^a_{(p)} u^b_{(p)} \Gamma^c{}_{ab} e_c = 0$$

$$\Longrightarrow u^a_{(p)} e_a \left(u^c_{(p)} \right) + u^a_{(p)} u^b_{(p)} \Gamma^c{}_{ab} = 0,$$

where $\Gamma^c{}_{ab}$ are the connection coefficients with respect to the basis, e_a, defined by $\nabla_{e_a} e_b = \Gamma^c{}_{ab} e_c$ (To avoid confusion, the $\Gamma^c{}_{ab}$ will never be evaluated so that components, $\Gamma^1{}_{11}$ for example, will always refer to the connection with respect to the coordinate basis.). Note also that $e_a \left(u^c_{(p)} \right)$ is a function defined on M. To proceed, we need to determine the relationship between $\Gamma^k{}_{ij}$ and $\Gamma^c{}_{ab}$. Now,

$$\begin{aligned} \nabla_{e_a} e_b &= e_a{}^i \nabla_i \left(e_b{}^j \partial_j \right) \\ &= e_a{}^i \partial_i \left(e_b{}^j \right) \partial_j + e_a{}^i e_b{}^j \Gamma^k_{ij} \partial_k \end{aligned}$$

and since $\partial_i = e^a{}_i e_a$, we therefore have

$$\Gamma^c_{ab} = e_a{}^i \partial_i \left(e_b{}^j \right) e^c{}_j + e_a{}^i e_b{}^j \Gamma^k_{ij} e^c{}_k.$$

Inserting this result into our geodesic equation gives

$$u^a_{(p)} e_a{}^i \partial_i \left(u^c_{(p)} \right) + u^a_{(p)} u^b_{(p)} \left(e_a{}^i \partial_i \left(e_b{}^j \right) e^c{}_j + e_a{}^i e_b{}^j \Gamma^k_{ij} e^c{}_k \right) = 0,$$

then since $v^i_{(p)} = dx^i_{(p)}/d\tau$, we have $u^a_{(p)} e_a{}^i \partial_i = d/d\tau$ resulting in,

$$\frac{du^c_{(p)}}{d\tau} + u^b_{(p)} e^c{}_k \frac{de_b{}^k}{d\tau} + u^a_{(p)} u^b_{(p)} e^c{}_k e_a{}^i \Gamma^k{}_{ij} e_b{}^j = 0.$$

By the normalisation condition, $u^a_{(p)}u_{(p)a} = -1$, we automatically have $u^4_{(p)}$ as a function of $u^1_{(p)}$ and so therefore we only need to solve the r component of this equation. Thus for $c = 1$,

$$\frac{du^1_{(p)}}{d\tau} - \frac{u^1_{(p)}}{2}\frac{d\alpha}{d\tau} + u^1_{(p)}u^1_{(p)}e^{-\alpha/2}\Gamma^1{}_{11} + 2u^1_{(p)}u^4_{(p)}e^{-\gamma/2}\Gamma^1{}_{14} + u^4_{(p)}u^4_{(p)}e^{\alpha/2-\gamma}\Gamma^1{}_{44} = 0, \tag{6.4.4}$$

meaning that in order for us to write our geodesic equation in terms of known quantities, we need to determine the connection coefficients. This, in fact, is easy because the metric in curvature coordinates is a diagonal matrix. Thus by simply inserting the form for g_{ij} that we have into the definition of $\Gamma^i{}_{jk}$, we obtain the following results:

$$\Gamma^1{}_{11} = \frac{1}{2}\alpha_1, \quad \Gamma^1{}_{14} = \frac{1}{2}\alpha_4 \quad \text{and} \quad \Gamma^1{}_{44} = \frac{1}{2}e^{\gamma-\alpha}\gamma_1.$$

Equation [6.4.4] then becomes

$$\begin{aligned}\frac{du^1_{(p)}}{d\tau} = {} & \frac{1}{2}u^1_{(p)}u^1_{(p)}e^{-\alpha/2}\alpha_1 + \frac{1}{2}u^1_{(p)}u^4_{(p)}e^{-\gamma/2}\alpha_4 - \frac{1}{2}u^1_{(p)}u^1_{(p)}e^{-\alpha/2}\alpha_1 - u^1_{(p)}u^4_{(p)}e^{-\gamma/2}\alpha_4 \\ & - \frac{1}{2}u^4_{(p)}u^4_{(p)}e^{-\alpha/2}\gamma_1\end{aligned}$$

$$\Longrightarrow \frac{du^1_{(p)}}{d\tau} = -\frac{1}{2}u^1_{(p)}u^4_{(p)}e^{-\gamma/2}\alpha_4 - \frac{1}{2}u^4_{(p)}u^4_{(p)}e^{-\alpha/2}\gamma_1.$$

This equation is now beginning to look like the form that we require. The left hand side is giving us the derivative of $u^1_{(p)}$, albeit with respect to τ (although this is not a problem), whereas the right hand side is a function of the mass descriptors, the metric components and derivatives of the metric components. In fact the troublesome metric derivatives can be removed using the $G^1{}_1$ and $G^1{}_4$ components of the Einstein tensor (equations [6.2.3] and [6.2.5]) in conjunction with the energy-momentum tensor (equations [6.3.3] and [6.3.5]). In other words, since

$$\alpha_4 = -re^{(\gamma+\alpha)/2}\kappa\sum_{p=1}^{3}\mu_{(p)}u^1_{(p)}u^4_{(p)}$$

and

$$\gamma_1 = \frac{e^\alpha}{r} - \frac{1}{r} + re^\alpha\kappa\sum_{p=1}^{3}\mu_{(p)}u^1_{(p)}u^1_{(p)},$$

our geodesic equation becomes

$$\frac{du^1_{(p)}}{d\tau} = \frac{1}{2}u^1_{(p)}u^4_{(p)}re^{\alpha/2}\kappa\sum_{q=1}^{3}\mu_{(q)}u^1_{(q)}u^4_{(q)} - \frac{1}{2r}u^4_{(p)}u^4_{(p)}e^{\alpha/2}\left(1-e^{-\alpha}\right)$$
$$- \frac{1}{2}u^4_{(p)}u^4_{(p)}re^{\alpha/2}\kappa\sum_{q=1}^{3}\mu_{(q)}u^1_{(q)}u^1_{(p)}.$$

We can remove the derivative with respect to τ and replace it with one with respect to t simply by dividing by $dt/d\tau = v^4_{(p)} = u^4_{(p)}e^{-\gamma/2}$. In addition, the summation signs can be removed if we allow (p, q, r) to represent some permutation of $(1, 2, 3)$. Thus finally,

$$\frac{du^1_{(p)}}{dt} = -\frac{1}{2}e^{(\alpha+\gamma)/2}\left(\frac{u^4_{(p)}}{r}\left(1-e^{-\alpha}\right) - \kappa r\mu_{(q)}u^1_{(q)}\left(u^4_{(q)}u^1_{(p)} - u^4_{(p)}u^1_{(q)}\right)\right.$$
$$\left. - \kappa r\mu_{(r)}u^1_{(r)}\left(u^4_{(r)}u^1_{(p)} - u^4_{(p)}u^1_{(r)}\right)\right). \quad [6.4.5]$$

To summarise, the above analysis results in an equation that relates the time derivative of $u^1_{(p)}$ to known quantities. If we wish, we could evaluate both sides of this equation at $t = t_n$ so that the result is immediately applicable in the numerical scheme summarised by equation [6.4.3]. In essence then, given $r_{(p,n)}$, $u^1_{(p,n)}$, $\mu_{(p,n)}$, $e^{\alpha(n)}$ and $e^{\gamma(n)}$, we have developed numerical techniques with which we can calculate $r_{(p,n+1)}$ and $u^1_{(p,n+1)}$. This completes parts 4 and 5 of the list of procedures that need to be implemented in a computer program.

Before we move on, we should say that equation [6.4.5] is quite important in its own right. It is analogous to the force equation presented in §4.1 (and indeed reduces to it in the low velocity limit as the next chapter will show), but is clearly far more complicated. It is quite believable, from a superficial glance, that there may indeed be certain additional terms that are determining the physics of cusp formation. That is to say, if we could perform an asymptotic analysis in much the same way as we did for the Newtonian equations of motion, we might find extra, non-linear terms that remain at the end of this process. On a less profound level, we can see from equation [6.4.5] that the terms containing $\mu_{(p)}$ cancel. This means that as $r_{(p)}$ approaches the caustic it is unaffected by its own local density except via its own integrated coulomb field. That is to say, the only effect that $\mu_{(p)}$ has on a geodesic, $r_{(p)}$ say, is due to the integrated mass between 0 and $r_{(p)}$ distorting the spacetime metric.

The last thing to point out regarding this equation, is that as noticed by Clarke and O'Donnell [CO], it is possible for the derivative to remain bounded even on the caustic. To explain, suppose that p, q and r equal 1, 2 and 3 respectively. Then as $r_{(1)}$ approaches the caustic, the term on the right hand side of equation [6.4.5] containing $\mu_{(2)}$ has the potential to become infinite. Let us consider this term. We have

$$-\kappa r \mu_{(2)} u^1_{(2)} \left(u^4_{(2)} u^1_{(1)} - u^4_{(1)} u^1_{(2)} \right),$$

or equivalently,

$$-\kappa r \mu_{(2)} u^1_{(2)} \left(u^1_{(1)} \sqrt{1 + u^1_{(2)} u^1_{(2)}} - u^1_{(2)} \sqrt{1 + u^1_{(1)} u^1_{(1)}} \right).$$

Now, there eventually becomes a point (on the caustic) where $r_{(1)}$ and $r_{(2)}$ coexist. Thus, although $\mu_{(1)}$ and $\mu_{(2)}$ both tend towards infinity, $u^1_{(1)} - u^1_{(2)}$ tends to zero and it becomes possible for the above product to be finite on the caustic.

§6.5. Approximations to the dust continuum and determination of the density function.

We have now come to a point in our discussion where we cannot proceed any further without discussing the density function. The reason for this is that in doing so, we define a quantity known as the *conserved mass* which crops up in the processes to determine the metric coefficients; the last two blanks in our list of procedures. In this sense, therefore, $\mu_{(p)}$ is more fundamental and so we discuss this next.

This section is likely to be of some length for not only do we need to define the density and how it relates to the concept of conserved mass, we must define the representation for the dust continuum that we wish to use in our computer model. This, of course, leads us into the discussion on how to manipulate our information regarding mass when our geodesics reach the caustic; joining conditions in other words. Here we will have to talk specifically about certain techniques used in our computer program to track particles as they move and cross, as well as keeping a record of which dust they are a member. We therefore leave this to the end of this section.

We begin by supposing that $M_{(p)}$ represents the contribution to the total mass in a region, U, of our spacetime, made by particles that are members of dust p. Then, as required, if U represents a $t = \text{const}$ time slice orthogonal to $v_{(p)}$, we can construct a relationship between these quantities and each $\mu_{(p)}$ (Note that the evolution equations discussed in the previous chapter essentially approximate our spacetime by a series of $t = \text{const}$ hypersurfaces.). To do this we shall firstly define the idea of a conserved 3-form on M that represents the mass flux of a *single dust* having a *unique flow vector*, v. This allows us to define as an integral, the relationship between M_s, the total mass in U due to our single dust, and μ, its density. Having done this we then modify this definition so that it becomes applicable to the case that we have, namely a multi-dust region in M. The result will ultimately constitute the functional definition for the densities that we have been neglecting to specify.

Now to determine this 3-form we note that for a single dust,

$$T^{ij}{}_{;j} = \mu v^j v^i_{;j} + \left(\mu v^j\right)_{;j} v^i,$$

which by equation [6.1.2] and [6.1.3] implies that,

$$\left(\mu v^j\right)_{;j} = 0.$$

We now write the left hand side of the above equation in terms of the co-vector, $v_\flat$. Thus,

$$\begin{aligned}
\left(\mu g^{jk} v_k\right)_{;j} &= \frac{1}{\sqrt{-g}} \partial_j \left(\sqrt{-g}\mu g^{jk} v_k\right) \\
&= - * \left(\frac{1}{g} \partial_j \left(\sqrt{-g}\mu g^{jk} v_k\right) dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4\right) \\
&= - * \left(\frac{1}{3!} \partial_j \left(\sqrt{-g}\mu g^{lk} v_k\right) \varepsilon_{lmnp} dx^j \wedge dx^m \wedge dx^n \wedge dx^p\right) \\
&= - * d \left(\frac{\sqrt{-g}}{3!} \mu g^{lk} v_k \varepsilon_{lmnp} dx^m \wedge dx^n \wedge dx^p\right) \\
&= - * d * \left(\mu v_k dx^k\right) \\
&= - * d * \left(\mu v_\flat\right),
\end{aligned}$$

so that

$$d * \left(\mu v_\flat\right) = 0,$$

meaning that $* \left(\mu v_\flat\right)$ is a conserved 3-form [N]. It is this that represents the mass flux 3-form for a single dust. To see this we now show that $* \left(\mu v_\flat\right)$ can also be written as

$\mu i_v \alpha_M$, where α_M represents the standard volume form on M. The quantity $i_v \alpha_M$ is said to be the restriction of α_M to a hypersurface orthogonal to v, i.e. U, and thus the integral of $\mu i_v \alpha_M$ is the total mass in this region. This calculation also consolidates the two definitions for $M_{(k)}(U)$ given by equations [4.5.1] and [4.5.2]. So,

$$\begin{aligned}
*\left(\mu v_\flat\right) &= *\left(\mu g_{ik} v^i dx^k\right) \\
&= \frac{\sqrt{-g}}{3!} \mu g_{ik} v^i \varepsilon^k{}_{pqr} dx^p \wedge dx^q \wedge dx^r \\
&= \frac{\sqrt{-g}}{3!} \mu v^i \varepsilon_{ipqr} dx^p \wedge dx^q \wedge dx^r \\
&= \mu i_v \alpha_M \\
&= \mu \alpha_s,
\end{aligned}$$

as required.

We can summarise the first part of this section by saying that the conservation equation for a single dust (namely [6.1.3]) implies the existence of a conserved 3-form. The integral of this 3-form over a surface in M, orthogonal to v, is given by

$$M_s = \int_U *\left(\mu v_\flat\right) = \int_U \mu i_v \alpha_M = \int_U \mu \alpha_s,$$

and represents the mass in U due to a single dust with unique flow vector. Finally we note that if U always corresponds to a $t =$ const surface, then M_s is a constant, a result of the fact that $*\left(\mu v_\flat\right)$ is a conserved 3-form. We shall use this feature to remove the integral signs, thus obtaining the density function.

The next question is how do we modify this definition to account for the fact we wish to consider a multi-dust region in our spacetime. For this, it is worth bearing in mind that this situation is equivalent to considering a region that contains a single dust with a non-unique velocity vector. To solve this problem we recall §4.5, which discussed how to define the notion of density in a Newtonian formulation of the problem. Because of the fact that π was not orientation preserving when restricted to $S \subset TM$, we needed to use the idea of a of pseudo-1-form on M to define $\rho_{(p)}$ so that it transformed between TM and M in the correct manner. For the General Relativistic case we have a similar scenario and we therefore define $\mu_{(p)}$ to reflect this. Thus, as in [4.5.2], we define the total mass in $U \subset M$ due to dust p to be the integral of the proper density over a $t = \text{const}$ hypersurface, orthogonal to $v_{(p)}$. This means that

$$M_{(p)} = \int_U \mu_{(p)}\alpha_{(p)}, \qquad [6.5.1]$$

where $\alpha_{(p)}$ is the volume form on this surface such that $\alpha_{(p)} = (-1)^{p+1} i_{v_{(p)}}\alpha_M$. As in the Newtonian case, the factor of $(-1)^{p+1}$ is important for it takes into account the fact that π is not orientation preserving (ref. §4.5) and insures that the $\alpha_{(p)}$ all have the same orientation with respect to α_M.

We now proceed to remove the integral signs in equation [6.5.1]. The previous calculation showed that for a general hypersurface orthogonal to $v_{(p)}$,

$$i_{v_{(p)}}\alpha_M = \frac{\sqrt{-g}}{3!}\varepsilon_{ijkl}v^i_{(p)}dx^j \wedge dx^k \wedge dx^l.$$

If we make this a $t = \text{const}$ hypersurface, then the above becomes

$$i_{v_{(p)}}\alpha_M = \sqrt{-g}v^4_{(p)}dx^1 \wedge dx^2 \wedge dx^3.$$

This means that for this particular choice for U, the integral in [6.5.1] can be formulated in terms of the more familiar integral over $\mathbb{R}^4$,

$$\begin{aligned} M_{(p)} &= (-1)^{p+1}\int_U \sqrt{-g}\mu_{(p)}v^4_{(p)}dx^1 \wedge dx^2 \wedge dx^3 && [6.5.2] \\ &= (-1)^{p+1}\int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} \sqrt{-g}\mu_{(p)}v^4_{(p)}\, d\hat{r}\, d\hat{\theta}\, d\hat{\phi} \\ &= (-1)^{p+1}\int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} \mu_{(p)}v^4_{(p)}e^{\alpha/2}e^{\gamma/2}\hat{r}^2\sin\hat{\theta}\, d\hat{r}\, d\hat{\theta}\, d\hat{\phi} \\ &= 4\pi(-1)^{p+1}\int_0^{\infty} \mu_{(p)}v^4_{(p)}e^{\alpha/2}e^{\gamma/2}\hat{r}^2\, d\hat{r}, && [6.5.3] \end{aligned}$$

and this is essentially the relationship between $M_{(p)}$ and $\mu_{(p)}$ spoken of earlier. This result can be developed further; suppose we *define* $M_{(p)}(r)$ as the integral,

$$M_{(p)}(r) = 4\pi(-1)^{p+1}\int_0^{r} \mu_{(p)}v^4_{(p)}e^{\alpha/2}e^{\gamma/2}\hat{r}^2\, d\hat{r}.$$

Then this evaluates the mass in a region, $[0, r]$, of a $t = \text{const}$ hypersurface. Now, since geodesics in dust p never intersect (an artifact of the multi-dust model that we have set up), it follows that $M_{(p)}\left(r_{(p,n)}\right)$ is a conserved quantity. This is because the region enclosed by $r_{(p,n)}$ is comoving with respect to dust p. In other words, since there is zero flux of dust p across the timelike surface swept out by the boundary of this region, i.e. $r = r_{(p,n)}$, the mass inside, i.e. $M_{(p)}\left(r_{(p)}\right)$, must remain constant. Now consider a finite number, N, of geodesics, indexed by i and denoted by $r_{(p,n,i)}$. We shall order these curves such that $0 < r_{(p,0,i)} < r_{(q,0,i+1)} < r_{(r,0,N)}$ on the initial time slice, but place no restriction on their order for arbitrary t. p, q and r represent dust numbers and are not necessarily different, however, it is important to realise that as these curves may cross, it is likely that these dust numbers will change. Returning to our discussion on conserved quantities, it follows that $M_{(p)}\left(r_{(p,0,i)}\right)$ is conserved as is,

$$M_{(p,i)} = M_{(p)}\left(r_{(p,0,i)}\right) - M_{(q)}\left(r_{(q,0,i-1)}\right). \qquad [6.5.4]$$

The latter we shall call the *conserved mass* between the ith geodesic and its nearest neighbour in the direction of the origin *on the initial time slice.* In most cases, at any time, the dustnumbers, p and q, of *adjacent* geodesics will be the same. The only exception to this is near the caustic where, due to the crossing of shells, we might have, $r_{(1,n,i-1)} > r_{(2,n,i)}$, whereas initially we had, $r_{(1,0,i-1)} < r_{(1,0,i)}$. The $M_{(p,i)}$ are important and we shall return to these later when we discuss ways of discretising the dust continuum to implement in a computer program.

The next step in this process of obtaining an expression for $\mu_{(p)}$ is to remove the integral sign in equation [6.5.3]. As mentioned, because geodesics corresponding to dust p never intersect, we can conclude that

$$M_{(p)}\left(r_{(p)}\right) = (-1)^{p+1} 4\pi \int_0^{r_{(p)}} \mu_{(p)} v_{(p)}^4 e^{\alpha/2} e^{\gamma/2} \hat{r}^2 \, d\hat{r}$$

is a constant. This equation can, of course, be rewritten in terms of a differential equation in $M_{(p)}(r)$. We obtain the result:

$$\begin{aligned} \mu_{(p)}\left(r_{(p)}\right) &= \frac{(-1)^{p+1}}{4\pi v_{(p)}^4 e^{\alpha/2} r_{(p)}^2 e^{\gamma/2}} \frac{dM_{(p)}}{dr}\left(r_{(p)}\right) \\ &= \frac{1}{4\pi u_{(p)}^4 e^{\alpha/2} r_{(p)}^2} \left| \frac{dM_{(p)}}{dr}\left(r_{(p)}\right)\right|, \end{aligned} \qquad [6.5.5]$$

which defines our density function, thus completing one of the aims of this section. If we are associating with each geodesic a conserved mass (as in equation [6.5.4]) then approximately we have

$$\mu_{(p)}\left(r_{(p,n,i)}\right) \approx \frac{1}{4\pi u_{(p,n,i)}^4 e^{\alpha(n,i)/2} r_{(p,n,i)}^2} \left| \frac{M_{(p,i)}}{r_{(p,n,i)} - r_{(q,n,i-1)}} \right|, \qquad [6.5.6]$$

where again q is most likely equal to p, the exception being near the caustic. The introduction of the modulus signs allows us to remove the $(-1)^{p+1}$ factor. We can do this since our initial conditions (which we define later) dictate that $M_{(p)}$ increases as r increases away from the origin. When geodesics cross, however, we can see that the conserved masses, $M_{(2,i)}$, although positive, decrease as r increases thus making $dM_{(2)}/dr$ negative. It follows that $(-1)^{p+1}dM_{(p)}/dr$ is always a positive quantity and we illustrate this by inserting the modulus signs. We shall use this approximation wherever $\mu_{(p)}$ occurs in the either the geodesic equation, [6.4.5], or the metric defining equations, [6.2.6] and [6.2.7].

With the ideas developed above in mind, it becomes a relatively simple matter to visualise the representation for the dust continuum that we shall use in our computer program. We discretise our continuum by considering a finite number of reference geodesics and associate with each, three numbers representing the position, velocity and a quantity that is the conserved mass between it and its nearest neighbour on the $t = 0$ time slice. We can then picture the interaction between dusts if we imagine a series of small springs attached to each other, end to end, along a line. The reference points would be represented by the joins and the conserved masses, the springs themselves. A compressed spring would mean a region of high density, a stretched spring, low density. As we move each join axially, we model the movement of our reference points, and if we fold the line of springs back on itself, we model the crossing of adjacent geodesics and the formation of the caustic. In the case where we have a fold, any join will coincide with a spring that is part of the line going the other way. Where a spring and join coincide, we have the situation where the density of dust p say, is influencing the movement of particles on the reference geodesic corresponding to dust q $(q \neq p)$. This is essentially how we model the interaction of shells in our computer model and we stress again the fact that the strength of this kind of interaction is controlled through the non-zero value for the density function, $\mu_{(p)}$, evaluated at $r_{(q)}$ (recall equation [6.4.5]).

In the case where $\mu_{(p)}$ is influencing the movement of dust particles along $r_{(q,n,i)}$, we will need to calculate $dM_{(p)}/dr\left(r_{(q,n,i)}\right)$. To do this, we simply determine the value of j such that $r_{(p,n,j-1)} \leq r_{(q,n,i)} \leq r_{(p,n,j)}$ for $p = 1, 3$, or $r_{(2,n,j-1)} \geq r_{(q,n,i)} \geq r_{(2,n,j)}$ for $p = 2$, and thus

$$\frac{dM_{(p)}}{dr}\left(r_{(q,n,i)}\right) = \frac{M_{(p,j)}}{r_{(p,n,j)} - r_{(p,n,j-1)}}. \qquad [6.5.7]$$

The reason behind the change in direction for the inequalities in the $p = 2$ case is to account for the fact that adjacent geodesics have crossed. Without this proviso, the computer could never satisfy all the requirements stipulated by the combination of inequalities and suffixes. If an i cannot be found then the right hand side is set to zero. This models the possibility that r may not be within the 3-dust region.

Having defined the density function and this notion of conserved masses, we need to describe how this information is carried through a caustic. There are, of course, two cases; the first is the simplest and corresponds to a trajectory passing through the caustic at an angle with the tangent. In this case we do nothing, the conservation of matter tells us that what goes in must come out, and since no geodesics from the same dust cross, the dust numbers do not change. The only physical effect is that the acceleration may see a discontinuity due to the high densities associated with dusts that are travelling parallel to the caustic curve.

The other case takes a little more care and to describe this we need to define a few quantities. The C programming language provides a means to aggregate variables of different types such as integers and doubles. These groups are called *structures.* As each shell has associated with it a conserved mass, position and tetrad velocity, we create a structure called a *particle* to hold this information. For N dust shells in total, we simply create an N-dimensional array of these *particles.* With this notation it is important to realise that since *particle[i].position* ($i = 1, \ldots, N$) holds the value of $r_{(p,n,i)}$ and that we initialise this array at $t = 0$, then i essentially orders the reference geodesics with respect to their initial position. In this sense, the i's in *particle[i].position* and $r_{(p,n,i)}$ are equivalent.

We are now able to discuss how the conserved mass is treated on the caustic. In fact, with the above numerical variables, this process becomes virtually trivial. We first of all define another integer known as the *dustnumber* for each geodesic and store this within the *particle* structure. With regards to the numerics then, the caustic is defined where two adjacent geodesics cross and to model the passage of particles from one dust to another, we simple change the *dustnumber* for the appropriate geodesic. We illustrate by example (and apologise for this is where the discussion becomes specific in the sense of implied boundary conditions). On the right caustic we might initially have two adjacent geodesics such that, $r_{(3,0,i)} < r_{(3,0,i+1)}$ (This inequality could be replaced by *particle[i].position<particle[i+1].position* if we wished to explain

the example in terms of computer variables.). Then if at $t = t_n$, $r_{(3,n,i)} < r_{(3,n,i+1)}$, whereas at a time step later, $r_{(3,n+1,i)} > r_{(3,n+1,i+1)}$, we relabel the dust number of the $(i+1)$th geodesic to be 2 (equivalent to setting *particle[i+1].dustnumber* equal to 2). Likewise on the left caustic, for if at $t = t_n$, $r_{(1,n,i)} < r_{(1,n,i+1)}$, whereas for $t = t_{n+1}$, $r_{(1,n+1,i)} > r_{(1,n+1,i+1)}$, we relabel the dust number of the ith geodesic to be 2. Returning to the original question of how the conserved mass is treated for particles 'touching' the caustic, we find the answer is again do nothing. The conserved mass between the ith and $(i+1)$th geodesic is always the mass between these curves regardless of whether they have crossed or not and thus regardless of their dust number. The only complication that could arise is due to the fact that the 'direction' of the conserved mass flips from going to the left of the geodesic, to going to the right when two curves cross. This would provide a negative $dM_{(p)}/dr$ and thus possibly a negative density, however, this is avoided by the modulus sign in [6.5.5].

Of course it is possible that during any time step two or more particles may be exchanged between dusts. This case is still valid for it simply means that the shell crossings are happening faster than our smallest time step. To check for this we must, for the $p = 1$ case, sequentially repeat the above process checking from $r_{(1,0,n_1)}$ to $r_{(1,0,1)}$ in sequence. Here n_1 corresponds to the largest integer such that $r_{(1,n,n_1)}$ exists and we should understand that this is a dynamic variable; n_1 reduces by 1 each time a reference geodesic 'touches' the caustic. For the $p = 3$ case, we have a similar situation; we repeat the process from $r_{(3,0,n_3)}$ sequentially to $r_{(3,0,N)}$, where now n_3 is the smallest integer such that $r_{(3,n,n_3)}$ exists.

The last point to mention regarding the assignment of dust numbers is that shell crossing can only occur at $r_{(1,0,n_1)}$, if only a *single* particle from dust 1, for example, 'touches' the caustic during a time step, and at $r_{(1,0,n_1)}$ and $r_{(1,0,n_1-1)}$, if two particles are exchanged and so on. If we find ourselves in the regime where the geodesic corresponding to $p = 1$, i say, crosses that corresponding to $p = 1$, $i-1$, with geodesic $i+1$ still a member of dust 1, then we have an error. In physical terms this is a perfectly acceptable phenomena, however, this corresponds to a multi-dust region where $k > 3$. Since our model is rigid in the sense it cannot account for varying k-values, we must abort the calculation.

§6.6. Determination of metric coefficients.

It has taken a while, but we are now at a stage where we have defined all of the mass variables. The only two processes that we have left on our list to describe are methods to determine the metric coefficients, e^{α} and e^{γ}. We begin with the former. If we consider equation [6.5.5] alongside [6.3.6], we can see that again metric coefficients occur within the integral that determines e^{α}. When we had this problem before we were able to get around this by adopting a tetrad formalism. In this case, however, there is no option other than to reformulate the integral equation in [6.3.6] as a differential equation in e^{α}, which we must solve.

For the moment we assume that we are still considering a continuous rather than a discrete model. By inserting matter into equation [6.2.4] via the Einstein equation we have

$$\sum_{p=1}^{3} \frac{\kappa}{4\pi} u_{(p)}^{4} r^{-2} e^{-\alpha/2} \left| \frac{dM_{(p)}}{dr} \right| = r^{-2} - r^{-2} e^{-\alpha} \left(1 - r\alpha_1\right)$$
$$\Longrightarrow e^{-\alpha/2} f(r) = 1 - e^{-\alpha} + r\alpha_1 e^{-\alpha},$$

where

$$f(r) = \sum_{p=1}^{3} \frac{\kappa}{4\pi} u_{(p)}^{4} \left| \frac{dM_{(p)}}{dr} \right| .$$

Putting $y = re^{-\alpha}$ gives

$$\frac{dy}{dr} = e^{-\alpha} - r\alpha_1 e^{-\alpha},$$

and so the above equation transforms to

$$\sqrt{\frac{y}{r}} f(r) = 1 - \frac{y}{r} + \frac{y}{r} - \frac{dy}{dr}$$

$$\Longrightarrow \frac{dy}{dr} = 1 - \sqrt{\frac{y}{r}} f(r). \qquad [6.6.1]$$

If we solve this equation numerically, we have a method by which e^{α} can be determined. To do this, we again adopt the simple Euler method starting from $r = 0$. Thus if $y_{(n)\{i\}} = r_{(n)\{i\}} e^{-\alpha_{(n)\{i\}}}$, where the suffix $\{i\}$ notation represents the geodesics ordered by *current* position rather than initial position (cf. $r_{(n,j)}$ where the i and j might be different.), then

$$y_{(n)\{i+1\}} = y_{(n)\{i\}} + \left(r_{(n)\{i+1\}} - r_{(n)\{i\}}\right) \frac{dy}{dr}\left(r_{(n)\{i\}}\right), \qquad [6.6.2]$$

where

$$\frac{dy}{dr}\left(r_{(n)\{i\}}\right) = 1 - \sqrt{\frac{y_{(n)\{i\}}}{r_{(n)\{i\}}}} \sum_{p=1}^{3} \frac{\kappa}{4\pi} u^4_{(p,n,j)} \left|\frac{dM_{(p)}}{dr}\left(r_{(p,n,j)}\right)\right|. \qquad [6.6.3]$$

Here the curves, $r_{(p,n,j)}$, are determined by the requirement that j is the integer such that $r_{(p,n,j-1)} \leq r_{(n)\{i\}} \leq r_{(p,n,j)}$ for $p = 1,3$ or $r_{(2,n,j-1)} \geq r_{(n)\{i\}} \geq r_{(2,n,j)}$. The derivative, $dM_{(p)}/dr\left(r_{(p,n,j)}\right)$, is as given in [6.5.7], but if a j cannot be found then this contribution to the sum is set to zero. In any case we have

$$e^{-\alpha_{(n)\{i+1\}}} = \frac{y_{(n)\{i+1\}}}{r_{(n)\{i+1\}}}. \qquad [6.6.4]$$

Note that this method of numerical integration jumps from shell to shell according to each shell's position with respect to the origin and does not recognise which dust each shell is a member of (hence the reason for dropping the p suffix in the above). For example, as one performs this iteration for some time slice, t_n, the position of each consecutive shell would obey, $\cdots < r_{(n)\{i\}} < r_{(n)\{i+1\}} < r_{(n)\{i+2\}} < r_{(n)\{i+3\}} < \cdots$ for all i, whereas the corresponding dust numbers could look like, $\cdots \longrightarrow 3 \longrightarrow 3 \longrightarrow 3 \longrightarrow 2 \longrightarrow 3 \longrightarrow 2 \longrightarrow 2 \longrightarrow 3 \longrightarrow 2 \longrightarrow 2 \longrightarrow 2 \longrightarrow \cdots$. The crucial point is that when evaluating $f(r)$, and in particular $e^{\alpha_{(n)\{i\}}}$, even though the ith shell may be a member of dust p say, the other two dusts play a role in evaluating $e^{\alpha_{(n)\{i\}}}$ by contributing a term proportional to their density. We also note that this method can be extremely inaccurate since errors in $dy/dr\left(r_{(n)\{i\}}\right)$ are compounded. This can only be controlled by increasing the number of shells.

The final check to be made, if this method is to be used, is whether or not dy/dr has a well defined limit at $r = 0$. Now,

$$\lim_{r\to 0} \frac{dy}{dr} = 1 - \lim_{r\to 0} \sqrt{\frac{y}{r}} \lim_{r\to 0} f(r),$$

and since

$$\lim_{r\to 0} \sqrt{\frac{y}{r}} = \lim_{r\to 0} e^{-\alpha} = 1$$

by [6.2.2] and $\lim_{r \to 0} f(r)$ exists, it follows that dy/dr has a well defined limit as $r \longrightarrow 0$.

To enable us to implement the above procedure within a computer program it is clear we need to have the particles ordered with respect to their current position. We provide this information by constructing an array of N variables that 'point' to the information stored by a particular *particle.* This array of variables is called *pointers.* In terms of the C programming language, the quantity stored by each element of *pointers* is the address of the memory location that stores the information in the *particle* that it 'points' to and initially, we have *pointers[i]=&particles[i]* where the & stands for address.

We can access the information stored in a *particle* quite simply since we know that *pointers[i]→position* is equivalent to *particles[j].position* if *pointers[i]* 'points' to *particle[j]*, for example. This means that the actual evolution can therefore occur either by moving the *particles* or by moving the *pointers* and in fact we for convenience we shall choose the second method. To ensure that the i in *pointers[i]→position*, essentially orders the geodesics with respect to their current position, we perform a bubble sort which exchanges *pointers* between *particles* according to increasing radial position. Moreover, *pointers[i]→position* is clearly equivalent to the notation, $r_{(n)\{i\}}$, introduced earlier on.

The final point to make regarding the fact that the evolution occurs with respect to *pointers[i]* rather than *particles[j]* is that at some stage, when we check to see if trajectories cross for example, we need to know i given j and visa versa. To provide this information we introduce another integer variable within our *particle* structure called *pointernumber.* This variable is set to be equal to i (*pointers[i]→pointernumber* $= i$) prior to the simulation, but during the above bubble sort, we also exchange *pointernumbers*. This insures that the equation, *pointers[i]→pointernumber* $= i$, is always true. This gives us a route from knowing i to knowing j and back again if we so wish.

The last metric component, $e^{-\gamma}$, is much easier to determine; we simply perform the integration of [6.3.7] as all quantities are given. The technique used in the actual program is simply to approximate the curve within the integral sign by a series of rectangles, whose sides correspond to $r_{(n)\{i\}}$, which are then summed. We can see directly from equation [6.3.7] that again, at each rectangle being considered, the density of all dusts plays a role in calculating $e^{-\gamma}$.

§6.7. Initial conditions, summary and example of results.

This chapter has been dedicated to presenting the techniques used in a computer program designed to model the formation of caustics. We introduced certain variables, $r_{(p,n,i)}$, $u^1_{(p,n,i)}$ and $M_{(p,i)}$, which describe the dusts and formulated a numerical scheme to solve for these quantities. In this section we shall collect together all the ideas and talk our way through a single iteration referring to all relevant equations. We shall begin at the $t = t_n$ time step, stating what we know, and what we need to find out. This means that we will essentially revisit the list of procedures laid out in §6.4. This approach of summarising this chapter avoids the specification of initial conditions. These will be quickly presented towards the end.

1. At the beginning of each time step, we suppose that for each reference geodesic, $r_{(p,n)\{i\}}$, $u^1_{(p,n)\{i\}}$ and $M_{(p)\{i\}}$ are known. Note that this notation is that of §6.6 where the suffix $\{i\}$ labels the geodesics with respect to current position, equivalent to working with the *pointers* computer variable. Note also that $M_{(p)\{i\}} = M_{(p,i)}$ for the initial time slice. Now the first step is to calculate $e^{\alpha(n)}$ at $r = r_{(n)\{i\}}$ for every i. The method is an Euler technique, i.e.,

$$y_{(n)\{i+1\}} = y_{(n)\{i\}} + \left(r_{(n)\{i+1\}} - r_{(n)\{i\}}\right) \frac{dy}{dr}\left(r_{(n)\{i\}}\right) \qquad [6.6.2]$$

and

$$\frac{dy}{dr}\left(r_{(n)\{i\}}\right) = 1 - \sqrt{\frac{y_{(n)\{i\}}}{r_{(n)\{i\}}}} \sum_{p=1}^{3} \frac{\kappa}{4\pi} u^4_{(p,n,j)} \left| \frac{dM_{(p)}}{dr}\left(r_{(p,n,j)}\right)\right|. \qquad [6.6.3]$$

Here, for *pointers[i]→dustnumber* = p, we determine j such that $r_{(n)\{i\}} = r_{(p,n,j)}$ so that

$$\frac{dM_{(p)}}{dr}\left(r_{(p,n,j)}\right) = \frac{M_{(p,j)}}{r_{(p,n,j)} - r_{(q,n,j-1)}}.$$

For *pointers[i]→dustnumber* $\neq p$, we determine j such that, $r_{(p,n,j-1)} < r_{(n)\{i\}} < r_{(p,n,j)}$ for $p = 1, 3$, and $r_{(2,n,j-1)} > r_{(n)\{i\}} > r_{(2,n,j)}$ otherwise. In this case

$$\frac{dM_{(p)}}{dr}\left(r_{(p,n,j)}\right) = \frac{M_{(p,j)}}{r_{(p,n,j)} - r_{(p,n,j-1)}}.$$

The last step can be done because we introduced the integer, *pointernumber*, into the *particle* structure. Finally,

$$e^{-\alpha_{(n)\{i+1\}}} = \frac{y_{(n)\{i+1\}}}{r_{(n)\{i+1\}}}. \qquad [6.6.4]$$

2. Calculate $e^{\gamma(n)}$ at $r = r_{(n)\{i\}}$ (final paragraph of §6.6) for every i.

3. Calculate $\mu_{(n)\{i\}}$ for each i. This uses

$$\mu_{(p)}\left(r_{(p,n)\{i\}}\right) \approx \frac{1}{4\pi u^4_{(p,n)\{i\}} e^{\alpha_{(n)\{i\}}/2} r^2_{(p,n)\{i\}}} \left|\frac{dM_{(p)}}{dr}\left(r_{(p,n,j)}\right)\right|,$$

where $dM_{(p)}/dr\left(r_{(p,n,j)}\right)$ is defined in 1 above. The above equation for the density is equivalent to [6.5.6].

4. Calculate $u^1_{(p,n+1)\{i\}}$ for each i using

$$u^1_{(p,n+1)\{i\}} \approx u^1_{(p,n)\{i\}} + \left(t_{n+1} - t_n\right) \frac{du^1_{(p)\{i\}}}{dt}\left(t_n\right),$$

which is essentially equation [6.4.3], and

$$\begin{aligned} \frac{du^1_{(p,n)\{i\}}}{dt} = -\frac{1}{2} e^{(\alpha_{(n)\{i\}}+\gamma_{(n)\{i\}})/2} \Bigg(& \frac{u^4_{(p,n)\{i\}}}{r}\left(1 - e^{-\alpha_{(n)\{i\}}}\right) \\ & - \kappa r \mu_{(q,n,j)} u^1_{(q,n,j)} \left(u^4_{(q,n,j)} u^1_{(p,n)\{i\}} - u^4_{(p,n)\{i\}} u^1_{(q,n,j)}\right) \\ & - \kappa r \mu_{(r,n,k)} u^1_{(r,n,k)} \left(u^4_{(r,n,k)} u^1_{(p,n)\{i\}} - u^4_{(p,n)\{i\}} u^1_{(r,n,k)}\right)\Bigg), \end{aligned}$$

where j and k are defined such that, $r_{(q,n,j-1)} < r_{(p,n)\{i\}} < r_{(q,n,j)}$ and $r_{(r,n,k-1)} < r_{(p,n)\{i\}} < r_{(r,n,k)}$ for $q, r = 1, 3$, and $r_{(2,n,j-1)} > r_{(p,n)\{i\}} > r_{(2,n,j)}$ and $r_{(2,n,k-1)} < r_{(p,n)\{i\}} > r_{(2,n,k)}$ otherwise. This equation is equivalent to [6.4.5].

5. Calculate $r_{(p,n+1)\{i\}}$ for each i. This process uses

$$r_{(p,n+1)\{i\}} \approx r_{(p,n)\{i\}} + \left(t_{n+1} - t_n\right) \frac{u^1_{(p,n)\{i\}}}{\sqrt{1 + u^1_{(p,n)\{i\}} u^1_{(p,n)\{i\}}}} e^{\left(\gamma_{(n)\{i\}} - \alpha_{(n)\{i\}}\right)/2},$$

which is equivalent to [6.4.2].

Of course, we cannot immediately proceed to the calculation of the $n+2$ quantities because we have to check if any geodesics have crossed. Thus we should add:

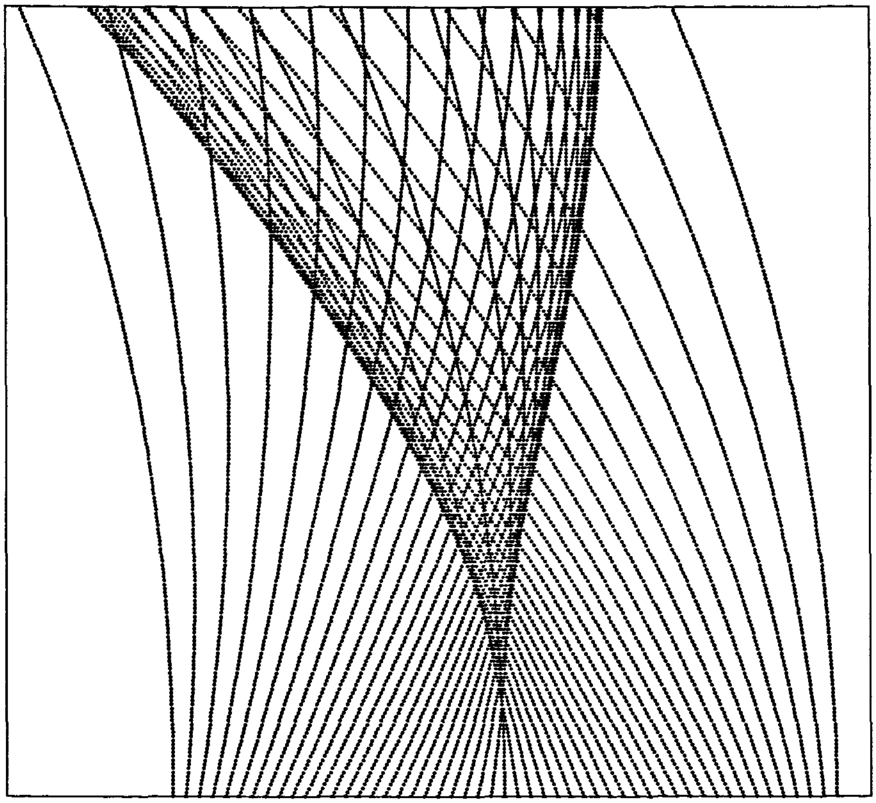
6. Check for any geodesic crossings and adjust *dustnumber* accordingly. These processes are described in §6.5, paragraph 13.

We now quickly present the choice of initial conditions used in our simulation. The idea here is to copy as far as possible, the initial conditions obtained from the zero gravity case (ref. chapter 3). Thus we chose $(r_c, 0)$ to be the point of cusp formation, and define our N geodesics such that they were distributed with a finite interval between each, centred around r_c. We supposed a uniform mass distribution throughout M. This means that if D represents a constant density per unit area, then the mass of each shell is given by $M_{(i)} = D \cdot (r_{(0)\{i\}})^2$. Finally, we chose the velocity so as to approximately mirror that of the gravity free initial conditions. That is to say,

$$u^1_{(0)\{i\}} = -V \cdot \left(r_{(0)\{i\}} - O\right)^{1/3} \left(1 - \exp\left(6\left(\left|r_{(0)\{i\}} - r_c\right| - c\right)\right)\right).$$

Here V is some user defined constant that has units of velocity. The choice of c in the above exponential defines a multiplicative factor that modulates the $\left(r_{(0)\{i\}} - O\right)^{1/3}$ term, so that for $r_{(0)\{i\}} \approx r_c$, $\left(1 - \exp\left(6\left(\left|r_{(0)\{i\}} - r_c\right| - c\right)\right)\right) \approx 1$, whereas for $\left|r_{(0)\{i\}} - r_c\right| \approx c$, $\left(1 - \exp\left(6\left(\left|r_{(0)\{i\}} - r_c\right| - c\right)\right)\right) \approx 0$. This mimics the initial conditions for the simple gravity-free model (ref. chapt. 3) close to the cusp, but allows for the velocity to tail off as $\left|r_{(0)\{i\}} - r_c\right|$ becomes sufficiently large. This enables us to compare our results (fig. 5) with that presented in fig. 1.

Fig. 5. Caustic produced by a spherically symmetric dust.



CHAPTER 7. MATHEMATICAL APPROACH TO CAUSTICS IN GENERAL RELATIVITY.

§7.1. Introduction.

The previous chapter described a model for caustic formation within the context of a spherically symmetric formulation of General Relativity. We can think of this as essentially being an extension to the multi-dust construction of the equivalent problem in Newtonian theory: equations [5.5.1]–[5.5.3] were simply replaced by their relativistic analogues. The differential equations that really constitute the model were presented ([6.1.1]–[6.1.3]), and a method by which a numerical solution could be obtained was proposed. This solution was based on the same initial conditions as that for the zero gravity caustic discussed in chapter 3, about a new radial origin at $r = r_c$, and assuming a uniform density distribution for the initial time slice.

The initial reason as to why we chose to change tack and construct a computer simulation of caustics within the framework of General Relativity was to try to grasp some sort of understanding of the processes at work during cusp formation. This implies the need for some sort of asymptotic analysis of equations [6.1.1]–[6.1.3] and the plan was to perform this investigation numerically. In the end we chose the more rigorous mathematical approach, which we shall now describe, and the numerical work really became an exercise in formulating the General Relativistic problem correctly. This explains the lack of results and conclusions in the previous chapter. Having said that, we do not wish to give the impression that the work presented in chapter 6 can be overlooked. In fact it is significant because it provides us with a deep understanding of how each term in our defining equations should be formulated mathematically. As examples of this we have the definition of the volume forms, $\alpha_{(p)}$, such that their orientation is always positive with respect to the orientation supplied by the standard volume form ($\alpha_{(p)} = (-1)^{p+1} i_{v_{(p)}} \alpha_M$ in other words), or the idea of writing $\mu_{(p)}$ in terms of the derivative $|dM_{(p)}/dr|$, and indeed the insertion of the modulus signs. There are other examples, but these are particularly important for it was only when we concentrated on formulating the $\mu_{(p)}$ in terms of derivatives of $M_{(p)}$, did we understand the real significance of the effect that a non-orientation preserving projection map, π, when restricted to $S \subset TM$, has on our equations. Specifically, this understanding forced us to interpret our density functions in terms of pseudo-$(m-1)$-forms on M, insert the $(-1)^{p+1}$ factor in the definitions of $\alpha_{(p)}$ and $\sigma_{(p)}$ (found in §4.5), and also the $(-1)^{i+1}$ correction in equation [4.5.3]. Thus, although not eluded to in any of the previous chapters, it was only when the numerical work was well under way did we begin to understand the intricacies involved with the density functions. This enabled us to backtrack and correct some of the work that had already been done, particularly on the tv space approach to similarity solutions in the Newtonian formulation of caustics.

As mentioned, the intention is to investigate the shape of the caustic as we consider a smaller and smaller neighbourhood containing the cusp. This would be equivalent to the analysis of chapter 5 where we investigated the possibility that cusp formation is independent of gravity. One might ask why repeat this analysis, or ask this question again, when we have already established that in the Newtonian case at least, gravity plays no part in cusp formation. Well the previous sentence essentially answers its own question. General Relativity is considered to give a far greater insight into the mechanisms behind the gravitational interaction than the simpler Newtonian description. The equations that we associate with any problem formulated within the framework of General Relativity are generally far more complicated and non-linear than their Newtonian equivalent. It is therefore conceivable that gravity can in fact play a part in cusp formation via the extra terms each equation has. Thus, to be sure our original conclusion is correct, we need to repeat the analysis of §5.3 but based on the equations supplied by General Relativity.

This chapter does just this. We choose the more rigorous method using the asymptotic solutions developed in chapter 5. That is to say, we construct rational algebraic curves along which we approach the cusp in a manner that allows us to define a limiting process based on the group of transformations given by [5.3.1]. In order for us to be able to define such a process, certain assumptions need to be made concerning the continuity of the metric components: i.e. we assume that α and γ are continuous functions. This requirement ensures that the curves we define are continuous particularly as we pass through the caustic. This will be explained in detail in §7.4. but we can summarise by stating that we require the densities to be integrable.

En route to doing this work, we shall consider another limit to our General Relativistic equations of motion. This is the Newtonian limit where we assume all velocities are small compared with c, the speed of light, but still keeping any mass contributions. This statement needs defining and to do this we must first of all convert equations [6.1.1]–[6.1.3] so that all variables are in terms of physical units. As is generally the case when discussing aspects of General Relativity, we assume that $c = G = 1$ for simplicity. This assumption is equivalent to setting all dimensions equal to that of length and the resulting equations are said to be written in terms of *geometrised* units. Thus to begin, we must reinsert the c's and G's. This process is discussed in the next section.

Having achieved this, we can now non-dimensionalise our General Relativistic equations of motion. This is done in the latter parts of the next section by replacing any variable, such as r for example, by a product of a constant that holds the units of that variable, and a dimensionless scaling factor. Thus in our example we might have $r = r_N L$, where L is a constant and having dimension length, and r_N is some non-dimensional parameter. If we introduce M and T to represent the dimensions of mass and time respectively, it becomes possible for each term in any equation, to separate and group together all the constants providing the units for that particular term, and all the dimensionless parameters that are describing the physics. If we multiply our equation by the relevant dimension, i.e L, T or M, we obtain dimensionless groups that can tell us which terms are significant in the differing velocity régimes. It is using these ideas that we shall define our notion of the Newtonian limit.

Without going into any detail, we find that all the resulting dimensionless groups can be written as some product of the following two non-dimensional quantities: GMT^2/L^3 and L/cT. Now, supposing that whilst keeping the first group finite, we allow the second group to tend towards zero. Physically, this means we assume that any velocity (with dimension L/T) is small when compared to that of light. The fact that the first group is finite allows any term that arises from quantities that are purely mass driven, and hence would contribute to any gravitational effects, to remain. For comparison, we can say that the opposite limit, i.e. allowing the first term to tend towards zero whilst keeping the second finite, gives the situation where on M we have caustics formed by massless particles moving with velocities that in some cases can be said to be a significant fraction of c. Now, it is the former that we are most interested in and so §7.3 and §7.4 essentially take the non-dimensionalised relativistic equations provided by §7.2, and determine those terms that remain after the limit, $L/cT \longrightarrow 0$, is taken (the definition of the Newtonian limit). The hope and expectation is that the resulting equations will be equivalent to [5.5.1]–[5.5.3]. If we do not obtain this result, then clearly the two sets of equations are describing different mechanisms of caustic formation. It may be that a General Relativistic formulation of spherically symmetric caustics provides us with a different caustic type [A] to that of the simple caustic or worse, that one or both sets of the modelling differential equations are wrong.

§7.5 and §7.6 complete this chapter by considering the asymptotic limit of the General Relativistic equations of motion. Because we have already spoken of this, and also because the techniques are so close to those considered in §5.3 and §5.5, we shall not elaborate any further. We shall, however, conclude by making a few observations regarding the results so far obtained. The expectation is that the procedures: let $L/cT \longrightarrow 0$, let $\varepsilon \longrightarrow \infty$, should commute. In our conclusion then, we say whether or not this is a true statement. We are also interested in the equations that the procedure, let $\varepsilon \longrightarrow \infty$, alone yields. Of course we expect the resulting equations to describe the physics of cusp formation and because they are derived from a more complicated theory of gravity, it is possible for them to be different to the asymptotic limit of the spherically symmetric Newtonian equations (§5.5). If this is the case, then we have the situation where our analysis suggests that gravity does in fact play a part in cusp formation; a clear contradiction to the conclusion of chapter 5 and an extremely important result.

In chapter 6 we found it useful to formulate our equations of motion with respect to a tetrad of orthonormal vectors acting as a basis for TM. In the following three sections we shall find it easier to use the ordinary coordinate basis so let us briefly reconstruct the geodesic equation in these terms. We have

$$\nabla_{v_{(p)}} v_{(p)} = 0$$

$$\Longrightarrow v^i_{(p)} \nabla_{\partial_i} \left(v^j_{(p)} \partial_j \right) = 0$$

$$\Longrightarrow v^i_{(p)} \partial_i \left(v^j_{(p)} \right) \partial_j + v^i_{(p)} v^j_{(p)} \Gamma^k{}_{ij} \partial_k = 0.$$

Since $v^i_{(p)} = dx^i_{(p)}/d\tau$, it follows that

$$\frac{dv^k_{(p)}}{d\tau} + \Gamma^k{}_{ij} v^i_{(p)} v^j_{(p)} = 0$$

$$\Longrightarrow \frac{dv^1_{(p)}}{d\tau} + \Gamma^1{}_{11} v^1_{(p)} v^1_{(p)} + 2\Gamma^1{}_{14} v^1_{(p)} v^4_{(p)} + \Gamma^1{}_{44} v^4_{(p)} v^4_{(p)} = 0.$$

If we work directly with the energy-momentum tensor given by equation [6.3.1], then the definition of the Christoffel symbol and equations [6.2.3]–[6.2.5] imply that

$$\Gamma^1{}_{11} = \frac{1}{2}\alpha_1 = \frac{\kappa}{2} r e^{\alpha} \sum_{p=1}^{3} \mu_{(p)} v^4_{(p)} v^4_{(p)} e^{\gamma} + \frac{1}{2r}\left(1 - e^{\alpha}\right),$$

$$\Gamma^1{}_{14} = \frac{1}{2}\alpha_4 = -\frac{r}{2} e^{\alpha} \kappa \sum_{p=1}^{3} \mu_{(p)} v^1_{(p)} v^4_{(p)} e^{\gamma},$$

and

$$\Gamma^1{}_{44} = \frac{1}{2} e^{\gamma-\alpha}\gamma_1 = \frac{1}{2r} e^{\gamma}\left(1 - e^{-\alpha}\right) + \frac{\kappa}{2} r e^{\gamma} \sum_{p=1}^{3} \mu_{(p)} v^1_{(p)} v^1_{(p)} e^{\alpha}.$$

Substituting these results into the above form for the geodesic equation finally gives

$$\frac{dv^1_{(p)}}{d\tau} = \frac{1}{2r}\left(e^{-\alpha} - 1\right) - \frac{\kappa}{2} r e^{\alpha+\gamma} \left\{ v^4_{(p)} v^4_{(p)} \sum_{q=1}^{3} \mu_{(q)} v^1_{(q)} v^1_{(q)} \right.$$
$$\left. -2 v^1_{(p)} v^4_{(p)} \sum_{q=1}^{3} \mu_{(q)} v^1_{(q)} v^4_{(q)} + v^1_{(p)} v^1_{(p)} \sum_{q=1}^{3} \mu_{(q)} v^4_{(q)} v^4_{(q)} \right\}. \qquad [7.1.1]$$

§7.2. Non-dimensionalising the General Relativistic equations of motion.

This section discusses the concept of non-dimensionalising the General Relativistic equations of motion. It essentially acts as a precursor to the following two sections in the sense that it formulates all our equations in a way so that they are immediately applicable in a procedure which determines their Newtonian limit. Now, before we can consider non-dimensionalising our equations, we need to ensure that they are written in terms of physical units. This requires that we reintroduce the c's and G's. To clarify this we should point out that, as is the case for the majority of problems in General Relativity, we usually simplify our equations by setting $c = G = 1$. This process is equivalent to supposing that all units have dimensions of length and we say that we work in *geometrised* units. It means, for example, that if in physical units a quantity has dimensions of $L^mT^nM^p$ then in geometrised units the same quantity has dimension, L^{m+n+p}.

We have a procedure, however, that enables us to reverse the above action [W]. We can see that $[c] = LT^{-1}$ and $[G/c^2] = LM^{-1}$; it follows therefore that

$$L^mT^nM^p\left[c^n\left(\frac{G}{c^2}\right)^p\right] = L^{m+n+p} \qquad [7.2.1]$$

and so to convert a quantity, A, written in geometrised units to physical units, we simply replace A by $Ac^{n-2p}G^p$ if, in physical units, $[A] = L^mT^nM^p$. It is best to illustrate this process by example. The Minkowski metric written in geometrised units is given by

$$ds^2 = dx^2 + dy^2 + dz^2 - dt^2.$$

From this it follows that $[t] = L$, since otherwise terms in the same equation would have different units. In physical units, however, $[t] = T$ and so by the procedure outlined above, we must replace t by ct to obtain

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2.$$

For a second example, consider the Schwarzschild radius,

$$r = 2M.$$

Again, our procedure implies that the conversion factors for r and M are 1 and G/c^2 respectively. Hence in physical units, the Schwarzschild radius is given by

$$r = \frac{2MG}{c^2}.$$

Let us now consider our particular equations. We begin with [6.1.1]. In its mixed form the only relevant non zero components of G are $G^1{}_1$, $G^4{}_4$ and $G^1{}_4$. Thus for $G^1{}_1 = -\kappa T^1{}_1$ we have

$$r^{-2} - r^{-2}e^{-\alpha}(1 + r\gamma_1) = -\kappa\sum_{p=1}^{3}\mu_{(p)}v^1_{(p)}v^1_{(p)}e^{\alpha}.$$

The conversion factors for r and $v^1_{(p)}$ are 1 and $1/c$ respectively since $[r] = L$ and $\left[v^1_{(p)}\right] = \left[dr_{(p)}/d\tau\right] = L/T$. The quantities, α and γ, are dimensionless (it doesn't make sense to exponentiate a quantity with dimensions) and they must have a conversion factor of 1. The fact that $\left[\mu_{(p)}\right] = ML^{-3}$ implies a conversion factor of G/c^2 so that in physical units the above Einstein equation becomes

$$r^{-2} - r^{-2}e^{-\alpha}(1 + r\gamma_1) = -\frac{\kappa G}{c^4}\sum_{p=1}^{3}\mu_{(p)}v^1_{(p)}v^1_{(p)}e^{\alpha}.$$

In a similar manner the remaining components of the Einstein equation are

$$r^{-2} - r^{-2}e^{-\alpha}(1 - r\alpha_1) = \frac{\kappa G}{c^2}\sum_{p=1}^{3}\mu_{(p)}v^4_{(p)}v^4_{(p)}e^{\gamma}$$

and

$$r^{-1}e^{-\alpha}\alpha_4 = \frac{\kappa G}{c^2}\sum_{p=1}^{3}\mu_{(p)}v^1_{(p)}v^4_{(p)}e^{\gamma}.$$

Slightly more complicated, although still using the same procedure, the geodesic equation becomes

$$\frac{dv^1_{(p)}}{d\tau} = \frac{c^2}{2r}\left(e^{-\alpha} - 1\right) - \frac{\kappa G}{2c^2}re^{\alpha+\gamma}\left\{v^4_{(p)}v^4_{(p)}\sum_{q=1}^{3}\mu_{(q)}v^1_{(q)}v^1_{(q)}\right.$$
$$\left. -2v^1_{(p)}v^4_{(p)}\sum_{q=1}^{3}\mu_{(q)}v^1_{(q)}v^4_{(q)} + v^1_{(p)}v^1_{(p)}\sum_{q=1}^{3}\mu_{(q)}v^4_{(q)}v^4_{(q)}\right\},$$

whereas the continuity equation $\left(T^{ij}_{(p);j} = 0 \text{ or } [6.1.3]\right)$ is simply

$$\frac{1}{\sqrt{-g}}\left(\mu_{(p)}\sqrt{-g}v^i_{(p)}\right)_{,i} = 0,$$

thus completing the first stage.

Now that all the relevant equations are written in terms of physical units, we can consider non-dimensionalising them. As hinted at in the introduction, to do this we introduce a length, a time and a mass parameter denoted by L, T and M respectively, which have, for example, units of metres, seconds and kilogrammes. This enables us to define dimensionless variables, r_N, t_N and $\mu_{N(p)}$, according to

$$r = r_N L, \qquad [7.2.2]$$

$$t = t_N T, \qquad [7.2.3]$$

and

$$\mu_{(p)} = \mu_{N(p)} \frac{M}{L^3}. \qquad [7.2.4]$$

For the velocity, however, things are not so simple. This is because we essentially have two parameters, L/T and c, which can be used to non-dimensionalise any velocity component. In other words, since $\left[dr_{(p)}/d\tau\right] = LT^{-1}$, we could have either

$$v^1_{(p)} = \frac{L}{T}\frac{dr_{N(p)}}{d\tau_N}$$

or

$$v^1_{(p)} = c\frac{dr_{N(p)}}{d\tau_N},$$

and it would be naive of us to assume any of these without further investigation. Now, the metric (equation [6.2.1]) implies that

$$\begin{aligned} c^2 d\tau^2 &= -g_{ij} dx^i dx^j \\ &= -e^\alpha dr^2 - r^2 d\sigma^2 + c^2 e^\gamma dt^2. \end{aligned}$$

Thus we have

$$d\tau = dt\sqrt{e^\gamma - \frac{e^\alpha}{c^2}\left(\frac{dr}{dt}\right)^2},$$

since for spherical symmetry, θ and ϕ are not functions of time. It follows that

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{e^\gamma - \frac{L^2}{c^2 T^2}\frac{e^\alpha}{c^2}\left(\frac{dr_N}{dt_N}\right)^2}},$$

and this clearly represents the non-dimensionalised form for the time component of any 4-velocity we might have. In order to simplify the above expression we define the dimensionless variable, f, to be the denominator in the above equation. Thus,

$$f^2 = e^\gamma - \frac{L^2}{c^2 T^2} e^\alpha \left(\frac{dr_N}{dt_N}\right)^2$$

and

$$\frac{dt}{d\tau} = \frac{1}{f}. \qquad [7.2.5]$$

Finally, we can use this result to non-dimensionalise any radial component of the velocity. We have

$$\begin{aligned}\frac{dr}{d\tau} &= \frac{dr}{dt}\frac{dt}{d\tau} \\ &= \frac{L}{T}\frac{1}{f}\frac{dr_N}{dt_N}.\end{aligned} \qquad [7.2.6]$$

Using these results, we can now complete this section by non-dimensionalising the Einstein, geodesic and continuity equations. It is a process where we simply replace our usual variables with their dimensionless equivalent using [7.2.2]–[7.2.6]. We obtain:

$$r_N^{-2} - r_N^{-2}e^{-\alpha}\left(1 + r_N\frac{\partial\gamma}{\partial r_N}\right) = -\frac{\kappa GLM}{c^4T^2}\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f_{(p)}^2}v_{(p)}v_{(p)}e^{\alpha}, \qquad [7.2.7]$$

$$r_N^{-2} - r_N^{-2}e^{-\alpha}\left(1 - r_N\frac{\partial\alpha}{\partial r_N}\right) = \frac{\kappa GM}{c^2L}\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f_{(p)}^2}e^{\gamma}, \qquad [7.2.8]$$

$$r_N^{-1}e^{-\alpha}\frac{\partial\alpha}{\partial t_N} = \frac{\kappa GM}{c^2L}\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f_{(p)}^2}v_{(p)}e^{\gamma}, \qquad [7.2.9]$$

$$\begin{aligned}\frac{d}{dt_N}\left(\frac{1}{f_{(p)}}v_{(p)}\right) = {} & \frac{c^2T^2}{L^2}\frac{1}{2r_N}f_{(p)}\left(e^{-\alpha} - 1\right) \\ & - \frac{\kappa GM}{c^2L}\frac{r_N}{2}e^{\alpha+\gamma}\frac{1}{f_{(p)}}\left\{\sum_{q=1}^{3}\mu_{N(q)}\frac{1}{f_{(q)}^2}v_{(q)}v_{(q)}\right. \\ & \left. -2v_{(p)}\sum_{q=1}^{3}\mu_{N(q)}\frac{1}{f_{(q)}^2}v_{(q)} + v_{(p)}v_{(p)}\sum_{q=1}^{3}\mu_{N(q)}\frac{1}{f_{(q)}^2}\right\}\end{aligned} \qquad [7.2.10]$$

and

$$\frac{1}{\sqrt{-g_N}}\frac{\partial}{\partial t_N}\left(\mu_{N(p)}\sqrt{-g_N}\frac{1}{f_{(p)}}\right) + \frac{1}{\sqrt{-g_N}}\frac{\partial}{\partial r_N}\left(\mu_{N(p)}\sqrt{-g_N}\frac{v_{(p)}}{f_{(p)}}\right) = 0. \qquad [7.2.11]$$

In the above we have assumed the shorthand notation, $v_{(p)} = dr_{N(p)}/dt_N$. This in turn defines the quantity, $f_{(p)}$, by

$$f_{(p)}^2 = e^{\gamma} - \frac{L^2}{c^2T^2}e^{\alpha}v_{(p)}v_{(p)}. \qquad [7.2.12]$$

We have also implicitly defined $g_N = r_N^4\sin^2\theta e^{\alpha}e^{\gamma}$.

§7.3. Newtonian limit of the geodesic equation.

In this section we shall discuss the Newtonian limit of the geodesic equation. This forms part of an important process that seeks to determine the dominant characteristics of our equations of motion ([6.1.1]–[6.1.3]) in the limit of small velocities. We shall make this statement more precise in a moment, however, it is important to realise the significance of this calculation. From a physical point of view, the situation where we only consider relatively slow moving, massive particles, can quite easily fall within the validity of a Newtonian-like description. It follows therefore, that we expect our General Relativistic equations in the limit of small velocities to be identical to those of [5.5.1]–[5.5.3]. This is the expectation; if for some reason there is a difference, then in order for us to present a reasonably complete study of spherically symmetric caustic formation, we will need to investigate why. It might be, for example, that the two models are describing different caustic types [A], or even that one or both of the defining differential equations are wrong. Clearly we need to check this.

The idea of considering only slow moving bodies in General Relativity we shall call the Newtonian limit. To define this, we first of all note that the process of non-dimensionalising our differential equations highlights a series of *dimensionless groups*: GLM/c^4T^2, GM/c^2L and L/cT (ref. equations [7.2.7]–[7.2.12]). By looking carefully we can see that there are in fact only two groups from which all the others can be determined. These are GMT^2/L^3 and L/cT. This can be seen if we write

$$\frac{GM}{c^2L} = \frac{GMT^2}{L^3} \times \left(\frac{L}{cT}\right)^2$$

and

$$\frac{GLM}{c^4T^2} = \frac{GM}{c^2L} \times \left(\frac{L}{cT}\right)^2 .$$

Thus it becomes possible to formally define the Newtonian limit as the end product of a process that continuously reduces the dimensionless group, L/cT, to zero whilst keeping GMT^2/L^3 finite. Physically, this is telling us that in the limit, velocities with dimension L/T are small in magnitude when compared to c. The last condition is important for it ensures that in the limit we consider the movement of massive dust particles and hence a gravitational interaction. This section performs this calculation on the geodesic equation, [7.2.10], and attempts to determine those terms that are significant in the sense that they remain after the limit has been taken.

The assumption that is implied throughout this analysis is that all quantities can be expressed as a power series in L/cT. We shall also make use of the Lesbesgue dominated convergence theorem which allows us to exchange limits and integral signs. With this approach we can easily take each quantity found within [7.2.10] and determine its behaviour as we allow this group to tend to zero. We first of all choose $f^2_{(p)}$. From [7.2.12] we have

$$\begin{aligned}\lim_{L/cT\to 0} f^2_{(p)} &= \lim_{L/cT\to 0}\left\{e^{\gamma} - \frac{L^2}{c^2T^2}e^{\alpha}v_{(p)}v_{(p)}\right\}\\ &= \lim_{L/cT\to 0} e^{\gamma}.\end{aligned} \qquad [7.3.1]$$

If we integrate equation [7.2.8] with respect to r_N we obtain the non-dimensionalised form of [6.2.6] and so

$$\begin{aligned}\lim_{L/cT\to 0} e^{-\alpha} &= \lim_{L/cT\to 0}\left\{1 - \frac{\kappa GM}{c^2L}\frac{1}{r_N}\int_0^{r_N} r^2\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f^2_{(p)}}e^{\gamma}\,dr\right\} \qquad [7.3.2]\\ &= 1.\end{aligned}$$

Similarly, if we subtract [7.2.8] from [7.2.7] and integrate, we get the non-dimensional equivalent of [6.2.7] and

$$\begin{aligned}\lim_{L/cT\to 0} e^{\gamma} &= \lim_{L/cT\to 0}\left\{e^{-\alpha}\exp\left\{\frac{\kappa GM}{c^2L}\int_0^{r_N} re^{\alpha}\left\{\sum_{p=1}^{3}\frac{L^2}{c^2T^2}\mu_{N(p)}\frac{1}{f^2_{(p)}}v_{(p)}v_{(p)}e^{\alpha}\right.\right.\right.\\ &\qquad\left.\left.\left.+\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f^2_{(p)}}e^{\gamma}\right\}dr\right\}\right\}\\ &= \lim_{L/cT\to 0}\left\{e^{-\alpha}\exp\left\{\frac{\kappa GM}{c^2L}\int_0^{r_N} re^{\alpha}\sum_{p=1}^{3}\mu_{N(p)}\left(\frac{2}{f^2_{(p)}}e^{\gamma}-1\right)dr\right\}\right\} \qquad [7.3.3]\\ &= 1. \qquad [7.3.4]\end{aligned}$$

Here we have used the results of [7.3.1] and [7.3.2] and the fact that the velocities are normalised. Consequently,

$$\lim_{L/cT\to 0} f_{(p)} = 1. \qquad [7.3.5]$$

If we return to the geodesic equation, [7.2.10], and consider in the first instance the limit of the right hand side, then using [7.3.1]–[7.3.5],

$$\begin{aligned}
\lim_{L/cT\to 0} \frac{d}{dt_N}\left(\frac{1}{f_{(p)}}v_{(p)}\right) &= \lim_{L/cT\to 0}\left\{\frac{c^2T^2}{L^2}\frac{1}{2r_N}f_{(p)}\left(e^{-\alpha}-1\right)\right. \\
&\quad -\frac{\kappa GM}{c^2L}\frac{r_N}{2}e^{\alpha+\gamma}\frac{1}{f_{(p)}}\left\{\sum_{q=1}^{3}\mu_{N(q)}\frac{1}{f_{(q)}^2}v_{(q)}v_{(q)}\right. \\
&\quad \left.\left.+2v_{(p)}\sum_{q=1}^{3}\mu_{N(q)}\frac{1}{f_{(q)}^2}v_{(q)}+v_{(p)}v_{(p)}\sum_{q=1}^{3}\mu_{N(q)}\frac{1}{f_{(q)}^2}\right\}\right\} \\
&= \lim_{L/cT\to 0}\frac{c^2T^2}{L^2}\frac{1}{2r_N}\left(e^{-\alpha}-1\right), \\
&= \lim_{L/cT\to 0}\frac{c^2T^2}{L^2}\frac{1}{2r_N}\left\{-\frac{\kappa GM}{c^2L}\frac{1}{r_N}\int_0^{r_N} r^2\sum_{q=1}^{3}\mu_{N(q)}\frac{1}{f_{(q)}^2}e^{\gamma}\,dr\right\} \\
&= -\frac{\kappa GMT^2}{L^3}\frac{1}{2r_N^2}\int_0^{r_N} r^2\sum_{q=1}^{3}\mu_{N(q)}\,dr. \qquad [7.3.6]
\end{aligned}$$

We now expand the left hand side and consider how these terms behave as $L/cT \longrightarrow 0$. We have

$$\lim_{L/cT\to 0}\left\{\frac{1}{f_{(p)}}\frac{dv_{(p)}}{dt_N}-\frac{1}{f_{(p)}^2}v_{(p)}\frac{df_{(p)}}{dt_N}\right\} = -\frac{\kappa GMT^2}{L^3}\frac{1}{2r_N^2}\int_0^{r_N} r^2\sum_{q=1}^{3}\mu_{N(q)}\,dr$$

$$\begin{aligned}
\Longrightarrow \lim_{L/cT\to 0}&\left\{\frac{1}{f_{(p)}}\frac{dv_{(p)}}{dt_N}-\frac{1}{f_{(p)}^2}v_{(p)}\frac{\partial f_{(p)}}{\partial t_N}-\frac{1}{f_{(p)}^2}v_{(p)}v_{(p)}\frac{\partial f_{(p)}}{\partial r_{N(p)}}\right\} \\
&= -\frac{\kappa GMT^2}{L^3}\frac{1}{2r_N^2}\int_0^{r_N} r^2\sum_{q=1}^{3}\mu_{N(q)}\,dr. \qquad [7.3.7]
\end{aligned}$$

Now from the definition of $f_{(p)}$ (equation [7.2.12]) it follows that

$$2f_{(p)}\frac{\partial f_{(p)}}{\partial t_N} = e^{\gamma}\frac{\partial\gamma}{\partial t_N}-\frac{L^2}{c^2T^2}e^{\alpha}\frac{\partial\alpha}{\partial t_N}v_{(p)}v_{(p)}-\frac{L^2}{c^2T^2}e^{\alpha}v_{(p)}\frac{\partial v_{(p)}}{\partial t_N}$$

$$\Longrightarrow \lim_{L/cT\to 0}\frac{\partial f_{(p)}}{\partial t_N} = \lim_{L/cT\to 0}\frac{\partial\gamma}{\partial t_N}-v_{(p)}v_{(p)}\lim_{L/cT\to 0}\frac{L^2}{c^2T^2}\frac{\partial\alpha}{\partial t_N}.$$

Similarly,

$$\lim_{L/cT\to 0}\frac{\partial f_{(p)}}{\partial r_{N(p)}} = \lim_{L/cT\to 0}\frac{\partial\gamma}{\partial r_{N(p)}}-v_{(p)}v_{(p)}\lim_{L/cT\to 0}\frac{L^2}{c^2T^2}\frac{\partial\alpha}{\partial r_{N(p)}}. \qquad [7.3.8]$$

From [7.2.7] it follows that

$$\lim_{L/cT\to 0}\frac{\partial\gamma}{\partial r_N}=\lim_{L/cT\to 0}\left\{\frac{\kappa GLM}{c^4T^2}r_Ne^{\alpha}\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f_{(p)}^2}v_{(p)}v_{(p)}e^{\alpha}-\frac{1}{r_N}+\frac{e^{\alpha}}{r_N}\right\}$$
$$=0, \qquad [7.3.9]$$

using the result of [7.3.2]. From [7.2.8] and again using [7.3.2] we have

$$\lim_{L/cT\to 0}\frac{\partial\alpha}{\partial r_N}=\lim_{L/cT\to 0}\left\{\frac{\kappa GM}{c^2L}r_Ne^{\alpha}\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f_{(p)}^2}e^{\gamma}+\frac{1}{r_N}-\frac{e^{\alpha}}{r_N}\right\}$$
$$=0, \qquad [7.3.10]$$

so that the right hand side of equation [7.3.8] vanishes, i.e.

$$\lim_{L/cT\to 0}\frac{\partial f_{(p)}}{\partial r_{N(p)}}=0. \qquad [7.3.11]$$

The limit of the total time derivative of $f_{(p)}$ is a little more complicated because there is no component of the Einstein equation that conveniently provides us with an expression for $\partial\gamma/\partial t$. To calculate this quantity we resort to differentiating the non-dimensional form of equation [6.2.8] with respect to t_N by brute force. Before we do this, however, we quickly calculate $\lim_{L/cT\to 0}\partial\alpha/\partial t_N$ as this will be needed. This information is given by [7.2.9]. Thus,

$$\lim_{L/cT\to 0}\frac{\partial\alpha}{\partial t_N}=\lim_{L/cT\to 0}\left\{\frac{\kappa GM}{c^2L}r_Ne^{\alpha}\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f_{(p)}^2}v_{(p)}e^{\gamma}\right\}$$
$$=0. \qquad [7.3.12]$$

We now calculate $\lim_{L/ct\to 0}\partial\gamma/\partial t_N$. To begin, we integrate equation [7.2.7] with respect to r_N to obtain

$$\gamma=\int_0^{r_N}\left\{\frac{e^{\alpha}-1}{r}-\frac{\kappa GLM}{c^4T^2}re^{\alpha}\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f_{(p)}^2}v_{(p)}v_{(p)}e^{\alpha}\right\}dr.$$

This is essentially the non-dimensional form of [6.2.8] spoken of earlier. If we differentiate this with respect to time then

$$\lim_{L/cT\to 0}\frac{\partial\gamma}{\partial t_N}=\lim_{L/cT\to 0}\int_0^{r_N}\left\{\frac{e^{\alpha}}{r}-\frac{\kappa GLM}{c^4T^2}re^{\alpha}\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f_{(p)}^2}v_{(p)}v_{(p)}e^{\alpha}\right\}\frac{\partial\alpha}{\partial t_N}dr$$

$$+\lim_{L/cT\to 0}\int_0^{r_N}\left\{\frac{\kappa GML}{c^4T^2}re^\alpha\frac{\partial}{\partial t_N}\left\{\sum_{p=1}^{3}\mu_{N(p)}\frac{1}{f^2_{(p)}}v_{(p)}v_{(p)}e^\alpha\right\}\right\}dr,$$

and since the first term is zero by [7.3.12] and the fact that we have a L^4/c^4T^4 factor multiplying the second part of the integrand, we obtain

$$\lim_{L/cT\to 0}\frac{\partial\gamma}{\partial t_N}=\lim_{L/cT\to 0}\int_0^{r_N}\frac{\kappa GM}{c^2L}re^\alpha\left\{\sum_{p=1}^{3}\frac{\partial\mu_{N(p)}}{\partial t_N}\left(\frac{1}{f^2_{(p)}}e^\gamma-1\right)\right\}dr$$
$$+\lim_{L/cT\to 0}\int_0^{r_N}\frac{\kappa GML}{c^4T^2}re^\alpha\left\{\sum_{p=1}^{3}\mu_{N(p)}2\left(\frac{1}{f_{(p)}}v_{(p)}\right)\frac{\partial}{\partial t_N}\left(\frac{1}{f_{(p)}}v_{(p)}\right)e^\alpha\right\}dr$$
$$+\lim_{L/cT\to 0}\int_0^{r_N}\frac{\kappa GM}{c^2L}re^\alpha\left\{\sum_{p=1}^{3}\mu_{N(p)}\left(\frac{1}{f^2_{(p)}}e^\gamma-1\right)\frac{\partial\alpha}{\partial t_N}\right\}dr.$$

Here we have used equation [7.2.12] to make the substitution,

$$\frac{L^2}{c^2T^2}\frac{1}{f^2_{(p)}}v_{(p)}v_{(p)}e^\alpha=\frac{1}{f^2_{(p)}}e^\gamma-1.$$

The first and last terms of the equation for $\lim_{L/cT\to 0}\partial\gamma/\partial t_N$ are both zero due to the L^2/c^2T^2 found within the GM/c^2L factor as well as by equations [7.3.1] and [7.3.12]. It follows then that

$$\lim_{L/cT\to 0}\frac{\partial\gamma}{\partial t_N}$$
$$=\lim_{L/cT\to 0}\int_0^{r_N}\frac{\kappa GML}{c^4T^2}re^\alpha\left\{\sum_{p=1}^{3}\mu_{N(p)}2\left(\frac{1}{f_{(p)}}v_{(p)}\right)\frac{\partial}{\partial t_N}\left(\frac{1}{f_{(p)}}v_{(p)}\right)e^\alpha\right\}dr.$$

This in turn implies that

$$\lim_{L/cT\to 0}\frac{\partial\gamma}{\partial t_N}$$
$$=\lim_{L/cT\to 0}\int_0^{r_N}\frac{\kappa GML}{c^4T^2}re^\alpha\left\{\sum_{p=1}^{3}\mu_{N(p)}2\left(\frac{1}{f_{(p)}}v_{(p)}\right)\frac{d}{dt_N}\left(\frac{1}{f_{(p)}}v_{(p)}\right)\right\}$$
$$-\lim_{L/cT\to 0}\int_0^{r_N}\frac{\kappa GML}{c^4T^2}re^\alpha\left\{\sum_{p=1}^{3}\mu_{N(p)}2\left(\frac{1}{f_{(p)}}v_{(p)}\right)v_{(p)}\frac{\partial}{\partial r_{(p)}}\left(\frac{1}{f_{(p)}}v_{(p)}\right)e^\alpha\right\}dr$$
$$=\lim_{L/cT\to 0}\int_0^{r_N}\frac{\kappa GML}{c^4T^2}re^\alpha\left\{\sum_{p=1}^{3}\mu_{N(p)}2\left(\frac{1}{f_{(p)}}v_{(p)}\right)\frac{d}{dt_N}\left(\frac{1}{f_{(p)}}v_{(p)}\right)\right\}$$

$$- \lim_{L/cT \to 0} \int_0^{r_N} \frac{\kappa GML}{c^4T^2} r e^{\alpha} \left\{ \sum_{p=1}^{3} \mu_{N(p)} 2 \left(\frac{1}{f_{(p)}} v_{(p)} \right) v_{(p)} \frac{1}{f_{(p)}} \frac{\partial v_{(p)}}{\partial r_{(p)}} e^{\alpha} \right\} dr$$

$$+ \lim_{L/cT \to 0} \int_0^{r_N} \frac{\kappa GML}{c^4T^2} r e^{\alpha} \left\{ \sum_{p=1}^{3} \mu_{N(p)} 2 \left(\frac{1}{f_{(p)}} v_{(p)} \right) v_{(p)} \frac{1}{f_{(p)}^2} v_{(p)} \frac{\partial f_{(p)}}{\partial r_{(p)}} e^{\alpha} \right\} dr.$$

The first term must be zero since by the geodesic equation ([7.3.6]) the limit of the term, $d\left(v_{(p)} f_{(p)}^{-1}\right)/dt_N$, is finite allowing the L^4/c^4T^4 factor within the GML/c^4T^2 group to dominate. For similar reasons ($\partial v_{(p)}/\partial r_{(p)}$, by assumption, is finite and $\lim_{L/cT \to 0} \partial f_{(p)}/\partial r_{(p)} = 0$) the second and third terms are zero and so finally we have

$$\lim_{L/cT \to 0} \frac{\partial \gamma}{\partial t_N} = 0, \qquad [7.3.13]$$

which is down to the L^4/c^4T^4 factor dominating. This, along with [7.3.12], implies that

$$\lim_{L/cT \to 0} \frac{\partial f_{(p)}}{\partial t_N} = 0, \qquad [7.3.14]$$

which when coupled with [7.3.11] implies that the limit of the geodesic equation ([7.3.7]) is of the form,

$$\lim_{L/cT \to 0} \frac{1}{f_{(p)}} \frac{dv_{(p)}}{dt_N} = -\frac{\kappa GMT^2}{L^3} \frac{1}{2r_N^2} \int_0^{r_N} r^2 \sum_{q=1}^{3} \mu_{N(q)}\, dr.$$

If we allow our variables to re-absorb the constants that provide our units, i.e. L, T and M then since $\kappa = 8\pi$, we obtain

$$\frac{d^2 r_{(p)}}{dt^2} = -\frac{4\pi G}{r^2} \int_0^{r} s^2 \sum_{q=1}^{3} \mu_{(q)}\, ds,$$

which is exactly the Newtonian force equation written in spherical coordinates.

§7.4. Newtonian limit of $T^{ij}_{(p);j} = 0$.

To determine the limit of this equation we can make use of the many results obtained in the previous section. The non-dimensional form of $T^{ij}_{(p);j} = 0$ is given by

$$\frac{1}{\sqrt{-g_N}} \frac{\partial}{\partial t_N} \left(\mu_{N(p)} \sqrt{-g_N} \frac{1}{f_{(p)}} \right) + \frac{1}{\sqrt{-g_N}} \frac{\partial}{\partial r_N} \left(\mu_{N(p)} \sqrt{-g_N} \frac{v_{(p)}}{f_{(p)}} \right) = 0, \qquad [7.2.11]$$

and expanding this gives

$$\frac{\partial \mu_{N(p)}}{\partial t_N}\frac{1}{f_{(p)}}+\frac{1}{2g_N}\mu_{N(p)}\frac{\partial g_N}{\partial t_N}\frac{1}{f_{(p)}}-\frac{\mu_{N(p)}}{f_{(p)}^2}\frac{\partial f_{(p)}}{\partial t}+\frac{\partial \mu_{N(p)}}{\partial r_N}\frac{v_{(p)}}{f_{(p)}}+\frac{1}{2g_N}\mu_{N(p)}\frac{\partial g_N}{\partial r_N}\frac{v_{(p)}}{f_{(p)}}$$
$$+\mu_{N(p)}\frac{\partial v_{(p)}}{\partial r_N}\frac{1}{f_{(p)}}-\mu_{N(p)}\frac{v_{(p)}}{f_{(p)}^2}\frac{\partial f_{(p)}}{\partial r_N}=0,$$

$$\Longrightarrow \frac{\partial \mu_{N(p)}}{\partial t_N}\frac{1}{f_{(p)}}+\frac{1}{2}\frac{\mu_{N(p)}}{f_{(p)}}\left(\frac{\partial \alpha}{\partial t_N}+\frac{\partial \gamma}{\partial t_N}\right)-\frac{\mu_{N(p)}}{f_{(p)}^2}\frac{\partial f_{(p)}}{\partial t_N}+\frac{\partial \mu_{N(p)}}{\partial r_N}\frac{v_{(p)}}{f_{(p)}}$$
$$+\frac{2}{r_N}\mu_{N(p)}\frac{v_{(p)}}{f_{(p)}}+\frac{1}{2}\mu_{N(p)}\frac{v_{(p)}}{f_{(p)}}\left(\frac{\partial \alpha}{\partial r_N}+\frac{\partial \gamma}{\partial r_N}\right)$$
$$+\mu_{N(p)}\frac{\partial v_{(p)}}{\partial r_N}\frac{1}{f_{(p)}}-\mu_{N(p)}\frac{v_{(p)}}{f_{(p)}^2}\frac{\partial f_{(p)}}{\partial r_N}=0,$$

where we have that $g_N=-r^4\sin^2\theta e^{\alpha}e^{\gamma}$. Taking the limit of this equation as $L/cT\longrightarrow 0$ is easy for we can simply appeal to equations [7.3.9], [7.3.10], [7.3.11], [7.3.12], [7.3.13] and [7.3.14] to set the second, third, sixth and eighth terms equal to zero. Thus we have

$$\frac{\partial \mu_{N(p)}}{\partial r_N}v_{(p)}+\frac{2\mu_{N(p)}v_{(p)}}{r_N}+\mu_{N(p)}\frac{\partial v_{(p)}}{\partial r_N}+\frac{\partial \mu_{N(p)}}{\partial t_N}=0,$$

which again is precisely equivalent to the Newtonian conservation of matter equation written in spherical coordinates. Thus we conclude that the differential equations describing a spherically symmetric, General Relativistic formulation of caustic formation reduce to the corresponding Newtonian equations in the limit of small velocities (i.e. as $L/cT\longrightarrow 0$).

§7.5. Asymptotic behaviour of geodesic equation.

The conclusion at the end of the last section is an important result as it reassures us that our model is the correct one for caustic formation in General Relativity. In this section we consider a different limiting process. That is to say, using the concept of asymptotic solutions introduced in §5.2, we shall construct a coordinate system whose length scale increases unboundedly as we allow a parameter, ε, to tend to infinity. This technique describes a magnification type of process and enables us to probe a small area containing the cusp so that we can determine the essential physics of cusp formation. We have already explained the reason why we are doing this but since it is an extremely significant calculation, we shall reiterate. In chapter 5 we demonstrated that under a similar asymptotic analysis, the spherically symmetric Newtonian equations of motion reduced to those of the gravity free case formulated such that the solutions exhibit planar symmetry. This is an important result for it allows us to conclude that it is only the boundary conditions that determine whether or not a cusp is formed. The effect of gravity is to simply shape the caustic. If we perform the same asymptotic analysis based on the General Relativistic equations of motion then we can compare the two results. Of course the hope is that by taking the asymptotic limit ($\varepsilon \longrightarrow \infty$) and then the Newtonian limit ($L/cT \longrightarrow 0$) we should obtain the gravity free differential equations and thus confirm the conclusion of §5.5.

In previous chapters we have developed two different, but similar formulations for the General Relativistic problem. That is to say we have the tetrad formalism and the coordinate basis formalism. The first was developed during the discussion on the development of the computer program, the second was used exclusively during the calculation of the Newtonian limit. In this section we shall again resort to using the tetrad formalism since we wish to use the useful feature that it possesses, namely the lack of metric components within the mixed form for the energy-momentum tensor.

We shall begin this asymptotic analysis by considering the geodesic equation and making the transformation, $r = x + r_c$. This illustrates the fact that we intend to convert all equations in the General Relativistic picture so that the origin for the radial coordinate now occurs at some finite distance, r_c. Moreover, we assume that the point, $x = 0$, $t = 0$, corresponds to the point where the cusp of the caustic initially forms. A family of curves, parameterised by ε, can now be constructed using the transformation functions introduced in §4.2. We choose a fixed point $\mathbf{x} \in M$ and define a particular curve by $\tilde{\mathbf{x}} = g_{\mathbf{x}}(\varepsilon; \mathbf{x})$. We then model the magnification process mentioned above by stipulating that relative to $\tilde{t}\tilde{x}$ coordinates, $\lim_{\varepsilon\to\infty} |\mathbf{x}| = \lim_{\varepsilon\to\infty} \left|g_{\mathbf{x}}\left(\varepsilon^{-1}; \tilde{\mathbf{x}}\right)\right| = 0$. Since we shall choose $g_{\mathbf{x}}(\varepsilon; t, x) = (\varepsilon^{k_t} t, \varepsilon^{k_x} x)$, these curves become defined by $x/t^{\beta} = \text{const}$ where $\beta = k_x/k_t$. To complete the picture we assume that as the length and time scales of our $\tilde{t}\tilde{x}$ coordinate system increase, so do the mass descriptors. In other words we choose the general transformation group to be

$$g\left(\varepsilon; t, x_{(p)}, u^1_{(p)}, \mu_{(p)}, G, c\right) = \left(\varepsilon^{k_t} t, \varepsilon^{k_x} x_{(p)}, \varepsilon^{k_u} u^1_{(p)}, \varepsilon^{k_\mu} \mu_{(p)}, \varepsilon^{k_G} G, \varepsilon^{-k_u} c\right), \quad [7.5.1]$$

which, we can see, has essentially been lifted from the equivalent Newtonian analysis of §5.5. In order to compare the results here with those of §5.5, we shall furthermore assume the same similarity degrees. That is to say, we suppose that if $k_x = \beta k_t$ then $k_u = (\beta - 1)k_t$, $k_\mu = -(2+\gamma)k_t$ and $k_G = \gamma k_t$ for $\beta = 3/2$ and $\gamma < 0$. Finally, we note that the constant, k_t, becomes important for it determines whether g defines a magnifying or reducing process.

It is noticeable that in [7.5.1] all velocities are scaled. The reason for this is quite simply to make it work. In our first attempt at looking for asymptotic solutions the velocity of light was not scaled and the result was that we could not make the derivatives of the metric components exist in the limit. This was an undesirable feature for this, as we shall see, implies that the geodesic equation does not have a limiting form. To solve this problem we found it necessary to scale c and we do so in the opposite manner to all other velocities. That is to say, with $k_u > 0$ (a relation arising from the fact that we must require $k_t > 0$) $\varepsilon^{k_u} c$ increases indefinitely as $\varepsilon = 0$ is approached. The only physical interpretation of this is that in Newtonian theory there is no upper limit for the magnitude of any velocity vector. It seems, therefore, that by choosing to scale c we might somehow be coupling together the $\varepsilon \longrightarrow \infty$ and $L/cT \longrightarrow 0$ limiting processes. This rather muddies the implication in the first paragraph of this section that the two limits are distinct. It might be that our asymptotic solutions defined by [7.5.1] will be equivalent to the planar symmetric Newtonian solutions encountered earlier in chapter 4.

The last thing we must mention before starting with our asymptotic analysis is that we require our metric components to be continuous. The reason for this is that the curves predefined by the above group of transformations must, at some stage, cross the caustic. If we did not make this restriction on g_{ij}, then from the geodesic equation the r-component of $d\tilde{u}^i_{(p)}/d\tilde{t}$ possesses an impulsive part implying that $\tilde{u}^1_{(p)}$ would be discontinuous. It follows then that the limiting process cannot be constructed because the velocity at some point on the curve (where it crosses the caustic) becomes undefined meaning that as $\tilde{\mathbf{x}}$ approaches the cusp, it is unknown which value for $\tilde{u}^1_{(p)}$ to take when we come to scale it accordingly.

Let us now consider the behaviour of the modelling equations ([6.1.1]–[6.1.3]) under the transformation given by [7.5.1]. We start, in this section, with the geodesic equation,

$$\frac{du^1_{(p)}}{dt} = -\frac{1}{2}e^{(\alpha+\gamma)/2}\Big\{c(x+r_c)^{-1}\sqrt{c^2+u^1_{(p)}u^1_{(p)}}\left(1-e^{-\alpha}\right)$$
$$-\frac{\kappa G}{c^3}(x+r_c)\,\mu_{(q)}u^1_{(q)}\left(u^1_{(p)}\sqrt{c^2+u^1_{(q)}u^1_{(q)}}-u^1_{(q)}\sqrt{c^2+u^1_{(p)}u^1_{(p)}}\right)$$
$$-\frac{\kappa G}{c^3}(x+r_c)\,\mu_{(r)}u^1_{(r)}\left(u^1_{(p)}\sqrt{c^2+u^1_{(r)}u^1_{(r)}}-u^1_{(r)}\sqrt{c^2+u^1_{(p)}u^1_{(p)}}\right)\Big\}, \qquad [7.5.2]$$

written in terms of our fixed coordinate system, **x**. To determine the asymptotics of our system, i.e. to determine how the above equation changes as **x** moves towards the origin in $\tilde{t}\tilde{x}$ space, we need to relate all variables, f, to their value at $\tilde{\mathbf{x}}$. For those functions whose dependence on ε is not absolutely specified by [7.5.1], we suppose that they transform by virtue of the quantities of which they are functions of. This means that the dependency of $e^{-\alpha}$ on ε, for example, can be calculated since,

$$e^{-\alpha} = 1+\frac{\kappa G}{c^4}\frac{1}{x+r_c}\int_0^{x+r_c} r^2\sum_{p=1}^{3}\mu_{(p)}\left(c^2+u^1_{(p)}u^1_{(p)}\right)dr$$
$$= 1+\frac{\kappa\tilde{G}}{\tilde{c}^4}\frac{\varepsilon^{k_x-k_G-k_\mu-2k_u}}{\tilde{x}+\varepsilon^{k_x}r_c}\int_0^{\varepsilon^{-k_x}\tilde{x}+r_c} r^2\sum_{p=1}^{3}\tilde{\mu}_{(p)}\left(\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right)dr.$$

In order to take the limit of this equation, we have to be sure of the dependence on ε for each factor. In the above, the exponent of ε is given by $k_x-k_G-k_\mu-2k_u = (4-\beta)k_t$ which is positive, implying that this term at least becomes very large as ε increases. If this were indicative of $e^{-\alpha}$ as a whole then this is likely to be bad news for it is implying that the metric is unbounded in the limit contradicting the result from the Newtonian asymptotic analysis. Clearly then we need to be sure of how the integral behaves as ε increases. By making the transformation, $s=\varepsilon^{k_x}r$, we obtain

$$e^{-\alpha} = 1+\frac{\kappa\tilde{G}}{\tilde{c}^4}\frac{\varepsilon^{-2k_x-k_G-k_\mu-2k_u}}{\tilde{x}+\varepsilon^{k_x}r_c}\int_0^{\tilde{x}+\varepsilon^{k_x}r_c} s^2\sum_{p=1}^{3}\tilde{\mu}_{(p)}\left(\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right)ds. \qquad [7.5.3]$$

Since $k_x = \beta k_t > 0$, we have that as $\varepsilon \longrightarrow \infty$, the above integral in some sense represents the total mass in our spacetime. We suppose that this is finite and since $-2k_x-k_G-k_\mu-2k_u = 4(1-\beta)k_t < 0$, we have

$$\lim_{\varepsilon\to\infty} e^{-\alpha} = 1.$$

Next we can look at the time component of the metric. We have from [6.3.7]

$$
\begin{aligned}
e^{\gamma} &= e^{-\alpha}\exp\left\{\frac{\kappa G}{c^4}\int_0^{x+r_c} re^{\alpha}\sum_{p=1}^{3}\mu_{(p)}\left(c^2+2u^1_{(p)}u^1_{(p)}\right)dr\right\} \\
&= e^{-\alpha}\exp\left\{\varepsilon^{-2k_x-k_G-k_\mu-2k_u}\frac{\kappa\tilde{G}}{\tilde{c}^4}\int_0^{\tilde{x}+\varepsilon^{k_x}r_c} se^{\alpha}\sum_{p=1}^{3}\tilde{\mu}_{(p)}\left(\tilde{c}^2+2\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right)ds\right\},
\end{aligned}
$$

which by a similar argument to that for $e^{-\alpha}$ implies that

$$
\lim_{\varepsilon\longrightarrow\infty} e^{\gamma} = 1.
$$

We conclude, therefore, that in the limit the metric is Minkowskian. This may seem an odd result considering the fact that in §6.2 we show that the metric components at a point $(t, r) \in M$ are functions of the total mass enclosed by the shell of radius r. It turns out, however, that the above limiting process requires that the cusp position be rescaled according to $\varepsilon^{k_x} r_c$ for each ε. By assumption, $k_x > 0$ and so it follows that in the limit as $\varepsilon \longrightarrow \infty$, the position of the cusp increases resulting in a situation that locally looks more and more planar. In this kind of scenario we can assume that the gravitational force is zero corresponding to a flat metric as proved.

Let us now return to the discussion on the ε-dependence of equation [7.5.2]. Clearly, the behaviour of all terms on the right hand side of the geodesic equation has been determined. However, before we can consider the limiting form for the geodesic equation, we need to investigate how $du^1_{(p)}/dt$ transforms under g. Since this quantity is essentially the acceleration along a geodesic, it must somehow be related to $d^2x_{(p)}/dt^2$. We assume that this quantity must survive as $\varepsilon \longrightarrow \infty$ for otherwise we do not anticipate a sensible result. Now,

$$
\begin{aligned}
\frac{du^1_{(p)}}{dt} &= \frac{d}{dt}\left(v^1_{(p)}e^{\alpha/2}\right) \\
&= \frac{d}{dt}\left(\frac{dr_{(p)}}{dt}v^4_{(p)}e^{\alpha/2}\right) \\
&= \frac{d}{dt}\left(\frac{dr_{(p)}}{dt}u^4_{(p)}e^{(\alpha-\gamma)/2}\right) \\
&= \frac{d^2r_{(p)}}{dt^2}u^4_{(p)}e^{(\alpha-\gamma)/2} + \frac{dr_{(p)}}{dt}\frac{du^4_{(p)}}{dt}e^{(\alpha-\gamma)/2} + \frac{1}{2}\frac{dr_{(p)}}{dt}u^4_{(p)}e^{(\alpha-\gamma)/2}\left(\frac{d\alpha}{dt}-\frac{d\gamma}{dt}\right) \\
&= \frac{d^2r_{(p)}}{dt^2}u^4_{(p)}e^{(\alpha-\gamma)/2} + \frac{u^1_{(p)}}{u^4_{(p)}}\frac{du^4_{(p)}}{dt} + \frac{u^1_{(p)}}{2}\left(\frac{d\alpha}{dt}-\frac{d\gamma}{dt}\right)
\end{aligned}
$$

$$= \frac{d^2 r_{(p)}}{dt^2} u^4_{(p)} e^{(\alpha-\gamma)/2} + \frac{1}{c^2} \frac{u^1_{(p)} u^1_{(p)}}{u^4_{(p)} u^4_{(p)}} \frac{du^1_{(p)}}{dt} + \frac{u^1_{(p)}}{2} \left(\frac{d\alpha}{dt} - \frac{d\gamma}{dt} \right),$$

where in the last step we have used $du^4_{(p)}/dt = u^1_{(p)}/c^3 u^4_{(p)} \cdot du^1_{(p)}/dt$; this is obtainable by differentiating the normalisation condition. Thus, in terms of our fixed coordinate system we finally obtain

$$\frac{du^1_{(p)}}{dt} = \frac{1}{c^3} \left(c^2 + u^1_{(p)} u^1_{(p)} \right)^{3/2} \frac{d^2 r_{(p)}}{dt^2} e^{(\alpha-\gamma)/2} + \frac{1}{2c^2} u^1_{(p)} \left(c^2 + u^1_{(p)} u^1_{(p)} \right) \left(\frac{d\alpha}{dt} - \frac{d\gamma}{dt} \right). \tag{7.5.4}$$

We now combine equations [7.5.2] and [7.5.4]. This gives us

$$\begin{aligned} \frac{d^2 x_{(p)}}{dt^2} = & -\frac{c}{2} u^1_{(p)} e^{(\gamma-\alpha)/2} \left(c^2 + u^1_{(p)} u^1_{(p)} \right)^{-1/2} \left(\frac{d\alpha}{dt} - \frac{d\gamma}{dt} \right) \\ & - \frac{c^4}{2} e^{\gamma} (x + r_c)^{-1} \left(c^2 + u^1_{(p)} u^1_{(p)} \right)^{-1} \left(1 - e^{-\alpha} \right) \\ & + \frac{1}{2} \kappa G \left(c^2 + u^1_{(p)} u^1_{(p)} \right)^{-3/2} (x + r_c) \\ & \quad \times e^{\gamma} \mu_{(q)} u^1_{(q)} \left(u^1_{(p)} \sqrt{c^2 + u^1_{(q)} u^1_{(q)}} - u^1_{(q)} \sqrt{c^2 + u^1_{(p)} u^1_{(p)}} \right) \\ & + \frac{1}{2} \kappa G \left(c^2 + u^1_{(p)} u^1_{(p)} \right)^{-3/2} (x + r_c) \\ & \quad \times e^{\gamma} \mu_{(r)} u^1_{(r)} \left(u^1_{(p)} \sqrt{c^2 + u^1_{(r)} u^1_{(r)}} - u^1_{(r)} \sqrt{c^2 + u^1_{(p)} u^1_{(p)}} \right), \end{aligned}$$

which in terms of our $\tilde{t}\tilde{x}$ coordinates becomes

$$\begin{aligned} \frac{d^2 \tilde{x}_{(p)}}{d\tilde{t}^2} = & -\frac{\tilde{c}}{2} \varepsilon^{k_x - k_u - k_t} \tilde{u}^1_{(p)} e^{(\gamma-\alpha)/2} \left(\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)} \right)^{-1/2} \left(\frac{d\alpha}{d\tilde{t}} - \frac{d\gamma}{d\tilde{t}} \right) \\ & - \frac{\tilde{c}^4}{2} \varepsilon^{k_x - 2k_t + 2k_u} e^{\gamma} \left(\varepsilon^{-k_x} \tilde{x} + r_c \right)^{-1} \left(\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)} \right)^{-1} \left(1 - e^{-\alpha} \right) \\ & + \frac{1}{2} \kappa \tilde{G} \varepsilon^{k_x - k_G - k_\mu - 4k_u - 2k_t} \left(\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)} \right)^{-3/2} \left(\varepsilon^{-k_x} \tilde{x} + r_c \right) \\ & \quad \times e^{\gamma} \tilde{\mu}_{(q)} \tilde{u}^1_{(q)} \left(\tilde{u}^1_{(p)} \sqrt{\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(q)} \tilde{u}^1_{(q)}} - \tilde{u}^1_{(q)} \sqrt{\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)}} \right) \\ & + \frac{1}{2} \kappa \tilde{G} \varepsilon^{k_x - k_G - k_\mu - 4k_u - 2k_t} \left(\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)} \right)^{-3/2} \left(\varepsilon^{-k_x} \tilde{x} + r_c \right) \\ & \quad \times e^{\gamma} \tilde{\mu}_{(r)} \tilde{u}^1_{(r)} \left(\tilde{u}^1_{(p)} \sqrt{\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(r)} \tilde{u}^1_{(r)}} - \tilde{u}^1_{(r)} \sqrt{\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)}} \right). \end{aligned}$$

We can now take the limit as ε tends to infinity. We note that $k_x - k_u - k_t = 0$, $k_x - 2k_t + 2k_u = (3\beta - 4)k_t > 0$ and $k_x - k_G - k_\mu - 4k_u - 2k_t = (4 - 3\beta)k_t < 0$

so clearly, the last two terms tend towards zero. The first two terms require more consideration and so we write,

$$\frac{d^2\tilde{x}_{(p)}}{d\tilde{t}^2} = -\frac{1}{2}\tilde{u}^1_{(p)} \lim_{\varepsilon\to\infty}\left(\frac{d\alpha}{d\tilde{t}} - \frac{d\gamma}{d\tilde{t}}\right) - \frac{\tilde{c}^2}{2r_c}\lim_{\varepsilon\to\infty}\left(\varepsilon^{k_x-2k_t+2k_u}\left(1-e^{-\alpha}\right)\right).$$

For the second term on the right hand side of the above equation we have

$$\begin{aligned}
&\lim_{\varepsilon\to\infty}\left(\varepsilon^{k_x-2k_t+2k_u}\left(1-e^{-\alpha}\right)\right)\\
&= \lim_{\varepsilon\to\infty}\frac{\kappa\tilde{G}}{\tilde{c}^4}\frac{\varepsilon^{-k_x-k_G-k_\mu-2k_t}}{(\tilde{x}+\varepsilon^{k_x}r_c)}\int_0^{\tilde{x}+\varepsilon^{k_x}r_c} s^2\sum_{p=1}^{3}\tilde{\mu}_{(p)}\left(\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right)ds\\
&= 0,
\end{aligned}$$

and thus to complete the analysis we need to determine the dependence on ε of the derivatives of α and γ.

To find this information we resort to Einstein's equation. This gives the ε-dependence for $\partial\alpha/\partial\tilde{x}$, $\partial\alpha/\partial\tilde{t}$ and $\partial\gamma/\partial\tilde{x}$ immediately, but for $\partial\gamma/\partial\tilde{t}$ it is a little more complicated. From [6.2.5] we have,

$$\frac{\partial\alpha}{\partial t} = -\frac{\kappa G}{c^3}(x+r_c)\,e^{(\alpha+\gamma)/2}\sum_{p=1}^{3}\mu_{(p)}u^1_{(p)}\sqrt{c^2+u^1_{(p)}u^1_{(p)}}$$

$$\Longrightarrow \frac{\partial\alpha}{\partial\tilde{t}} = -\frac{\kappa\tilde{G}}{\tilde{c}^3}\varepsilon^{-k_G-k_\mu-3k_u-k_t}\left(\varepsilon^{-k_x}\tilde{x}+r_c\right)e^{(\alpha+\gamma)/2}\sum_{p=1}^{3}\tilde{\mu}_{(p)}\tilde{u}^1_{(p)}\sqrt{\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}}$$

$$\Longrightarrow \lim_{\varepsilon\to\infty}\frac{\partial\alpha}{\partial\tilde{t}} = 0 \qquad [7.5.5]$$

(since $-k_G-k_\mu-3k_u-k_t = (4-3\beta)k_t < 0$). Similarly [6.2.4] gives,

$$\frac{\partial\alpha}{\partial x} = \frac{1-e^\alpha}{x+r_c} + \frac{\kappa G}{c^4}(x+r_c)\,e^\alpha\sum_{p=1}^{3}\mu_{(p)}\left(c^2+u^1_{(p)}u^1_{(p)}\right)$$

$$\begin{aligned}
\Longrightarrow \frac{\partial\alpha}{\partial\tilde{x}} &= \frac{1-e^\alpha}{\tilde{x}+\varepsilon^{k_x}r_c}\\
&\quad+\frac{\kappa\tilde{G}}{\tilde{c}^4}\varepsilon^{-k_x-k_G-k_\mu-2k_u}\left(\varepsilon^{-k_x}\tilde{x}+r_c\right)e^\alpha\sum_{p=1}^{3}\tilde{\mu}_{(p)}\left(\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right).
\end{aligned} \qquad [7.5.6]$$

To determine the limit of the right hand side of this equation we need to investigate how $(1-e^{\alpha})$ depends on ε. We define A_ε by

$$A_\varepsilon = \frac{\kappa\tilde{G}}{\tilde{c}^4}\frac{1}{(\tilde{x}+\varepsilon^{k_x}r_c)}\int_0^{\tilde{x}+\varepsilon^{k_x}r_c} s^2\sum_{p=1}^{3}\tilde{\mu}_{(p)}\left(\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right)ds,$$

so that $\lim_{\varepsilon\to\infty} A_\varepsilon$ exists and is constant. Then equation [7.5.3] becomes

$$e^{-\alpha} = 1+\varepsilon^{-2k_x-k_G-k_\mu-2k_u}A_\varepsilon$$

$$\Longrightarrow e^{\alpha} = \frac{1}{1+\varepsilon^{-2k_x-k_G-k_\mu-2k_u}A_\varepsilon}$$

$$\begin{aligned}\Longrightarrow e^{\alpha}-1 &= \frac{1}{1+\varepsilon^{-2k_x-k_G-k_\mu-2k_u}A_\varepsilon}-1\\ &= \frac{1-\left(1+\varepsilon^{-2k_x-k_G-k_\mu-2k_u}A_\varepsilon\right)}{1+\varepsilon^{-2k_x-k_G-k_\mu-2k_u}A_\varepsilon}\\ &= \frac{\varepsilon^{-2k_x-k_G-k_\mu-2k_u}A_\varepsilon}{1+\varepsilon^{-2k_x-k_G-k_\mu-2k_u}A_\varepsilon}\end{aligned}$$

$$\Longrightarrow \lim_{\varepsilon\to\infty}(e^{\alpha}-1)=0,$$

so that finally, using the fact that $-k_x-k_G-k_\mu-2k_u=(4-3\beta)k_t<0$, equation [7.5.6] becomes

$$\lim_{\varepsilon\to\infty}\frac{\partial\alpha}{\partial\tilde{x}}=0.$$

It now follows that

$$\lim_{\varepsilon\to\infty}\frac{d\alpha}{d\tilde{t}}=\lim_{\varepsilon\to\infty}\frac{\partial\alpha}{\partial\tilde{t}}+\lim_{\varepsilon\to\infty}\frac{\partial\tilde{x}}{\partial\tilde{t}}\frac{\partial\alpha}{\partial\tilde{x}}=0.$$

Next we consider the total time derivative for γ. From [6.2.3] we have,

$$\frac{\partial\gamma}{\partial x}=\frac{e^{\alpha}-1}{x+r_c}+\frac{\kappa G}{c^4}(x+r_c)\,e^{\alpha}\sum_{p=1}^{3}\mu_{(p)}u^1_{(p)}u^1_{(p)}$$

$$\Longrightarrow \frac{\partial\gamma}{\partial\tilde{x}}=\frac{e^{\alpha}-1}{\tilde{x}+\varepsilon^{k_x}r_c}+\frac{\kappa\tilde{G}}{\tilde{c}^4}\varepsilon^{-k_G-k_\mu-6k_u}\left(\varepsilon^{-k_x}\tilde{x}+r_c\right)e^{\alpha}\sum_{p=1}^{3}\tilde{\mu}_{(p)}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}$$

$$\Longrightarrow \lim_{\varepsilon\to\infty}\frac{\partial\gamma}{\partial x}=0,$$

since $-k_G - k_\mu - 6k_u = (8-6\beta)k_t = -k_t < 0$. Next we consider $\partial\gamma/\partial\tilde{t}$. This is a much more complicated calculation for we do not have an Einstein equation to resort to. Instead we are forced to differentiate [6.3.7] by force. We have

$$\gamma = -\alpha + \frac{\kappa G}{c^4}\int_0^{x+r_c} re^\alpha \sum_{p=1}^{3} \mu_{(p)} \left(c^2 + 2u^1_{(p)}u^1_{(p)}\right)\, dr$$

$$\begin{aligned}\implies \frac{\partial\gamma}{\partial t} = \frac{\partial\alpha}{\partial t} &+ \frac{\kappa G}{c^4}\int_0^{x+r_c} re^\alpha \frac{\partial\alpha}{\partial t}\sum_{p=1}^{3} \mu_{(p)} \left(c^2 + 2u^1_{(p)}u^1_{(p)}\right)\, dr \\ &+ \frac{\kappa G}{c^4}\int_0^{x+r_c} re^\alpha \sum_{p=1}^{3} \frac{\partial\mu_{(p)}}{\partial t} \left(c^2 + 2u^1_{(p)}u^1_{(p)}\right)\, dr \\ &+ \frac{4\kappa G}{c^4}\int_0^{x+r_c} re^\alpha \sum_{p=1}^{3} \mu_{(p)}u^1_{(p)}\frac{\partial u^1_{(p)}}{\partial t}\, dr.\end{aligned}$$

Transforming to $\tilde{t}\tilde{x}$ coordinates gives

$$\begin{aligned}\frac{\partial\gamma}{\partial\tilde{t}} = \frac{\partial\alpha}{\partial\tilde{t}} &+ \frac{\kappa\tilde{G}}{\tilde{c}^4}\varepsilon^{-k_G-k_\mu-2k_u-2k_x}\int_0^{\tilde{x}+\varepsilon^{k_x}r_c} se^\alpha\frac{\partial\alpha}{\partial\tilde{t}}\sum_{p=1}^{3}\tilde{\mu}_{(p)}\left(\tilde{c}^2 + 2\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right)\, ds \\ &+ \frac{\kappa\tilde{G}}{\tilde{c}^4}\varepsilon^{-k_G-k_\mu-2k_u-2k_x}\int_0^{\tilde{x}+\varepsilon^{k_x}r_c} se^\alpha\sum_{p=1}^{3}\frac{\partial\mu_{(p)}}{\partial\tilde{t}}\left(\tilde{c}^2 + 2\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right)\, ds \\ &+ \frac{4\kappa\tilde{G}}{\tilde{c}^4}\varepsilon^{-k_G-k_\mu-6k_u-2k_x}\int_0^{\tilde{x}+\varepsilon^{k_x}r_c} se^\alpha\sum_{p=1}^{3}\tilde{\mu}_{(p)}\tilde{u}^1_{(p)}\frac{\partial\tilde{u}^1_{(p)}}{\partial\tilde{t}}\, ds,\end{aligned}$$

so that $-k_G-k_\mu-2k_u-2k_x = 4(1-\beta)k_t < 0$ and $-k_G-k_\mu-6k_u-2k_x = 8(1-\beta)k_t < 0$ imply that

$$\lim_{\varepsilon\to\infty}\frac{\partial\gamma}{\partial\tilde{t}} = 0 \qquad [7.5.7]$$

$$\implies \lim_{\varepsilon\to\infty}\frac{d\gamma}{d\tilde{t}} = \frac{\partial\gamma}{\partial\tilde{t}} + \frac{\partial\tilde{x}}{\partial\tilde{t}}\frac{\partial\gamma}{\partial\tilde{x}} = 0.$$

Thus finally we can conclude that the asymptotic form of the geodesic equation is

$$\frac{d^2x_{(p)}}{dt^2} = 0. \qquad [7.5.8]$$

§7.6. Asymptotic behaviour of $T^{ij}_{(p);j} = 0$.

We know that written in terms of our static **x** variable, [6.1.3] implies that

$$\frac{1}{\sqrt{-g}}\left(\mu_{(p)}\sqrt{-g}v^i_{(p)}\right)_{,i} = 0.$$

If we expand this equation then we obtain

$$\frac{\partial\mu_{(p)}}{\partial r}v^1_{(p)} + \frac{1}{2g}\mu_{(p)}\frac{\partial g}{\partial r}v^1_{(p)} + \mu_{(p)}\frac{\partial v^1_{(p)}}{\partial r} + \frac{\partial\mu_{(p)}}{\partial t}v^4_{(p)} + \frac{1}{2g}\mu_{(p)}\frac{\partial g}{\partial t}v^4_{(p)} + \mu_{(p)}\frac{\partial v^4_{(p)}}{\partial t} = 0$$

$$\Longrightarrow \frac{\partial\mu_{(p)}}{\partial r}v^1_{(p)} + \frac{1}{2}\mu_{(p)}v^1_{(p)}\left(\frac{\partial\alpha}{\partial r} + \frac{\partial\gamma}{\partial r}\right) + \frac{2}{r}\mu_{(p)}v^1_{(p)} + \mu_{(p)}\frac{\partial v^1_{(p)}}{\partial r}$$
$$+ \frac{\partial\mu_{(p)}}{\partial t}v^4_{(p)} + \frac{1}{2}\mu_{(p)}v^4_{(p)}\left(\frac{\partial\alpha}{\partial t} + \frac{\partial\gamma}{\partial t}\right) + \mu_{(p)}\frac{\partial v^4_{(p)}}{\partial t} = 0,$$

recalling that $g = -r^4\sin^2\theta e^\alpha e^\gamma$. We now transform coordinates so that $r = x + r_c$. Then,

$$\frac{\partial\mu_{(p)}}{\partial x}\frac{dx_{(p)}}{dt}v^4_{(p)} + \frac{1}{2}\mu_{(p)}\frac{dx_{(p)}}{dt}v^4_{(p)}\left(\frac{\partial\alpha}{\partial x} + \frac{\partial\gamma}{\partial x}\right) + \frac{2\mu_{(p)}v^4_{(p)}}{x + r_c}\frac{dx_{(p)}}{dt} + \mu_{(p)}\frac{\partial}{\partial x}\left(\frac{dx_{(p)}}{dt}v^4_{(p)}\right)$$
$$+ \frac{\partial\mu_{(p)}}{\partial t}v^4_{(p)} + \frac{1}{2}\mu_{(p)}v^4_{(p)}\left(\frac{\partial\alpha}{\partial t} + \frac{\partial\gamma}{\partial t}\right) + \mu_{(p)}\frac{\partial v^4_{(p)}}{\partial t} = 0$$

$$\Longrightarrow \frac{\partial\mu_{(p)}}{\partial x}\frac{dx_{(p)}}{dt} + \frac{1}{2}\mu_{(p)}\frac{dx_{(p)}}{dt}\left(\frac{\partial\alpha}{\partial x} + \frac{\partial\gamma}{\partial x}\right) + \frac{2\mu_{(p)}}{x + r_c}\frac{dx_{(p)}}{dt} + \mu_{(p)}\frac{\partial}{\partial x}\left(\frac{dx_{(p)}}{dt}\right)$$
$$+ \frac{\mu_{(p)}}{v^4_{(p)}}\frac{dx_{(p)}}{dt}\frac{\partial v^4_{(p)}}{\partial x} + \frac{\partial\mu_{(p)}}{\partial t} + \frac{1}{2}\mu_{(p)}\left(\frac{\partial\alpha}{\partial t} + \frac{\partial\gamma}{\partial t}\right) + \frac{\mu_{(p)}}{v^4_{(p)}}\frac{\partial v^4_{(p)}}{\partial t} = 0.$$

Now if y represents either x or t then

$$\frac{1}{v^4_{(p)}}\frac{\partial v^4_{(p)}}{\partial y} = \frac{1}{u^4_{(p)}}\frac{\partial u^4_{(p)}}{\partial y} - \frac{1}{2}\frac{\partial\gamma}{\partial y}$$
$$= \frac{u^1_{(p)}}{c^2 + u^1_{(p)}u^1_{(p)}}\frac{\partial u^1_{(p)}}{\partial y} - \frac{1}{2}\frac{\partial\gamma}{\partial y}.$$

Thus,

$$\frac{\partial \mu_{(p)}}{\partial x}\frac{dx_{(p)}}{dt} + \frac{1}{2}\mu_{(p)}\frac{dx_{(p)}}{dt}\frac{\partial \alpha}{\partial x} + \frac{2\mu_{(p)}}{x+r_c}\frac{dx_{(p)}}{dt} + \mu_{(p)}\frac{\partial}{\partial x}\left(\frac{dx_{(p)}}{dt}\right)$$
$$+ \frac{\mu_{(p)}u^1_{(p)}}{c^2+u^1_{(p)}u^1_{(p)}}\frac{dx_{(p)}}{dt}\frac{\partial u^1_{(p)}}{\partial x} + \frac{\partial \mu_{(p)}}{\partial t} + \frac{1}{2}\mu_{(p)}\frac{\partial \alpha}{\partial t} + \frac{\mu_{(p)}u^1_{(p)}}{c^2+u^1_{(p)}u^1_{(p)}}\frac{\partial u^1_{(p)}}{\partial t} = 0.$$

We now transform to our dynamic coordinate system, $\tilde{\mathbf{x}}$. The result is,

$$\frac{\partial \tilde{\mu}_{(p)}}{\partial \tilde{x}}\frac{d\tilde{x}_{(p)}}{d\tilde{t}} + \frac{1}{2}\tilde{\mu}_{(p)}\frac{d\tilde{x}_{(p)}}{d\tilde{t}}\frac{\partial \alpha}{\partial \tilde{x}} + \frac{2\tilde{\mu}_{(p)}}{\tilde{x}+\varepsilon^{k_x}r_c}\frac{d\tilde{x}_{(p)}}{d\tilde{t}}$$
$$+ \tilde{\mu}_{(p)}\frac{\partial}{\partial \tilde{x}}\left(\frac{d\tilde{x}_{(p)}}{d\tilde{t}}\right) + \varepsilon^{-4k_u}\frac{\tilde{\mu}_{(p)}\tilde{u}^1_{(p)}}{\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}}\frac{d\tilde{x}_{(p)}}{d\tilde{t}}\frac{\partial \tilde{u}^1_{(p)}}{\partial \tilde{x}}$$
$$+ \frac{\partial \tilde{\mu}_{(p)}}{\partial \tilde{t}} + \frac{1}{2}\tilde{\mu}_{(p)}\frac{\partial \alpha}{\partial \tilde{t}} + \varepsilon^{-4k_u}\frac{\tilde{\mu}_{(p)}\tilde{u}^1_{(p)}}{\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}}\frac{\partial \tilde{u}^1_{(p)}}{\partial \tilde{t}} = 0,$$

which becomes in the limit,

$$\frac{\partial \tilde{\mu}_{(p)}}{\partial \tilde{x}}\frac{d\tilde{x}_{(p)}}{d\tilde{t}} + \tilde{\mu}_{(p)}\frac{\partial}{\partial \tilde{x}}\left(\frac{d\tilde{x}_{(p)}}{d\tilde{t}}\right) + \lim_{\varepsilon\to\infty}\left\{\varepsilon^{-4k_u}\frac{\tilde{\mu}_{(p)}\tilde{u}^1_{(p)}}{\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}}\frac{d\tilde{x}_{(p)}}{d\tilde{t}}\frac{\partial \tilde{u}^1_{(p)}}{\partial \tilde{x}}\right\}$$
$$+ \frac{\partial \tilde{\mu}_{(p)}}{\partial \tilde{t}} + \lim_{\varepsilon\to\infty}\left\{\varepsilon^{-4k_u}\frac{\tilde{\mu}_{(p)}\tilde{u}^1_{(p)}}{\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}}\frac{\partial \tilde{u}^1_{(p)}}{\partial \tilde{t}}\right\} = 0. \qquad [7.6.1]$$

Here we have used the fact that $\lim_{\varepsilon\to\infty}(x+\varepsilon^{k_x}r_c)^{-1} = 0$ as well as equations [7.5.5] and [7.5.7]. To be sure of how the above equation behaves as $\varepsilon \longrightarrow \infty$ we need to investigate the ε-dependence of $\partial\tilde{u}^1_{(p)}/\partial\tilde{x}$ and $\partial\tilde{u}^1_{(p)}/\partial\tilde{t}$. The dependence of these quantities can easily be determined from the normalisation condition. We have, for y representing t or x,

$$\frac{\partial u^1_{(p)}}{\partial y} = \frac{1}{c^3}\left(c^2+u^1_{(p)}u^1_{(p)}\right)^{3/2}\frac{\partial}{\partial y}\frac{dx_{(p)}}{dt}e^{(\alpha-\gamma)/2} + \frac{1}{2c^2}u^1_{(p)}\left(c^2+u^1_{(p)}u^1_{(p)}\right)\left(\frac{\partial \alpha}{\partial y} - \frac{\partial \gamma}{\partial y}\right),$$

(ref. equation [7.5.4]) which imply that

$$\frac{\partial \tilde{u}^1_{(p)}}{\partial \tilde{y}} = \frac{1}{\tilde{c}^3}\varepsilon^{k_t+k_u-k_x}\left(\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right)^{3/2}\frac{\partial}{\partial \tilde{y}}\frac{d\tilde{x}_{(p)}}{d\tilde{t}}e^{(\alpha-\gamma)/2}$$
$$+ \frac{1}{2\tilde{c}^2}\tilde{u}^1_{(p)}\left(\tilde{c}^2+\varepsilon^{-4k_u}\tilde{u}^1_{(p)}\tilde{u}^1_{(p)}\right)\left(\frac{\partial \alpha}{\partial \tilde{y}} - \frac{\partial \gamma}{\partial \tilde{y}}\right)$$

(again noting that $k_x - k_t - k_u = 0$). Inserting these results into equation [7.6.1] gives

$$
\begin{aligned}
&\frac{\partial \tilde{\mu}_{(p)}}{\partial \tilde{x}} \frac{d\tilde{x}_{(p)}}{d\tilde{t}} + \tilde{\mu}_{(p)} \frac{\partial}{\partial \tilde{x}} \left(\frac{d\tilde{x}_{(p)}}{d\tilde{t}} \right) \\
&+ \lim_{\varepsilon \to \infty} \left\{ \varepsilon^{-4k_u} \frac{\tilde{\mu}_{(p)} \tilde{u}^1_{(p)}}{\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)}} \frac{d\tilde{x}_{(p)}}{d\tilde{t}} \left\{ \frac{1}{\tilde{c}^3} \left(\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)} \right)^{3/2} \frac{\partial}{\partial \tilde{x}} \frac{d\tilde{x}_{(p)}}{d\tilde{t}} e^{(\alpha - \gamma)/2} \right. \right. \\
&\qquad \left. \left. + \frac{1}{2\tilde{c}^2} \tilde{u}^1_{(p)} \left(\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)} \right) \left(\frac{\partial \alpha}{\partial \tilde{x}} - \frac{\partial \gamma}{\partial \tilde{x}} \right) \right\} \right\} \\
&\quad + \frac{\partial \tilde{\mu}_{(p)}}{\partial \tilde{t}} \\
&\quad + \lim_{\varepsilon \to \infty} \left\{ \varepsilon^{-4k_u} \frac{\tilde{\mu}_{(p)} \tilde{u}^1_{(p)}}{\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)}} \left\{ \frac{1}{\tilde{c}^3} \left(\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)} \right)^{3/2} \frac{\partial}{\partial \tilde{t}} \frac{d\tilde{x}_{(p)}}{d\tilde{t}} e^{(\alpha - \gamma)/2} \right. \right. \\
&\qquad \left. \left. + \frac{1}{2\tilde{c}^2} \tilde{u}^1_{(p)} \left(\tilde{c}^2 + \varepsilon^{-4k_u} \tilde{u}^1_{(p)} \tilde{u}^1_{(p)} \right) \left(\frac{\partial \alpha}{\partial \tilde{t}} - \frac{\partial \gamma}{\partial \tilde{t}} \right) \right\} \right\} \\
&\quad = 0,
\end{aligned}
$$

which finally becomes

$$\frac{\partial \tilde{\mu}_{(p)}}{\partial \tilde{x}} \frac{d\tilde{x}_{(p)}}{d\tilde{t}} + \tilde{\mu}_{(p)} \frac{\partial}{\partial \tilde{x}} \left(\frac{d\tilde{x}_{(p)}}{d\tilde{t}} \right) + \frac{\partial \tilde{\mu}_{(p)}}{\partial \tilde{t}} = 0. \qquad [7.6.2]$$

Equations [7.5.8] and [7.6.2] complete an extremely interesting calculation for we have shown that given the transformation group,

$$g\left(\varepsilon; t, x_{(p)}, u^1_{(p)}, \mu_{(p)}, G, c\right) = \left(\varepsilon^{k_t} t, \varepsilon^{k_x} x_{(p)}, \varepsilon^{k_u} u^1_{(p)}, \varepsilon^{k_\mu} \mu_{(p)}, \varepsilon^{k_G} G, \varepsilon^{-k_u} c\right), \qquad [7.5.1]$$

our General Relativistic equations of motion ([6.1.1]–[6.1.3]) reduce to the *planar symmetric*, gravity free equations of caustic formation in Newtonian theory. We notice that G plays the same role as in the Newtonian analysis. The resulting asymptotic solutions are exactly the same as the solutions obtained in §5.3. In other words, under an appropriate Galilean transformation and choice of boundary conditions, the asymptotic solution is given by $x = q(t, v)$ with the tangent bundle surface, S, reducing to S_q. This result corroborates the conclusion of chapter 5, which is that gravity does not play a part in caustic formation.

We also have the result that the asymptotic limit of the full General Relativistic equations is identical to the asymptotic limit of the equations obtained after L/cT has been set to zero. This result differs from the expectation at the beginning of this chapter where we supposed that the two limiting processes are distinct and thus could commute. It rather suggests that by choosing to scale the velocity of light we have somehow incorporated the Newtonian limit within the procedure, let $\varepsilon \longrightarrow \infty$. To explain we note that according to the transformation specified by equation [7.5.1],

$$\lim_{\varepsilon \to \infty} \frac{L}{cT} = \lim_{\varepsilon \to \infty} \varepsilon^{k_t - k_x - k_u} \frac{\tilde{L}}{\tilde{c}\tilde{T}} = 0$$

(since $k_t - k_x - k_u = 2(1-\beta)k_t < 0$), which is the Newtonian limit and thus highlights the coupled nature of our General Relativistic transformation group.

CHAPTER 8. TOWARDS AN EXISTENCE PROOF FOR THE NEWTONIAN EQUATIONS OF MOTION.

§8.1. Introduction.

In this chapter we begin the ground work in the setup of an existence proof for a solution to the Newtonian differential equations describing caustic formation (ref. §4.1). Although a complete proof cannot be given here in this thesis, the results we shall present are used to prove existence for a certain class of solution in [SC]. The method, which we are working towards, adopts a contraction mapping type of argument. We can think of this as a mathematical way of specifying an iterative procedure that tends to the required solution. For example, if J represents our contraction mapping, $x(t)$ the solution to a differential equation, and $x_0(t)$ the initial, and *approximate* solution that might be our first guess, then $x(t) = \lim_{n\to\infty} J^n[x_0](t)$, where $J^2[x_0] = J\,[J[x_0]]$, $J^3[x_0] = J\,[J\,[J[x_0]]]$ and so on. At this point we refer the reader to appendix 2, which defines the concept of a contraction mapping. We also give an example of how these ideas can be used to establish the existence of a unique solution to $dy/dx = f(x,y)$ with initial condition, $y(x_0) = y_0$. This example is useful for it highlights certain aspects of this approach that need to be considered with care.

The basic idea begins by reformulating the Newtonian equations of motion in terms of an integral, or set of integral equations. This allows a function, J, to be defined that is a map from some, as yet undefined, metric space such that the solution to equations [4.1.1]–[4.1.3] (or equivalently [4.5.3]–[4.5.5]) correspond to a fixed point of J. To clarify what we mean, we illustrate using the example mentioned above. In appendix 2 we show that $dy/dx = f(x,y)$, $y(x_0) = y_0$, is equivalent to

$$y(x) = y(x_0) + \int_{x_0}^{x} f(t, y(t))\,dt. \qquad [8.1.1]$$

In this case, given that $x, x_0 \in [a,b]$ and $\phi \in C[a,b]$, we could define J to be the map, $J\colon C[a,b] \longrightarrow C[a,b]$,

$$J[\phi](x) = y(x_0) + \int_{x_0}^{x} f(t, \phi(t))\,dt,$$

so that the above integral equation, [8.1.1], is equivalent to $J[y] = y$.

Another example of how we might define J to form a contraction mapping is that where the above ideas are applied to the tv space formulation of our Newtonian problem. We have for the multi-dust region in M,

$$\frac{\partial F}{\partial v_i} = (-1)^{i+1}\sigma_i + \frac{\partial x}{\partial v_i}\sum_{j\neq i}(-1)^{j+1}\left(\sigma_j\left(\frac{\partial x}{\partial v_j}\right)^{-1}\circ\phi_j\right), \qquad [4.5.3]$$

$$v_i - \frac{\partial x}{\partial t} = F\frac{\partial x}{\partial v_i}, \qquad [4.5.4]$$

$$\frac{\partial \sigma_i F}{\partial v_i} + \frac{\partial \sigma_i}{\partial t} = 0, \qquad [4.5.5]$$

where the arguments of all equations are (t, v_i). The function, ϕ_j, defined by

$$\phi_j(t, v_i) = \{(t, v_j)|v_j \neq v_i,\, x(t, v_i) = x(t, v_j)\},$$

highlights the fact that in calculating the acceleration there are contributions to the total density from the three velocities that satisfy $x = x(t, v_i)$ (This harks back to §2.1 where we talked about disjoint regions in TM contributing to the gravitational force.). Now suppose that we are solving these equations for some initial conditions on x and σ_i at time, s (the boundary condition on F is that $F(t,0) = 0$). Let $\Psi(s,t)\colon \mathbb{R} \longrightarrow \mathbb{R}$ be the diffeomorphism from the v-axis at time, s, to the v-axis at time, $t > s$ defined by the geodesic flow, $Z = \partial/\partial t + F\partial/\partial v$. This vector field is equivalent to the Z of §3.3 except that now we assume that the tangent bundle surface, S, has local coordinates, (t, v). This means that if $x(t)$ represents a specific integral curve of Z then $\Psi(s,t)(u) = v$ (really $\Psi(s,t;F)(u)$ since in actual fact we have

$$\frac{\partial}{\partial t}\Psi(s,t)(u) = F(t, \Psi(s,t)(u)),$$

implying that the form of Ψ is dependent on the given F) where $u = dx/dt(s)$ and $v = dx/dt(t)$. In addition we also have $\Psi(s,s)$ representing the identity and $\Psi(s,t)^{-1} = \Psi(t,s)$, the inverse function.

We can we now rewrite equations [4.5.3]–[4.5.5]. We have from [4.5.5]

$$\frac{d\sigma_i}{dt}(t, \Psi(s,t)(u)) = -\sigma_i(t, \Psi(s,t)(u))\frac{\partial F}{\partial v}(t, \Psi(s,t)(u))$$

$$\Longrightarrow \sigma_i(t,v) = \sigma_i(s,\Psi(t,s)(v)) - \int_s^t \sigma_i(t',\Psi(t,t')(v))\frac{\partial F}{\partial v}(t',\Psi(t,t')(v))\,dt' = J_1[x,\sigma_i,F](t,v).$$

Similarly, [4.5.4] gives

$$\frac{dx}{dt}(t,\Psi(s,t)(u)) = \Psi(s,t)(u)$$

$$\Longrightarrow x(t,v) = x(s,\Psi(t,s)(v)) + \int_s^t \Psi(t,t')(v)\,dt' = J_2[x,\sigma_i,F](t,v),$$

whereas equation [4.5.3] is equivalent to

$$F(t,v) = -\int_0^v \sigma_i(t,v') + \frac{\partial x}{\partial v}(t,v')\sum_{j\neq i}\left(\sigma_j\left(\frac{\partial x}{\partial v}\right)^{-1}\circ\phi_j\right)(t,v')\,dv' = J_3[x,\sigma_i,F](t,v).$$

It follows then that we can construct a map (that maps some, as yet, undefined metric space to itself) by the equation,

$$(x,\sigma_i,F) = (J_1[x,\sigma_i,F],J_2[x,\sigma_i,F],J_3[x,\sigma_i,F]) = J[x,\sigma_i,F].$$

The above equation completes the second example illustrating how a differential equation or set of equations can be reformulated so that they suggest a possible candidate for a contraction mapping. It might be possible to proceed with this development in our quest for an existence proof, however, we find that by reformulating our Newtonian equations using a *Lagrange* coordinate system, we obtain a much simplified set of differential equations with which to work with. We shall concentrate on this new formalism for the rest of this chapter and indeed thesis.

The next step in any contraction mapping proof involves defining the metric space. We neglected to stress this in our second example because we were primarily concerned with ensuring that the reader understands the motivation behind the definition of J. To complete the specification of this map we must of course state the metric space upon which J acts. That is to say we need to define the class of functions which J takes as its arguments, as well as the metric itself. In reality this process is the hardest part. The reason for this is that in order for us to determine the appropriate metric space, we must have a good idea of what we expect the solution to look like. Moreover, even if this is the case and we know to some degree of certainty where the function or its derivatives become unbounded, for example, the process of defining the metric space is one of trial and error. Having said this intuition is a great help in reducing the number of possible permutations to a reasonable number.

The final stage is to show that J does in fact define a contraction mapping. Although this is not necessarily easy, it can, in general, always be done provided the metric space can be made sufficiently small enough. A good example of what we mean is that given in appendix 2. There we supposed that f in [8.1.1] was Lipschitz continuous ($|f(x, y_1) - f(x, y_2)| \leq K|y_1 - y_2|$) in order to prove the theorem. The constant, K, was chosen such that $K < 1/\delta(b-a)$ where $\delta = d(y_1, y_2)$. Had we chosen K unwisely so that it did not satisfy this condition then our defined J would not have been a contraction. In this case we have restricted our class of functions by reducing K. Another example is that where we restrict the size of the region on which we are trying to prove existence. Suppose that we were stuck with a K such that $K > 1/\delta(b-a)$. Then defining a' and b' so that $K < 1/\delta(b' - a')$ again means that J is a contraction, provided of course we still have $x_0 \in [a', b']$.

The above summarises the procedure that we hope to take. The rest of this chapter will be dedicated to defining both the map, which stands as our candidate for a contraction mapping, and the metric space. To begin, the next section introduces the Lagrangian coordinate system, (t, X), that we will adopt. We do this whilst presenting the exact solution that is to be found in the external region, i.e. those points not enclosed by the cocaustic. The significance of the X variable is three-fold. Firstly, and primarily, we will show that for the case of a single dust with initial conditions $x = q(t, X)$, X is proportional to the force at (t, X) (or equivalently, $x_X(t)$). We can think of this as its definition and it is the job of §8.3 to show how this generalises to the case where we have a system of several superimposed dusts. Its second importance is its relation to the velocity on the initial time slice; we have $v(0, X) = X$ (For the rest of this thesis, we shall assume initial conditions that coincide with the special model used to illustrate the zero gravity case. This means that we also have the relationship, $x(0, X) = -X^3$.). Finally, we have that geodesics, $x_X(t) = x(t, X)$ say, become functions of time, labelled by the X variable, and correspond to straight, vertical lines in tX space. In this sense, therefore, we have a comoving coordinate system.

In §8.3 we use the ideas developed in §8.2 to transform our differential equations describing caustic formation so that they are now written in terms of Lagrangian coordinates. This involves the introduction of three new variables, $X_i(t, X)$ ($i = 1, 2, 3$), that are defined by $x(t, X_i(t, X)) = x(t, X)$. That is to say, X_i *labels* those geodesics that are coincident at $x_X(t)$ in tx space. In some sense this is quite a complicated quantity for we can use it, quite unambiguously, as both a dependant or independent variable. To explain, we point out that the subscript i has an additional meaning for we choose, $X_1 \leq -X_c \leq X_2 \leq X_c \leq X_3$, where $X = X_c(t)$ represents the equation of the caustic in tX space. This implies that the i also corresponds to the dust number in much the same way as the i in v_i of §4.5. It follows, by the fact that we assume a multi-dust spacetime, that in each region $x_X(t)$ is unique and moreover, that $X_i(t, X) = X$ when X satisfies the above condition for the ith region. This illustrates its behaviour as an independent variable

The two reasons why we adopt this approach are as follows: the first is that the differential equations that describe the motion of particles get completely decoupled from those that describe the changes in the density. This has a secondary effect (also the second reason), which is that the force at any point is given by a vastly simplified expression. To illustrate, we state without proof the force equation in this new coordinate system,

$$F(t, X) = A(X_1(t, X) - X_2(t, X) + X_3(t, X)). \qquad [8.1.2]$$

The constant, A, is defined so that $AX_1(t, X)$, for example, represents the *mass of dust 1* enclosed between $x_X(t)$ and the origin in tx space. The right hand side of equation [8.1.2] therefore represents the *total mass* between $x_X(t)$ and $x = 0$. Finally we complete this section by suggesting a map, J such that $x(t, X) = J[x](t, X)$, as a candidate for a contraction mapping.

In sections 8.4, 8.5, and 8.6 we discuss the behaviour of the type of solution that we are trying to prove existence of. This is based on the assumption that the solution can be thought of as a perturbation on q (the solution for $G = 0$) in some sufficiently small region of tX space containing the cusp. By assumption, the continuity of the X_i dictate that x must be at least C^2 in the time coordinate. However, due to the step function-like behaviour in the acceleration of any particle as it crosses the cocaustic (remember that in tx space the caustic and cocaustic are identical and that the former corresponds to points where there are unbounded densities), we expect the differentiability with respect to X of $x(t, X)$, a geodesic, to be less than C^2.

The caustic is another problem area and to examine the solution's behaviour close to these curves we split the multi-dust region into three parts. These are represented by those points that correspond to: $0 \leq X \leq X_c$, $X_c \leq X \leq X_0$ and $X_0 \leq X \leq X_{cc}$. The boundary, X_0, between the last two regions is simply there to illustrate that the region close to the cocaustic should never stretch to the caustic and visa versa. We shall thus leave this undefined. Lastly we point out that by the symmetry inherited from q, we need only consider the half problem.

The discussion concerning the differentiability of x consists of a series of calculations that concentrate on investigating the continuity of $\partial^2 x/\partial X^2$. We allow q to provide a basis from which we can begin this calculation in the sense that we shall assume that $\partial x/\partial X$ is continuous everywhere as is $\partial^2 x/\partial X^2$ except at points corresponding to the caustic or cocaustic. To perform this investigation, we differentiate $x = J[x]$ twice with respect to X to obtain $\partial^2 x/\partial X^2 = \partial^2/\partial X^2 J[x]$. The right hand side of this equation is now the second order differential of a double iterated integral. We can interchange the two operations to get an expression for $\partial^2/\partial X^2 J[x]$ as an iterated integral of terms involving the second order derivative of x. Specifically we obtain

$$\frac{\partial^2 x}{\partial X^2}(t,X) = -6X + A\int_0^t\int_0^{t'} \frac{\partial^2 X_1}{\partial X^2}(t'',X) - \frac{\partial^2 X_2}{\partial X^2}(t'',X) + \frac{\partial^2 X_3}{\partial X^2}(t'',X)\, dt'' dt'. \tag{8.1.3}$$

The next stage is to express the integrand as a power series expansion in either $|t - k_c|$ or $|t - k_{cc}|$ (here $k_c(X) = X_c^{-1}(X)$ and $k_{cc}(X) = X_{cc}^{-1}(X)$ so that $t = k_c$ and $t = k_{cc}$ are alternative representations of the equations of the caustic and cocaustic respectively) depending on if we are close to the caustic or cocaustic. In constructing these series, we illustrate the expected singular nature of $\partial^2 x/\partial X^2$ by pulling out a factor of $|t - k|^{-p_k}$, i.e. by writing,

$$\frac{\partial^2 x}{\partial X^2}(t,X) = \sum_{n=0}^{\infty} \alpha_{(n)}(|t-k|,X)|t-k|^{n-p_k} = |t-k|^{-p_k}\alpha(|t-k|,X).$$

Here k represents either the caustic or cocaustic depending on which region of our tX plane we are considering and α is assumed to be analytic. The constant, p_k, is then determined so that we have consistency between the left and right hand sides of equation [8.1.3]. Assuming that $p_k > 0$ means that $\partial^2 x/\partial X^2$ is less than C^2 meaning that we may be forced to work with a complicated metric space.

In the final section we take the results so far obtained and propose a metric space in which we hope to find the solution. The aim is to base the set of functions that we are considering on

$$\mathcal{V} = \left\{ f\colon B_T \longrightarrow \mathbb{R} \mid f(t,X) = -f(t,-X), \quad \tfrac{\partial f}{\partial X}(0,0) = 0, \quad \|f\|_{\mathcal{V}} < \infty \right\} \qquad [8.1.4]$$

where

$$\|f\|_{\mathcal{V}} = \sup_{(t,X)\in B_T} \left\{ \left|\frac{\partial^2 f}{\partial X^2}\right| + \left|\frac{\partial^2 f}{\partial t^2}\right| + \left|\frac{\partial^2 f}{\partial X \partial t}\right| \right\}$$

and

$$B_T = \left\{ (t,X) \in \mathbb{R}^2 \mid |X| < T, \quad 0 \le t < T \right\}.$$

B_T is the domain in which we hope to prove existence and we can control the size of this by varying T. $\|f\|_{\mathcal{V}}$, which we shall use to define the metric function, is simply the standard C^2 norm. If we are unlucky and it turns out that we must expect x to be less than C^2, then we may have to introduce a weighting that modifies the above norm to

$$\|f\|_{\mathcal{V}} = \sup_{(t,X)\in B_T} \left\{ \frac{1}{|t-k|^p} \left|\frac{\partial^2 f}{\partial X^2}\right| + \frac{1}{|t-k|^{p'}} \left|\frac{\partial^2 f}{\partial X^2}\right| + \left|\frac{\partial^2 f}{\partial t^2}\right| + \left|\frac{\partial^2 f}{\partial X \partial t}\right| \right\},$$

where p and p' 'measure' the rate at which $\partial^2 f/\partial X^2$ becomes unbounded as (t,X) approaches these curves. Finally the conditions placed on the space of functions in the above definition for $\mathcal{V}$ reflect the assumption that x always *looks like* q. To express the fact that we look for a solution that is *near* to q we shall restrict $\mathcal{V}$ still further by considering

$$\mathcal{V}_a = \{ f \in \mathcal{V} \mid \|f - q\|_{\mathcal{V}} < a \}.$$

§8.2. Exact solution for the external region.

Working in tv space we have for a single dust,

$$\frac{d\sigma}{dt} = -\sigma \frac{\partial F}{\partial v} \qquad [8.2.1]$$

where we note that in terms of tv coordinates, $d/dt = \partial/\partial t + F\partial/\partial v$. This represents the relationship between the force and the matter in the region external to that enclosed by the cocaustic for we realise that here we have a well defined fluid flow vector. Let us now introduce comoving coordinates. We make the transformation by the assumption that for any function, f say, $f(t,X) = f(t, v(t,X))$ with $v(0,X) = X$. It follows then, using the results summarised in appendix 1, that

$$\left(\frac{\partial F}{\partial v}\right)_t = \left(\frac{\partial F}{\partial X}\right)_t \left(\frac{\partial v}{\partial X}\right)_t^{-1} \qquad [8.2.2]$$

and

$$\left(\frac{\partial f}{\partial t}\right)_X = \left(\frac{\partial f}{\partial v}\right)_t \left(\frac{\partial v}{\partial t}\right)_X + \left(\frac{\partial f}{\partial t}\right)_v = \left.\frac{d}{dt}\right|_{X=\text{const}} f(t,v),$$

so that

$$\left(\frac{\partial F}{\partial X}\right)_t = \left(\frac{\partial}{\partial X}\left(\frac{dv}{dt}\right)\right)_t = \left(\frac{\partial}{\partial X}\left(\frac{\partial v}{\partial t}\right)_X\right)_t = \left(\frac{\partial}{\partial t}\left(\frac{\partial v}{\partial X}\right)_t\right)_X = \frac{d}{dt}\left(\frac{\partial v}{\partial X}\right)_t.$$

Applying these equations to [8.2.1] gives

$$\frac{d\sigma}{dt} = -\sigma \left(\frac{\partial v}{\partial X}\right)_t^{-1} \frac{d}{dt}\left(\frac{\partial v}{\partial X}\right)_t$$

$$\Longrightarrow \sigma(t,X) = \sigma_0(X)\left(\frac{\partial v}{\partial X}\right)_t^{-1}. \qquad [8.2.3]$$

We now assume zero gravity boundary conditions at $t = 0$ to determine $\sigma_0(X)$. This means that $\partial v/\partial X(0,X) = 1$. To obtain an expression for $\sigma(t,X)$ we have to refer to an asymptotic analysis of the Newtonian equations written in terms of (t,v) coordinates ([4.5.3]–[4.5.5]). This work has not been presented in this thesis, however, it has been done and we summarise the results as follows. The general approach follows very closely that described in §5.3. That is to say we consider the group of transformations,

$$g(\varepsilon; t, v_i, F, x, \sigma_i, G) = \left(\varepsilon^{k_t} t, \varepsilon^{\beta k_t} v_i, \varepsilon^{k_F} F, \varepsilon^{k_x} x, \varepsilon^{k_\sigma} \sigma_i, \varepsilon^{k_G} G\right),$$

in conjunction with equations [4.5.3]–[4.5.5] with G reinserted (this quantity only appears as a multiplicative factor in front of each σ_i in equation [4.5.3]). For invariance of our equations we require that $k_F = (\beta - 1)k_t$, $k_x = (\beta + 1)k_t$, $k_\sigma = -(1+\gamma)k_t$ and $k_G = \gamma k_t$. Asymptotic solutions exist provided $k_t > 0$, $\beta > 0$ and $\gamma > 0$, resulting in a solution for $\sigma(t, v_i)$ given by [4.6.3] but modified slightly to give $\sigma_i = A_i t^{-(1+\gamma)k_i} |\xi_i|^{-(1+\gamma)/\beta}$. If we assume that we require non zero and finite behaviour on $\xi = 0$ then we must have $1 + \gamma = 0$. This means that $\sigma_i(t, v_i) = A_i$ represents the asymptotic solution for the density in a tv space formulation of caustic formation. Finally, using this result, it follows that we have $\sigma_0(X) = A$ where A is also a constant.

Next we consider,

$$\left(\frac{\partial F}{\partial v}\right)_t = \sigma.$$

In terms of tX coordinates this becomes

$$\left(\frac{\partial F}{\partial X}\right)_t = \sigma_0,$$

which if we integrate gives

$$F(t, X) = \int^X \sigma_0 dX + \alpha(t) = AX + \alpha(t),$$

where α is an arbitrary function. In section 5.4 we showed that the tx space formulation of the Newtonian equations of motion exhibit the symmetry,

$$\tilde{t} = t, \qquad \tilde{x} = x + H(t), \qquad \tilde{v} = v + H'(t), \qquad \tilde{F} = F + H''(t).$$

A similar symmetry exists (although we do not prove this here) in the tv space formulation. We therefore set $\alpha(t) = 0$ to obtain

$$F(t, X) = AX. \qquad [8.2.4]$$

This result is worth remarking on since we shall use it later on. Firstly, we note that since $d/dt|_{X=\text{const}} = (\partial/\partial t)_X$ we can define curves, $x_X(t) = x(t, X)$, that represent geodesics in tx space. Equation [8.2.4] is therefore telling us that along these curves, the gravitational force is constant. Moreover, since we are assuming zero gravity boundary conditions, the constant, A, can be chosen so that AX represents the mass enclosed by $x_0(0)$ and $x_X(0)$. This point is extremely important for we shall use this idea when we come back to considering caustic formation. The simplicity of [8.2.4] illustrates what we meant in the introduction when we stated that this Lagrange approach decouples those differential equations describing the motion of our particles from those describing the changes in the density. That's not to say that σ has no dependence on the motion of our fluid; the presence of the $(\partial v/\partial X)_t^{-1}$ term dictates the opposite. What we mean is that, as shown, the original differential equations can be manipulated into a form that has the density dependency removed from the force equation allowing a solution for σ, [8.2.3], to be obtained.

We complete this section by integrating equation [8.2.4] to obtain the solutions for x, v and consequently σ. Since

$$\left.\frac{d^2x}{dt^2}\right|_{X=\text{const}}(t,X) = \frac{\partial^2 x}{\partial t^2}(t,X) = \frac{\partial v}{\partial t}(t,X) = F(t,X),$$

we obtain firstly $v(t,X) = X(1+At)$, and then $x(t,X) = X(t + At^2/2) - X^3$, using of course the boundary conditions, $v(0,X) = X$ and $x(0,X) = -X^3$ (corresponding to $q(0,X) = -X^3$). With v defined we can now fully determine the density. Since $\partial v/\partial X = 1 + At$ we have $\sigma = A/(1+At)$. Thus the solution for a *self gravitating dust* with zero-gravity initial conditions is

$$x = Xt + \frac{1}{2}AXt^2 - X^3,$$

$$v = X(1+At)$$

and

$$\sigma = \frac{A}{1+At}.$$

In terms of a tv space formulation we obtain

$$x = -\frac{v^3}{(1+At)^3} + \frac{v}{(1+At)}\left(t + \frac{1}{2}At^2\right),$$

$$v = v_0(1+At)$$

and

$$\sigma = \frac{A}{1+At}.$$

§8.3. Lagrangian formulation of Newtonian caustics.

According to our tv space formulation of the problem we have essentially four functions, σ_i and x, described by [4.5.3]–[4.5.5]. To convert these equations so that each term is a function of t and X, our Lagrangian coordinates, we use the results of appendix 1. The first thing to notice, however, is that the analysis concerning σ for a single dust in the previous section holds in our multi-dust scenario. That is to say, we can immediately write the solution for σ_i as $\sigma_i(t,X) = A(\partial v_i/\partial X)_t^{-1}$. This is essentially because the aforementioned solution was obtained without reference to the specific form of the force equation (ref. §8.2). Furthermore, since $\partial^2 x/\partial t^2(t,X) = F(t,X)$, we need only investigate how the force equation transforms into these new coordinates. We have

$$\left(\frac{\partial F}{\partial v_i}\right)_t = (-1)^{i+1}\sigma_i + \sum_{j\neq i}(-1)^{j+1}\sigma_j\left(\frac{\partial v_j}{\partial v_i}\right)_t . \qquad [4.5.3]$$

This is a tv space formulation of equation [4.1.1] and tells us that the force at (t,v) is a function of the mass enclosed by $x(t,v)$. Here each v_i, defined by $x(t,v_i(t,v)) = x(t,v)$, represents the velocity fields of those geodesics that are coincident at a point in tx space (Recall that since we have chosen to model caustic formation using a multi-dust spacetime, the velocity field is unique having a one to one correspondence with geodesics on M.). The subscript has a further meaning for it labels those regions in TM corresponding to different dusts. That is to say, we have, $v_1 \leq -v_c \leq v_2 \leq v_c \leq v_3$, where $v = v_c(t)$ represents the equation of the caustic. We note that as v varies, depending on which region we happen to be considering at the time, there must exist a v_i such that $v_i(t,v) = v$.

Let us now introduce comoving coordinates, (t,X), such that if $f(t,v)$ is any function then we define $f(t,X) = f(t,v(t,X))$. with $v(0,X) = X$. An alternative choice for defining the relationship between X and v might be $v(k_{cc}(X),X) = X$. This essentially specifies initial conditions for v at the cocaustic. We let (t,X) be any point in our new tX space but to mimic the different regions in tv space mentioned earlier, we introduce three functions, $X_i(t,X)$, such that $X_1 \leq -X_c(t) \leq X_2 \leq X_c(t) \leq X_3$, where $X = X_c(t)$ (and its inverse, $t = k_c(X)$) represents the equation of the caustic. This means that again we have i labelling the different dusts.

We continue this construction by stipulating an additional requirement that can be thought of as a secondary definition of the X_i. That is to say we assume that $x(t, X_i(t,X)) = x(t,X)$. This means that $X_i(t,X)$ labels those geodesics that are coincident at $(t, x_X(t))$ in tx space. Having said this, however, it does not make sense to define geodesics such as $x_{X_i}(t)$. The reason for this is purely because X_i is a function of X; there is nothing to stop us considering curves such as $y(t, X_i(t,X))$ except that we cannot attach any physical meaning to them. In this sense, the X_i constitute quite confusing quantities, more so if we consider the fact that they can be thought of as both dependent and independent variables. That is to say, in all of the above we have considered them as dependent variables, however, we point out that due to the uniqueness of curves, $x_X(t)$, in any of the regions defined above, we must always have one X_i such that $X_i(t,X) = X$ depending on which region we are considering. This is an artifact of the fact that in TM the tangent bundle surface can be constructed by 'gluing together' three separate subsurfaces.

Now we move on to transforming [4.1.1]. Using the results of appendix 1 and following similar lines to the previous chapter we have

$$\left(\frac{\partial F}{\partial X_i}\right)_t \left(\frac{\partial v_i}{\partial X_i}\right)_t^{-1} = (-1)^{i+1}\sigma_{i0}(X_i)\left(\frac{\partial v_i}{\partial X_i}\right)_t^{-1} + \sum_{j\neq i}(-1)^{j+1}\sigma_{j0}(X_j)\left(\frac{\partial v_j}{\partial X_j}\right)_t^{-1}\left(\frac{\partial v_j}{\partial v_i}\right)_t .$$

Note how we have already used the X_i as an independent variable as we have assumed that a specific region of our tX space has been chosen, i.e., we have the relation, $X_i(t,X) = X$. This implies that

$$\left(\frac{\partial F}{\partial X_i}\right)_t = (-1)^{i+1}\sigma_{i0}(X_i) + \sum_{j\neq i}(-1)^{j+1}\sigma_{j0}(X_j)\left(\frac{\partial v_j}{\partial X_j}\right)_t^{-1}\left(\frac{\partial v_j}{\partial v_i}\right)_t\left(\frac{\partial v_i}{\partial X_i}\right)_t$$

$$\Longrightarrow \int^{X_i}\left(\frac{\partial F}{\partial X_i}\right)_t dX_i = \int^{X_i}(-1)^{i+1}A\,dX_i + \int^{X_i}\sum_{j\neq i}(-1)^{j+1}A\left(\frac{\partial v_j}{\partial X_j}\right)_t^{-1}\left(\frac{\partial v_j}{\partial v_i}\right)_t\left(\frac{\partial v_i}{\partial X_i}\right)_t dX_i + \alpha(t),$$

where $\alpha(t)$ is the arbitrary function of integration. As per the previous section, the force equation is again invariant under the addition of arbitrary functions of time. This means that the above then simplifies to

$$F(t,X_i) = \sum_j \int^{X_j(t,X_i)} (-1)^{j+1} A \, dX_j$$

$$\Longrightarrow F(t,X) = A(X_1(t,X) - X_2(t,X) + X_3(t,X)). \qquad [8.3.1]$$

In terms of $x(t,X)$, the desired solution, the above becomes the following second order partial differential equation,

$$\frac{\partial^2 x}{\partial t^2}(t,X) = A(X_1(t,X) - X_2(t,X) + X_3(t,X)), \qquad [8.3.2]$$

together with the initial conditions, $x = q(0,X)$ and $v = \partial q/\partial t(0,X)$. Of course we also have

$$\sigma_i(t,X) = A\left(\frac{\partial v_i}{\partial X_i}\right)_t^{-1},$$

but since this is a solution, albeit dependent on the knowledge of v_i, we realise that the problem of obtaining a solution to the Newtonian equations of motion boils down to solving just a single differential equation, namely [8.3.2].

Now we consider the beginnings of the construction of a contraction mapping proof. We first of all write [8.3.2] as an integral equation. One example is the following,

$$x(t,X) = -X^3 + A\int_0^t \int_0^{t'} X_1(t'',X) - X_2(t'',X) + X_3(t'',X)\, dt'' dt',$$

where we have considered initial conditions corresponding to $x = q(0,X)$. However, care must be taken in interpreting this equation because the integration with respect to the time coordinate must pass through the external region before it reaches the cocaustic (the equation of which is defined by $t = k_{cc}(X)$ or $X = X_{cc}(t)$). In this instance, therefore, we must remember that the X_i are defined by $x(t,X_i(t,X)) = x(t,X)$ for $|X| \leq X_{cc}(t)$, $X_1(t,X) = X$ for $X < -X_{cc}(t)$ with $X_2(t,X) = X_3(t,X) = 0$, and $X_3(t,X) = X$ for $X > X_{cc}(t)$ with $X_1(t,X) = X_2(t,X) = 0$. An alternative integral formulation might be

$$x(t,X) = Xk_{cc} + \frac{1}{2}AXk_{cc}^2 - X^3 + A\int_{k_{cc}}^t \int_{k_{cc}}^{t'} X_1(t'',X) - X_2(t'',X) + X_3(t'',X)\, dt'' dt'.$$

This reflects the assertion that we wish to specify boundary conditions on the cocaustic itself. That is to say, we have also used the solution to the external region to evolve the gravity free initial conditions specified at $t = 0$ up to the cocaustic. Now there are two things that we must be aware of when interpreting the above equation. Firstly, we only consider t such that $t \geq k_{cc}$. This avoids the confusion produced by the different definitions of X_i in the above. The second thing is that the symmetry of our system still allows the external solution to be valid, even though at points to the interior of (t, X) say, geodesics are passing through $X = 0$ (corresponding to $x_X(t) = 0$, the axis of symmetry). To explain, we notice that one of the conditions on x specified by equation [8.1.4] (we must have $x \in \mathcal{V}$ for our contraction mapping proof to work) is that $x(t, X) = -x(t, -X)$. It follows that as time evolves and a geodesic, $x_X(t)$ say crosses that of $X = 0$, its opposite partner, $x_{-X}(t)$ with the same mass associated with it appears from the other side. Thus provided $|X'| > X_{cc}$, the mass enclosed by $x_{X'}(t)$ is constant resulting in a uniform acceleration according to $F = -AX'$.

For simplicity we choose the first of the above integral equations to work with. This implies the following as a candidate for our contraction mapping,

$$J[f](t,X) = -X^3 + A\int_0^t \int_0^{t'} \chi[f](t'',X)\, dt''\, dt', \qquad [8.3.3]$$

where we have defined $\chi[f](t, X)$ to represent the right hand side of equation [8.3.2] evaluated for any trial or approximate solution, $f(t, X)$. The function f is taken to be a member of some undefined function/metric space. Our next task is to determine what this metric space is.

§8.4. Investigation to determine the differentiability of x.

Before we can define the function/metric space upon which J acts, we must investigate the behaviour of our solution, x. It is insufficient to simply assume each f in equation [8.3.3] to be at least C^n because without prior knowledge, it is difficult to say whether the iterative procedure defined by J will preserve this degree of differentiability (Remember, if we stipulate a certain degree of differentiability on f then we must show that $J[f]$ also exhibits this degree of differentiability in order for J to be a mapping from some metric space to itself.). This implies that by the term 'behaviour' we mean the degree of differentiability that we expect x to have. In addition to this, we may have the situation where x might be C^n but that we are required to know something about the $(n+1)$th order derivative. In this case by 'behaviour' we mean knowledge of those regions in our tX plane where $D^{n+1}(x)$ is discontinuous and possible the rate at which these functions become unbounded.

The general procedure in determining the behaviour of x, therefore, is to assume the lowest order of differentiability for x and then show, by explicit calculation, that $J[x]$ has similar behaviour. We do this using

$$\frac{\partial x}{\partial X} = t - 3X^2 + A\int_0^t \int_0^{t'} \frac{\partial \chi[x]}{\partial X}(t'', X)\, dt''dt',$$

$$\frac{\partial^2 x}{\partial X^2} = -6X + A\int_0^t \int_0^{t'} \frac{\partial^2 \chi[x]}{\partial X^2}(t'', X)\, dt''dt' \qquad [8.4.1]$$

and

$$\frac{\partial^2 x}{\partial t \partial X} = 1 + A\int_0^t \frac{\partial \chi[x]}{\partial X}(t', X)\, dt',$$

where we realise that the x on the left hand side of all of the above really represents $J[x]$.

Now we can automatically assume that x must be C^2 with respect to the time coordinate because the right hand side of equation [8.3.1] is continuous. This is an assumption based on the fact that we require the surface, $S \in TM$, generated by geodesics, $x_X(t)$, to be continuous. It remains therefore to determine the differentiability with respect to X. We shall assume that x is definitely C^1 in the X coordinate because we require that the velocity field associated with our dust particles to be continuous and this is as far as we can go. We don't know anything about the second derivative with respect to X except possibly that it might be continuous everywhere apart from those points that lie on either the cocaustic or caustic.

We shall base our analysis on these assumptions by proving that they are self consistent under a single application of J in the following manner. The idea is to replace the integrand of all of the above by Taylor expansions that reflect the assumptions made on the differentiability of x. For the reasons stated above, however, we will only consider [8.4.1]. The definition of $\chi[x]$ implies that we need to consider $\partial^2 X_i / \partial X^2$.

From the definition of X_i, namely $x(t, X_i(t, X)) = x(t, X)$, we find by differentiation that

$$\frac{\partial^2 X_i}{\partial X^2} = \frac{x''(X)}{x'(X_i)} - \frac{x'(X)^2 x''(X_i)}{x'(X_i)^3}. \qquad [8.4.2]$$

This equation illustrates exactly where we expect possible unbounded behaviour in x'': one can see that things only go bad when $X_i \longrightarrow X_c$ since in these regions $\lim_{X_i \to X_c} x'(t, X_i) = 0$. To proceed with our investigation we should suppose that our multi-dust region can be split into six parts. By the symmetry we expect x to have, however, we only need consider the positive half of the tX plane. This reduces the number of regions to three. We define these regions as those points corresponding to $0 \leq X \leq X_c$, $X \geq X_c$ close to the caustic and $X \leq X_{cc}$ close to the cocaustic. With these in mind, we then expand each quantity appearing within the integral on the right hand side of [8.4.1] in terms of $|t - k_c|$ or $|t - X_{cc}|$ (i.e. for $X \leq X_{cc}$ we might have

$$\frac{\partial^2 x}{\partial X^2}(t, X) = \sum_{n=0}^{\infty} \alpha_{(n)}(X)|t - k_{cc}|^{n-p} = |t - k_{cc}|^{-p} \alpha(|t - k_{cc}|, X)$$

for example) depending on which region we are considering. This will allow us to perform a single iteration defined by [8.4.1] so that we can determine the value of p by insuring that we have consistency between the left and right hand sides.

§8.5. Near the cocaustic.

We write equation [8.4.1] as

$$\frac{\partial^2 x}{\partial X^2} = -6X + A \int_0^t \int_0^{t'} \frac{\partial^2 X_1}{\partial X^2}(t'', X) - \frac{\partial^2 X_2}{\partial X^2}(t'', X)\, dt''dt', \qquad [8.5.1]$$

since in this region, $X_3(t, X) = X$. Now equation [8.4.2] tells us how to relate the derivatives of X_i to derivatives of x. In order to feed in the assumptions regarding x and its derivatives mentioned in the previous section, we replace all terms on the right hand side of [8.4.2] that are expected to produce singular behaviour near the caustic by an appropriate Taylor expansion. To illustrate what we mean by this we first of all notice that for X close to X_{cc}, $x(t, X_1(t, X)) = x(t, X)$ implies that X_1 must be close to $-X_c$ indicating that the $x'(X_1)$ terms on the right hand side of equation [8.4.2] will be causing a degree of singular behaviour as X approaches X_{cc}. To illustrate this we write, using Taylor's theorem with remainder,

$$\begin{aligned} x'(X_1) &= x'(-X_c) - |X_1 + X_c| x''(Y_a) \\ &= -|X_1 + X_c| x''(Y_a), \end{aligned} \tag{8.5.2}$$

since $x'(-X_c) = 0$. In the above, to simplify things, we have dropped the time dependence and defined $Y_a(t, X)$ such that $X_1 < Y_a < -X_c$. In addition, $x(t, X_1(t, X)) = x(t, X)$ implies that

$$x(-X_c) + \frac{1}{2}|X_1 + X_c|^2 x''(Y_b) = x(X_{cc}) - |X - X_{cc}| x'(Y_c),$$

$$\Longrightarrow |X_1 + X_c| = |X - X_{cc}|^{1/2} \left| \frac{2x'(Y_c)}{x''(Y_b)} \right|^{1/2} \tag{8.5.3}$$

since by definition of the cocaustic, $x(-X_c(t)) = x(X_{cc}(t))$. We also have $X_1 < Y_b(t, X) < -X_c$ and $X < Y_c(t, X) < X_{cc}$. Clearly, equations [8.5.2] and [8.5.3] determine exactly how quickly $(x'(X_1))^{-1}$ tends to infinity in the limit $X \longrightarrow X_c$. Inserting these equations into [8.4.2] gives

$$\frac{\partial^2 X_1}{\partial X^2}(X, t) = -\frac{x''(X)}{|X - X_{cc}|^{1/2} x''(Y_a)} \left| \frac{x''(Y_b)}{2x'(Y_c)} \right|^{1/2} + \frac{x'(X)^2 x''(X_1)}{|X - X_{cc}|^{3/2} x''(Y_a)^3} \left| \frac{x''(Y_b)}{2x'(Y_c)} \right|^{3/2}. \tag{8.5.4}$$

We do not need to concern ourselves with expanding the $x'(Y_c)$ since close to the cocaustic, x' is assumed to be non-zero and finite.

Of course it is not only the $(x'(X_1))^{-1}$ and $(x'(X_1))^{-3}$ terms that determine the singular behaviour of $\partial^2 X_1 / \partial X^2$, we also have the second order derivatives causing problems. The next stage, therefore, is to assume singular behaviour at both the caustic and cocaustic. The first assumption is possibly a little severe, however, since the analysis is not complicated any further by this assumption, we may as well include it. We assume that for Y_y close to the cocaustic, $x''(Y_y(t, X))$ for example, has the following dependency on X,

$$x''(Y_y(t, X)) = \sum_{n=0}^{\infty} \alpha_{(n)}(Y_y(t, X))|t - k_{cc}|^{n-p} = |t - k_{cc}|^{-p} \alpha_y(|t - k_{cc}|, X),$$

whereas for $Y_y < -X_c$ close to the caustic,

$$x''(Y_y(t,X)) = \sum_{n=0}^{\infty} \alpha_{(n)}(Y_y(t,X))|t-k_{cc}|^{n-q} = |t-k_{cc}|^{-q}\alpha_y(|t-k_{cc}|,X),$$

noting the replacement of p by q. Both of these are real constants satisfying $0 \le p,q < 1$. We draw attention to the α. These can be supposed to be analytic since we have pulled out the factors, $|t-k_{cc}|^{-p}$ and $|t-k_{cc}|^{-p}$, to represent the divergence of x'' in the two regions. We have also introduced a subscript y. This is simply a label that allows us to define a unique α_y according to the argument of x''. If α has no subscript then we take the argument of x'' to be X.

If we are supposing that α_y is analytic and arbitrary, then it does not make any sense to consider values of p and q less than zero since this replaces $\alpha_y(|t-k_{cc}|,X)$ by $|t-k_{cc}|\alpha_y(|t-k_{cc}|,X)$ for example, which is also analytic and so we remove this redundancy. The upper bound on p and q arise from the assumption that $\partial x/\partial X$ is continuous everywhere. To see this we write for X near the cocaustic,

$$\begin{aligned} x'(X) &= x'(X_{cc}) - |X-X_{cc}|x''(Y_d(t,X)) \\ &= x'(X_{cc}) - |t-k_{cc}|^{1-p}\alpha_d(|t-k_{cc}|,X)|k'_{cc}(Y_e)|^{-1}, \end{aligned}$$

where $X < Y_d(t,X), Y_e(t,X) < X_{cc}$ and the mean-value theorem has been used to write,

$$t - k_{cc}(X) = -(X - X_{cc}(t))k'_{cc}(Y_e).$$

A similar argument holds for X close to the caustic. Clearly then, for $p > 1$ the first derivative becomes unbounded at the cocaustic contradicting our assumptions.

Returning to the argument we have, upon using the mean value theorem,

$$\begin{aligned} \frac{\partial^2 X_1}{\partial X^2}(X,t) = &-\frac{|t-k_{cc}|^{-p}\alpha\,|k'_{cc}(Y_e)|^{1/2}}{|t-k_{cc}|^{1/2}|t-k_{cc}|^{-q}\alpha_a}\left|\frac{|t-k_{cc}|^{-q}\alpha_b}{2x'(Y_c)}\right| \\ &+\frac{x'(X)^2|t-k_{cc}|^{-q}\alpha_1\,|k'_{cc}(Y_e)|^{3/2}}{|t-k_{cc}|^{3/2}|t-k_{cc}|^{-3q}\alpha_a^3}\left|\frac{|t-k_{cc}|^{-q}\alpha_b}{2x'(Y_c)}\right|^{3/2} \end{aligned}$$

$$\begin{aligned} \Longrightarrow \frac{\partial^2 X_1}{\partial X^2}(X,t) = &-|t-k_{cc}|^{q/2-1/2-p}\frac{\alpha}{\alpha_a}\left|\frac{\alpha_b k'_{cc}(Y_e)}{2x'(Y_c)}\right|^{1/2} \\ &+|t-k_{cc}|^{q/2-3/2}x'(X)^2\frac{\alpha_1}{\alpha_a^3}\left|\frac{\alpha_b k'_{cc}(Y_e)}{2x'(Y_c)}\right|^{3/2}. \end{aligned} \qquad [8.5.5]$$

This represents the singular behaviour of $\partial^2 X_1/\partial X^2$ as $t \longrightarrow k_{cc}$ (or equivalently as $X \longrightarrow X_{cc}$). The coefficients of each of the $|t - k_{cc}|$ terms are continuous. This implies that we need k_{cc} to be at least C^1.

Let us now repeat the analysis and calculate an expression for $\partial^2 X_2/\partial X^2(X,t)$. Again for $X > 0$ near the cocaustic we expect X_2 to be close to $-X_c$. Using Taylor's theorem with remainder gives

$$x'(X_2) = |X_2 + k_{cc}| x'(Y_{\bar{a}}),$$

for $-X_c < Y_{\bar{a}}(t,X) < X_2$. Also, since $x(t, X_2(t,X)) = x(t,X)$,

$$|X_2 + X_c| = |X - X_{cc}|^{1/2} \left| \frac{2x'(Y_{\bar{c}})}{x''(Y_{\bar{b}})} \right|^{1/2},$$

where $-X_c < Y_{\bar{b}}(t,X) < X_2$ and $X < Y_{\bar{c}}(t,X) < X_{cc}$. Thus, using [8.4.2] we have

$$\frac{\partial^2 X_2}{\partial X^2}(X,t) = \frac{x''(X)}{|X - X_{cc}|^{1/2} x''(Y_{\bar{a}})} \left| \frac{x''(Y_{\bar{b}})}{2x'(Y_{\bar{c}})} \right|^{1/2} - \frac{x'(X)^2 x''(X_2)}{|X - X_{cc}|^{3/2} x''(Y_{\bar{a}})^3} \left| \frac{x''(Y_{\bar{b}})}{2x'(Y_{\bar{c}})} \right|^{3/2}. \qquad [8.5.6]$$

We note the similarities between the above and equation [8.5.4], which is due to the symmetry, $|X_1 + X_c| \approx |X_2 + X_c|$.

We now assume, for example, that

$$x''(Y_y(t,X)) = \sum_{n=0}^{\infty} \alpha_{(n)}(Y_y(t,X)) |t - k_{cc}|^{n-r} = |t - k_{cc}|^{-r} \alpha_y(|t - k_{cc}|, X)$$

in order to model the *possible* singular behaviour for some $Y_y > -X_c$ close to the caustic. With this assumption the above becomes

$$\frac{\partial^2 X_2}{\partial X^2}(X,t) = \frac{|t - k_{cc}|^{-p} \alpha \, |k'_{cc}(Y_e)|^{1/2}}{|t - k_{cc}|^{1/2} |t - k_{cc}|^{-r} \alpha_{\bar{a}}} \left| \frac{|t - k_{cc}|^{-r} \alpha_{\bar{b}}}{2x'(Y_{\bar{c}})} \right|^{1/2} - \frac{x'(X)^2 |t - k_{cc}|^{-r} \alpha_2 \, |k'_{cc}(Y_e)|^{3/2}}{|t - k_{cc}|^{3/2} |t - k_{cc}|^{-3r} \alpha_{\bar{a}}^3} \left| \frac{|t - k_{cc}|^{-r} \alpha_{\bar{b}}}{2x'(Y_{\bar{c}})} \right|^{3/2}$$

$$\Longrightarrow \frac{\partial^2 X_2}{\partial X^2}(X,t) = |t - k_{cc}|^{r/2 - 1/2 - p} \frac{\alpha}{\alpha_{\bar{a}}} \left| \frac{\alpha_{\bar{b}} k'_{cc}(Y_e)}{2x'(Y_{\bar{c}})} \right|^{1/2} - |t - k_{cc}|^{r/2 - 3/2} x'(X)^2 \frac{\alpha_2}{\alpha_{\bar{a}}^3} \left| \frac{\alpha_{\bar{b}} k'_{cc}(Y_e)}{2x'(Y_{\bar{c}})} \right|^{3/2}. \qquad [8.5.7]$$

Inserting [8.5.5] and [8.5.7] into [8.5.1] gives

$$\frac{\partial^2 x}{\partial X^2} = -6X + A\int_0^t\int_0^{t'}\Bigg\{ - |t'' - k_{cc}|^{q/2-1/2-p}\frac{\alpha}{\alpha_a}\left|\frac{\alpha_b k'_{cc}(Y_e)}{2x'(Y_c)}\right|^{1/2}$$
$$- |t'' - k_{cc}|^{r/2-1/2-p}\frac{\alpha}{\alpha_{\bar{a}}}\left|\frac{\alpha_{\bar{b}} k'_{cc}(Y_e)}{2x'(Y_{\bar{c}})}\right|^{1/2}$$
$$+ |t'' - k_{cc}|^{q/2-3/2}x'(X)^2\frac{\alpha_1}{\alpha_a^3}\left|\frac{\alpha_b k'_{cc}(Y_e)}{2x'(Y_c)}\right|^{3/2}$$
$$+ |t'' - k_{cc}|^{r/2-3/2}x'(X)^2\frac{\alpha_2}{\alpha_{\bar{a}}^3}\left|\frac{\alpha_{\bar{b}} k'_{cc}(Y_e)}{2x'(Y_{\bar{c}})}\right|^{3/2}\Bigg\}\, dt''dt'.$$

For convenience, we shall write the above as

$$\frac{\partial^2 x}{\partial X^2} = -6X + A\int_0^t\int_0^{t'} \Big\{|t'' - k_{cc}|^{q/2-1/2-p}\beta_1 + |t'' - k_{cc}|^{r/2-1/2-p}\beta_2$$
$$+|t'' - k_{cc}|^{q/2-3/2}\beta_3 + |t'' - k_{cc}|^{r/2-3/2}\beta_4\Big\}\, dt''dt',$$

or

$$\frac{\partial^2 x}{\partial X^2} = -6X + A\int^t\int^{t'} \Big\{|t'' - k_{cc}|^{q/2-1/2-p}\beta_1 + |t'' - k_{cc}|^{r/2-1/2-p}\beta_2$$
$$+|t'' - k_{cc}|^{q/2-3/2}\beta_3 + |t'' - k_{cc}|^{r/2-3/2}\beta_4\Big\}\, dt''dt' + \gamma,$$

where we have defined $\beta_i(|t-k_{cc}|, X)$ $(i = 1, \ldots, 4)$ to be the unknown analytic functions representing the coefficients of each $|t - k_{cc}|$ term in the above equation and $\gamma(|t - k_{cc}|, X)$, which is assumed to be analytic, to represent the functions of integration. If we now integrate this equation then provided t does not approach k_c we have

$$\frac{\partial^2 x}{\partial X^2} = -6X + A\Big\{|t - k_{cc}|^{q/2+3/2-p}\bar{\beta}_1 + |t - k_{cc}|^{r/2+3/2-p}\bar{\beta}_2$$
$$+|t - k_{cc}|^{q/2+1/2}\bar{\beta}_3 + |t - k_{cc}|^{r/2+1/2}\bar{\beta}_4\Big\} + \bar{\gamma}.$$

We again assume that the $\bar{\beta}_i(|t - k_{cc}|, X)$ and $\bar{\gamma}_i(|t - k_{cc}|, X)$ are analytic. Now we observe that the left hand side is $O(|t-k_{cc}|^{-p})$ and so for $p \neq 0$, we have a divergence as $t \longrightarrow k_{cc}$. On the other hand the terms on the right are always finite. We conclude therefore, that $p = 0$ and consequently that x'' is continuous near the cocaustic. This is remarkable because our worry was that this would not be the case, forcing us to use a complicated, weighted norm to define our metric. Instead the above result implies that we can use a norm that looks like

$$\|x\| = \sup_{(t,X)\in\mathcal{D}} \left\{ \left|\frac{\partial^2 x}{\partial X^2}\right| + \left|\frac{\partial^2 x}{\partial X^2}\right| + \left|\frac{\partial^2 x}{\partial t^2}\right| + \left|\frac{\partial^2 x}{\partial X \partial t}\right| \right\},$$

for some $\mathcal{D}$, or at least a derivative of this.

All things now point towards a function space that is C^2 in both the t and X coordinate. We must, however, check that there are no surprises at our other suspected problem area, the caustic.

§8.6. Near the caustic, $X > X_c$.

Since $X \geq X_c$, equation [8.5.1] is still valid. We shall consider the two regions of tX space separated by $t = k_c(X)$ to have a different dependence on $|t - k_c|$. That is to say, we shall assume singular behaviour at the caustic in much the same way as in the previous chapter, but allow the rates at which x'' diverges either side of this curve to be different. This was essentially done in §8.5 when we introduced the different q and r.

We begin using the same Taylor expansion methods described in the previous section and consider first of all $\partial^2 X_1/\partial X^{2^2}$. Since X is now close to the caustic, it follows from the definition of the X_i that X_1 must be close to the cocaustic. Thus in this case, both $x''(X_1)$ and $x'(X_1)$ are expected to be finite and non-zero (we have just proved this). The behaviour of $x'(X)$ can be determined. We have

$$x'(X) = |X - X_c| x''(Y_a),$$

where $X_c < Y_a(t, X) < X$ is different to the Y_a introduced in the previous section as are the Y_b's Y_c's etc. that will follow. This means that

$$\frac{\partial^2 X_1}{\partial X^2} = \frac{x''(X)}{x'(X_1)} - \frac{|X - X_c|^2 x''(Y_a)^2 x''(X_1)}{x'(X_1)^3}$$

If $Y_y > X_c$ is close to the caustic then we illustrate the possible unbounded nature of the second derivative in this region by writing,

$$x''(Y_y) = |t - k_c|^{-s} \alpha_y(|t - k_c|, X) \quad 0 < s < 1.$$

Then,

$$\frac{\partial^2 X_1}{\partial X^2} = |t - k_c|^{-s}\frac{\alpha}{x'(X_1)} - |t - k_c|^{2-2s}\frac{\alpha_a^2 x''(X_1)}{|k_c'(Y_b)|^2 x'(X_1)^3}, \qquad [8.6.1]$$

where again the subscripts to each α indicate the arguments for the original x'' from which it is derived. Also we have used the mean value theorem to write,

$$k_c(X) - t = -(X - X_c(t))k_c'(Y_b).$$

Here $X_c < Y_b(t, X) < X$.

Let us now consider $\partial^2 X_2/\partial X^2$. We have, remembering that for $X \geq X_c$ close to the caustic X_2 must also be close to the caustic,

$$x'(X_2) = -|X_2 - X_c|x''(Y_c),$$

where $X_2 < Y_c(t, X) < X_c$. Also, from $x(t, X_i(t, X)) = x(t, X)$, we have

$$x(X_c) + \frac{1}{2}|X_2 - X_c|^2 x''(Y_d) = x(X_c) + \frac{1}{2}|X - X_c|^2 x''(Y_e),$$

where $X_2 < Y_d(t, X) < X_c$ and $X_c < Y_e(t, X) < X$. This implies that

$$|X_2 - X_c| = |X - X_c|\left|\frac{x''(Y_e)}{x''(Y_d)}\right|^{1/2}.$$

Thus it follows that

$$\frac{\partial^2 X_2}{\partial X^2} = -\frac{x''(X)}{|X - X_c|x''(Y_c)}\left|\frac{x''(Y_d)}{x''(Y_e)}\right|^{1/2} + \frac{|X - X_c|^2 x''(Y_a)^2 x''(X_2)}{|X - X_c|^3 x''(Y_c)^3}\left|\frac{x''(Y_d)}{x''(Y_e)}\right|^{3/2}.$$

We are again required to feed in our assumptions regarding the expected singular behaviour. For any $Y_y < X_c$ close to the caustic we assume that the second derivative can be written as

$$x''(Y_y) = |t - k_c|^{-t}\alpha_y(|t - k_c|, X) \quad 0 < t < 1.$$

Thus,

$$\frac{\partial^2 X_2}{\partial X^2} = -\frac{|t - k_c|^{-s}\alpha|k_c'(Y_b)|}{|t - k_c||t - k_c|^{-t}\alpha_c}\left|\frac{|t - k_c|^{-t}\alpha_d}{|t - k_c|^{-s}\alpha_e}\right|^{1/2}$$
$$+ \frac{|k_c'(Y_b)||t - k_c|^{-2s}\alpha_a^2|t - k_c|^{-t}\alpha_2}{|t - k_c||t - k_c|^{-3t}\alpha_c^3}\left|\frac{|t - k_c|^{-t}\alpha_d}{|t - k_c|^{-s}\alpha_e}\right|^{3/2},$$

which finally becomes

$$\frac{\partial^2 X_2}{\partial X^2} = -|t-k_c|^{t/2-s/2-1}\frac{\alpha|k_c'(Y_b)|}{\alpha_c}\left|\frac{\alpha_d}{\alpha_e}\right|^{1/2} + |t-k_c|^{t/2-s/2-1}\frac{|k_c'(Y_b)|\alpha_a^2\alpha_2}{\alpha_c^3}\left|\frac{\alpha_d}{\alpha_e}\right|^{3/2}. \tag{8.6.2}$$

Inserting [8.6.1] and [8.6.2] into equation [8.4.1] gives

$$\begin{aligned}\frac{\partial^2 x}{\partial X^2} = -6X + A\int_0^t\int_0^{t'}\Bigg\{&|t''-k_c|^{-s}\frac{\alpha}{x'(X_1)} - |t''-k_c|^{2-2s}\frac{\alpha_a^2 x''(X_1)}{|k_c'(Y_b)|^2 x'(X_1)^3}\\ &+|t''-k_c|^{t/2-s/2-1}\frac{\alpha|k_c'(Y_b)|}{\alpha_c}\left|\frac{\alpha_d}{\alpha_e}\right|^{1/2}\\ &-|t''-k_c|^{t/2-s/2-1}\frac{|k_c'(Y_b)|\alpha_a^2\alpha_2}{\alpha_c^3}\left|\frac{\alpha_d}{\alpha_e}\right|^{3/2}\Bigg\}\,dt''dt',\end{aligned}$$

which we shall simplify by writing as

$$\begin{aligned}\frac{\partial^2 x}{\partial X^2} = -6X + A\int^t\int^{t'}\Big\{&|t''-k_c|^{-s}\beta_1 - |t''-k_c|^{2-2s}\beta_2\\ &+|t''-k_c|^{t/2-s/2-1}\beta_3\Big\}\,dt''dt' + \gamma.\end{aligned}$$

Here the β_i and γ are different functions to that defined in the previous section. They do, however, possess the same attributes, namely that they are analytic functions, and in particular, that γ still represents the arbitrary function of integration. Since x is C^2 in a region close to the caustic, the above representation holds for all $t < k_c$. Thus, if we integrate we obtain

$$\frac{\partial^2 x}{\partial X^2} = -6X + A\Big\{|t-k_c|^{-s+2}\bar\beta_1 - |t-k_c|^{4-2s}\bar\beta_2 + |t-k_c|^{t/2-s/2+1}\bar\beta_3\Big\} + \bar\gamma.$$

But again for $s \neq 0$ the left hand side is $O(|t-k_c|^{-s})$ and therefore unbounded as $t \longrightarrow k_c$ whereas the right hand side is finite. It follows that for consistency, $s = 0$ and we conclude that x'' is continuous for $X \geq X_c$.

§8.7. Near the caustic, $X < X_c$.

For the other side of the caustic we proceed in the same manner. Things are getting progressively easier for as we proceed to integrate expressions for x'' up the time axis, we seem to be proving that x is C^2 at all points below (t, X). The only function whose continuity is not known is $x''(Y_y)$ for $Y_y < X_c$ close to the caustic. We have, using Taylor's theorem,

$$x'(X) = -|X - X_c|x''(Y_a),$$

where $X < Y_a(t, X) < X_c$. Assuming that all second order derivatives this side of the caustic behave like

$$x''(Y_y) = \sum_{n=0}^{\infty} \alpha_{(n)}(Y_y)|t - k_c|^{n-u} = |t - k_c|^{-u}\alpha_y,$$

then for $\partial^2 X_i/\partial X^2$ (remembering that now $X = X_2$),

$$\frac{\partial^2 X_1}{\partial X^2} = |t - k_c|^{-u}\frac{\alpha}{x'(X_1)} - |t - k_c|^{2-2u}\frac{\alpha_a^2 x''(X_1)}{|k_c'(Y_b)|^2 x'(X_1)}, \qquad [8.7.1]$$

where we have used the mean value theorem to write,

$$t - k_c = -(X - X_c)k_c'(Y_b),$$

where $X < Y_b(t, X) < X_c$.

Similarly, we have from Taylor's theorem that

$$x'(X_3) = |X_3 - X_c|x''(Y_c)$$

and

$$|X_3 - X_c| = |X - X_c|\left|\frac{x''(Y_e)}{x''(Y_d)}\right|^{1/2}.$$

Thus it follows that

$$\frac{\partial^2 X_3}{\partial X^2} = \frac{|t - k_c|^{-u}\alpha|k_c'(Y_b)|}{|t - k_c|x''}\left|\frac{x''(Y_d)}{|t - k_c|^{-u}\alpha_e}\right|^{1/2} - \frac{|t - k_c|^2|k_c(Y_b)||t - k_c|^{-2u}\alpha_a x''(X_3)}{|t - k_c|^3 x''(Y_c)^3}\left|\frac{x''(Y_d)}{|t - k_c|^{-u}\alpha_e}\right|^{3/2}$$

$$\Longrightarrow \frac{\partial^2 X_3}{\partial X^2} = |t - k_c|^{-u/2-1}\frac{\alpha|k_c'(Y_b)|}{x''(Y_c)}\left|\frac{x''(Y_d)}{\alpha_e}\right|^{1/2} - |t - k_c|^{-u/2-1}\frac{|k_c(Y_b)|\alpha_a x''(X_3)}{x''(Y_c)^3}\left|\frac{x''(Y_d)}{\alpha_e}\right|^{3/2}. \qquad [8.7.2]$$

Finally, inserting [8.7.1] and [8.7.2] into [8.4.1] gives

$$\frac{\partial^2 x}{\partial X^2} = -6X + A\int_0^t\int_0^{t'}\left\{|t''-k_c|^{-u}\frac{\alpha}{x'(X_1)} - |t''-k_c|^{2-2u}\frac{\alpha_a^2 x''(X_1)}{|k_c'(Y_b)|^2 x'(X_1)}\right.$$
$$+|t''-k_c|^{-u/2-1}\frac{\alpha|k_c'(Y_b)|}{x''(Y_c)}\left|\frac{x''(Y_d)}{\alpha_e}\right|^{1/2}$$
$$\left.-|t''-k_c|^{-u/2-1}\frac{|k_c(Y_b)|\alpha_a x''(X_3)}{x''(Y_c)^3}\left|\frac{x''(Y_d)}{\alpha_e}\right|^{3/2}\right\}dt''dt'.$$

For simplicity we write this as

$$\frac{\partial^2 x}{\partial X^2} = -6X + A\int^t\int^{t'}\left\{|t''-k_c|^{-u}\beta_1 - |t''-k_c|^{2-2u}\beta_2\right.$$
$$\left.+|t''-k_c|^{-u/2-1}\beta_3\right\}dt''dt' + \gamma$$

$$\Longrightarrow \frac{\partial^2 x}{\partial X^2} = -6X + A\left\{|t-k_c|^{2-u}\bar\beta_1 - |t-k_c|^{4-2u}\bar\beta_2 + |t-k_c|^{1-u/2}\bar\beta_3\right\} + \bar\gamma.$$

Since the left hand side is $O(|t-k_c|^{-u})$ we conclude that to have consistency we must have $u=0$. The implications of this and the previous two section's conclusions imply that x is C^2 everywhere in the three-dust region. This will have profound implications on the simplicity of both the metric space that will be used and the actual calculation proving that J is a contraction.

§8.8. Metric space proposed for a contraction mapping proof of existence.

In the previous three sections we have shown that $\partial^2 x/\partial X^2$ is continuous everywhere in the three dust region. The fact that x is now C^2 means that we can focus our attention on a relatively simple formulation of the metric function. By this we mean something that looks like

$$d(f_1,f_2) = \|f_1 - f_2\|,$$

where

$$\|f\| = \sup_{(t,X)\in\mathcal{D}}\left\{\left|\frac{\partial^2\phi}{\partial X^2}\right| + \left|\frac{\partial^2\phi}{\partial t^2}\right| + \left|\frac{\partial^2\phi}{\partial X\partial t}\right|\right\},$$

which is the standard C^2 norm.

First of all we define the domain in which we hope to prove existence, i.e.,

$$B_T = \left\{(t, X) \in \mathbb{R}^2 \mid |X| < T, 0 \le t \le T\right\},$$

which, for suitably small $T > 0$, defines a box enclosing the cusp. Next, to model our assumption that the solution *looks like* q, which is the solution for zero gravitational constant, we define

$$\mathcal{V} = \left\{f: B_T \longrightarrow \mathbb{R} \mid f \text{ is } C^2, \quad f(t,X) = -f(t,-X), \quad \tfrac{\partial f}{\partial X}(0,0) = 0, \quad \|f\|_{\mathcal{V}} < \infty\right\}$$

as the space of functions within which we hope to prove existence. The norm defined in the above is given by

$$\|f\|_{\mathcal{V}} = \sup_{(t,X)\in B_T} \left\{\frac{1}{|X|}\left|\frac{\partial^2 \phi}{\partial X^2}\right| + \left|\frac{\partial^2 \phi}{\partial t^2}\right| + \left|\frac{\partial^2 \phi}{\partial X \partial t}\right|\right\}. \qquad [8.8.1]$$

If we restrict this further to the set of functions given by

$$\mathcal{V}_c = \{f \in \mathcal{V} \mid \|f - q\|_{\mathcal{V}} < c\},$$

then this confines our test solution, f, to a tube around q. The 'width' of this tube is determined by the constant c.

The above definitions embody all of the ideas so far presented in this chapter, however, the thoughts that led to the exact form of equation [8.8.1] have not been presented. We shall now do so. In defining our metric space we postulated that for a small enough region containing the cusp, any trial solution must be close to $x = q(t, X)$ where q is the solution for zero gravitational constant (ref. §3.1). To put meaning to this statement we required that

1. x is an odd function in X (as is q) and
2. that the bounds to be used for $D^2(x - q)$ should be linear in $|X|$.

Item 2 suggests that for $X \ge 0$,

$$\left|\frac{\partial^2 x}{\partial X^2} - \frac{\partial^2 q}{\partial X^2}\right| \le a_1 + a_2|X|, \qquad [8.8.2]$$

$$\left|\frac{\partial^2 x}{\partial X \partial t} - \frac{\partial^2 q}{\partial X \partial t}\right| \leq b_1 + b_2|X|, \qquad [8.8.3]$$

$$\left|\frac{\partial^2 x}{\partial t^2} - \frac{\partial^2 q}{\partial t^2}\right| \leq c_1 + c_2|X|. \qquad [8.8.4]$$

To show that these conditions imply bounds on x that are cubic in nature we must integrate the above and to do this we note that for all $f \in \mathcal{V}_c$,

$$\frac{\partial f}{\partial X}(t, X) = \int_0^X \frac{\partial^2 f}{\partial X^2}(0, X')\, dX' + \int_0^t \frac{\partial^2 f}{\partial t \partial X}(t', X)\, dt'$$

and

$$f(t, X) = \int_0^X \frac{\partial f}{\partial X}(t, X')\, dX'.$$

Thus, for $X > 0$ (we can obtain estimates for the $X < 0$ case by symmetry), equations [8.8.2] and [8.8.3] imply that

$$\begin{aligned}
\left|\frac{\partial f}{\partial X}(t, X) - \frac{\partial q}{\partial X}(t, X)\right| &\leq \int_0^X \left|\frac{\partial^2 f}{\partial X^2}(0, X') - \frac{\partial^2 q}{\partial X^2}(0, X')\right| dX' \\
&\quad + \int_0^t \left|\frac{\partial^2 f}{\partial t \partial X}(t', X) - \frac{\partial^2 q}{\partial t \partial X}(t', X)\right| dt' \\
&\leq \int_0^X a_1 + a_2 X\, dX' + \int_0^t b_1 + b_2 X\, dt' \\
&= a_1 X + \frac{1}{2} a_2 X^2 + b_1 t + b_2 X t.
\end{aligned}$$

This in turn gives

$$\begin{aligned}
|f(t, X) - q(t, X)| &\leq \int_0^X \left|\frac{\partial f}{\partial X}(t, X') - \frac{\partial q}{\partial X}(t, X')\right| dX' \\
&= \frac{1}{2} a_1 X^2 + \frac{1}{6} a_2 X^3 + b_1 X t + \frac{1}{2} b_2 X^2 t.
\end{aligned}$$

Item 1 in the above list forces $a_1 = b_2 = 0$. This finally gives,

$$(1 - b_1)Xt - \left(1 + \frac{a_2}{6}\right) X^3 \leq f(t, X) \leq (1 + b_1)Xt - \left(1 - \frac{a_2}{6}\right) X^3$$

for $X > 0$ and provided $a_2 < 6$. Thus for $X > 0$, f is bounded above and below by a cubic that looks like q. We can refine equations [8.8.2]–[8.8.4] still further since we can set $c_2 = 0$ without affecting the cubic structure of our upper and lower bounds. In addition, if we set $c = \max\{a_2, b_1, c_1\}$ then [8.8.2]–[8.8.4] become equivalent to $\|f - q\| < c$.

CHAPTER 9. ESTIMATES FOR $\partial^2\chi[f]/\partial X^2$ IN MULTI-DUST REGION.

§9.1. Introduction.

The next step is to show that J, as defined by [8.3.3], is a contraction mapping on the space of functions, $\mathcal{V}_c$. To do this, we need to estimate $\|J[f]\|_{\mathcal{V}}$. The only difficulty arises with the second X-derivative given by equation [8.4.1] which we restate below,

$$\frac{\partial^2 J[f]}{\partial X^2}(t,X) = -6X + A\int_0^t \int_0^{t'} \frac{\partial^2\chi[f]}{\partial X^2}(t'',X)\,dt''dt', \qquad [8.4.1]$$

where

$$\frac{\partial^2\chi[f]}{\partial X^2}(t,X) = \frac{\partial^2 X_1}{\partial X^2}(t,X) - \frac{\partial^2 X_2}{\partial X^2}(t,X) + \frac{\partial^2 X_3}{\partial X^2}(t,X)$$

and

$$\frac{\partial^2 X_i}{\partial X^2}(t,X) = \frac{f''(X)}{f'(X_i)} - \frac{f'(X)^2 f''(X_i)}{f'(X_i)^3}.$$

The last equation represents the difficulty entirely for it contains two divergent terms, $[f'(X_i)]^{-1}$ and $[f'(X_i)]^{-3}$. We complete this thesis by constructing estimates for the first of these quantities.

§9.2. Bounds for the caustic and its derivative.

In this section estimates for the caustic, $k_c(X)$, and its inverse, $X_c(t)$, will be obtained. We shall do this by first of all calculating bounds on the derivatives, $k_c'(X)$ and $X_c'(t)$, and then integrating these using the boundary conditions $k_c(0) = X_c(0) = 0$. This results in expressions that bound k_c and X_c away from zero; an important requirement for later on when $[f'(t,X_i)]^{-1}$ is considered as a function of t and X, terms like k_c^{-1} will appear.

We begin by considering the definition of the caustic, $f'(k_c,X) = 0$. By differentiating this expression k_c' can be defined in terms of 2nd order derivatives of f, i.e.,

$$k_c'(X) = -\frac{\partial^2 f/\partial X^2\,(k_c,X)}{\partial^2 f/\partial t\partial X\,(k_c,X)}.$$

Previous analysis has discussed the need to confine f about q, where $q = Xt - X^3$ is the cubic corresponding to caustic formation in the absence of gravity. This was achieved by stating that $\|f - q\|_{\mathcal{V}} \leq c$ resulting in

$$\left|\frac{\partial^2 f}{\partial X^2}(t,X) - \frac{\partial^2 q}{\partial X^2}(t,X)\right| \leq c|X| \quad \text{and} \quad \left|\frac{\partial^2 f}{\partial t \partial X}(t,X) - \frac{\partial^2 q}{\partial t \partial X}(t,X)\right| \leq c, \quad [9.2.1]$$

where c is an unrestricted constant that defines the size of our tube containing f. By the symmetry of f about the origin we only need consider that half of the tX plane corresponding to $X \geq 0$. From equations [9.2.1] we therefore have

$$-(6+c)X \leq \frac{\partial^2 f}{\partial X^2}(t,X) \leq -(6-c)X$$

and

$$1 - c \leq \frac{\partial^2 f}{\partial t \partial X}(t,X) \leq 1 + c.$$

Clearly then,

$$\frac{6-c}{1+c}X \leq k_c' \leq \frac{6+c}{1-c}X. \quad [9.2.2]$$

Integrating this expression between 0 and X with the boundary condition, $k_c(0) = 0$, gives

$$\int_0^X \frac{6-c}{1+c}X'\,dX' \leq \int_0^X k_c'(X')\,dX' \leq \int_0^X \frac{6+c}{1-c}X'\,dX'$$

$$\Longrightarrow \frac{6-c}{2\,(1+c)}X^2 \leq k_c \leq \frac{6+c}{2\,(1-c)}X^2. \quad [9.2.3]$$

To obtain estimates for X_c and X_c' we can use the inverse function theorem to bound X_c' from above and below and then follow a similar argument to the above. Now,

$$X_c'(t) = \left(k_c'(X_c(t))\right)^{-1}$$

$$\Longrightarrow \int_0^t \frac{1-c}{6+c}\,dt' \leq \int_0^t X_c'(t')X_c(t')\,dt' \leq \int_0^t \frac{1+c}{6-c}\,dt'$$

$$\Longrightarrow \sqrt{\frac{2(1-c)}{6+c}t} \leq X_c \leq \sqrt{\frac{2(1+c)}{6-c}t}, \quad [9.2.4]$$

using the boundary condition, $X_c(0) = 0$. Finally,

$$\frac{1-c}{6+c}\sqrt{\frac{6-c}{2(1+c)t}} \leq X_c' \leq \frac{1+c}{6-c}\sqrt{\frac{6+c}{2(1-c)t}}. \quad [9.2.5]$$

§9.3. Bounds for the cocaustic and its derivative.

We have already defined the equation for the cocaustic in tX space to be $t = k_{cc}(X)$ with its inverse, $X = X_{cc}(t)$. In this section bounds for both of these functions and their derivatives will be obtained. The argument will follow along similar lines as in the previous section: estimates for the derivatives will be obtained which will then be integrated to give estimates for the functions themselves. The analysis is complicated by the fact that we can only define the first derivative of the cocaustic as a composite function involving f and $X_c(t)$. Having to estimate the derivatives of both of these functions therefore results in bounds for the cocaustic that are perhaps less stringent than those for the caustic.

It is possible to find bounds for the second derivative of the cocaustic, however, this is a lengthy procedure and cannot be used to improve upon the estimates for the first derivative of the cocaustic or the function itself. Moreover it does not feature anywhere in the evaluation of estimates for $\partial^2 X_i / \partial X^2$.

To begin we have, by definition,

$$f\left(k_{cc}(X), -X_c(k_{cc}(X))\right) = f\left(k_{cc}(X), X\right).$$

Differentiating this gives

$$\begin{aligned} -\frac{\partial f}{\partial X}\left(k_{cc}, -X_c(k_{cc})\right) X_c'(k_{cc}) k_{cc}' + \frac{\partial f}{\partial t}\left(k_{cc}, -X_c(k_{cc})\right) k_{cc}' &= \frac{\partial f}{\partial X}\left(k_{cc}, X\right) \\ + \frac{\partial f}{\partial t}\left(k_{cc}, X\right) k_{cc}'. \end{aligned} \qquad [9.3.1]$$

The first term vanishes by the definition of the caustic and so

$$k_{cc}'(X) = -\frac{\partial f / \partial X\left(k_{cc}, X\right)}{\partial f / \partial t\left(k_{cc}, X\right) - \partial f / \partial t\left(k_{cc}, -X_{cc}(k_{cc})\right)}. \qquad [9.3.2]$$

By virtue of the assumption that f is confined to some 'tube' centred upon q (as defined above), we can show that

$$\left|\frac{\partial f}{\partial X}(t, X) - \frac{\partial q}{\partial X}(t, X)\right| \leq ct + \frac{1}{2} c X^2$$

and

$$\left|\frac{\partial f}{\partial t}(t,X) - \frac{\partial q}{\partial t}(t,X)\right| \leq c|X| \qquad [9.3.3]$$

(ref. §8.8), which are equivalent to

$$(1-c)\,t - \frac{1}{2}(6+c)X^2 \leq \frac{\partial f}{\partial X} \leq (1+c)\,t - \frac{1}{2}(6-c)X^2 \quad \forall X$$

and

$$(1-c)\,X \leq \frac{\partial f}{\partial t} \leq (1+c)\,X \quad \forall X \geq 0,$$

or

$$(1+c)\,X \leq \frac{\partial f}{\partial t} \leq (1-c)\,X \quad \forall X \leq 0.$$

For the moment, let us only consider an upper bound for k_c'. Now,

$$-\frac{\partial f}{\partial X}(k_{cc}, X) \leq -(1-c)\,k_{cc} + \frac{1}{2}(6+c)X^2$$

and

$$\begin{aligned}\frac{\partial f}{\partial t}(k_{cc}, X) - \frac{\partial f}{\partial t}(k_{cc}, -X_c(k_{cc})) &\geq (1-c)X - (1-c)\,(-X_c(k_{cc})) \\ &\geq (1-c)\left\{X + \sqrt{\frac{2(1-c)}{(6+c)}k_{cc}}\right\}.\end{aligned}$$

Thus from [9.3.2],

$$\begin{aligned}k_{cc}' &\leq \frac{6+c}{2(1-c)}\left\{\frac{X^2 - 2(1-c)(6+c)^{-1}k_{cc}}{X + \sqrt{2(1-c)(6+c)^{-1}k_{cc}}}\right\} \\ &= \frac{6+c}{2(1-c)}\left\{X - \sqrt{\frac{2(1-c)}{6+c}k_{cc}}\right\} \\ &\leq \frac{6+c}{2(1-c)}X.\end{aligned}$$

Integrating this equation with respect to X and using the boundary condition for the cocaustic at the origin, namely $k_{cc}(0) = 0$, gives

$$k_{cc} \leq \frac{6+c}{4(1-c)}X^2. \qquad [9.3.4]$$

For a bound going the other way,

$$-\frac{\partial f}{\partial X}(k_{cc}, X) \geq -(1+c)k_{cc} + \frac{1}{2}(6-c)X^2$$

and

$$\begin{aligned}\frac{\partial f}{\partial t}(k_{cc}, X) - \frac{\partial f}{\partial t}(k_{cc}, -X_c(k_{cc})) &\leq (1+c)X - (1+c)\left(-X_c(k_{cc})\right) \\ &\leq (1+c)\left\{X + \sqrt{\frac{2(1+c)}{6-c}k_{cc}}\right\}\end{aligned}$$

together imply that

$$k'_{cc} \geq \frac{6-c}{2(1+c)}\left\{1 - \sqrt{\frac{(1+c)(6+c)}{2(1-c)(6-c)}}\right\}X,$$

using equation [9.3.4].

We can think of this estimate as being the most general lower bound for k'_{cc}. However, this expression must be used to determine bounds for the inverse function, X_{cc}, and it would be desirable if we could simplify it. c is an arbitrary constant such that $0 < c < 1$ (the condition that $\partial f/\partial t$ must still exhibit q-like behaviour, ref. [9.3.3]) and essentially determines the thickness of the 'tube' that confines f close to q. Reducing the value of this constant tightens this tube and improves all the estimates for f, the caustic, the cocaustic and all the corresponding derivatives. We shall find later on when bounds for $[f'(X_i(t,X))]^{-1}$ are calculated for various regions in the positive half of the tX plane, that the range of admissible values for c must be reduced. Consequently there is no reason to stop us from modifying this range even at this early stage, so as to simplify the lower bound for k'_{cc}.

We suppose therefore, that c satisfies,

$$\frac{(1+c)(6+c)}{2(1-c)(6-c)} < \frac{5}{9}. \qquad [9.3.5]$$

With this choice we have $1 - \sqrt{5/9} > 1/4$ so that

$$k'_{cc} > \frac{6-c}{8(1+c)}X. \qquad [9.3.6]$$

Now equation [9.3.5] clearly restricts the value of c. The extent to which it does this can be determined by solving $c^2 - 133c + 6 > 0$, which is equivalent to [9.3.5]. Since $x^2 - 133x + 6 = 0$ has solutions,

$$x = \frac{133 \pm \sqrt{17665}}{2},$$

we can say that provided

$$\frac{133 - \sqrt{17665}}{2} = \frac{1596 - \sqrt{2543760}}{24} > \frac{1596 - \sqrt{2544025}}{24} = \frac{1}{24} > c,$$

then both [9.3.5] and [9.3.6] are true statements. This means that we can write,

$$\frac{6-c}{8(1+c)}X < k'_{cc} \leq \frac{6+c}{2(1-c)}X, \qquad [9.3.7]$$

which if we integrate using the assumption that $k_{cc}(0) = 0$ gives

$$\frac{6-c}{16(1+c)}X^2 < k_{cc} \leq \frac{6+c}{4(1-c)}X^2. \qquad [9.3.8]$$

Again, to find estimates for X_{cc} and its derivative we use the inverse function theorem. Equation [9.3.7] therefore implies that

$$\frac{2(1-c)}{(6+c)X_{cc}} \leq X'_{cc} < \frac{8(1+c)}{(6-c)X_{cc}}. \qquad [9.3.9]$$

Integrating this with the boundary condition $X_{cc}(0) = 0$ gives

$$\int_0^t \frac{2(1-c)}{6+c}\,dt' \leq \int_0^t X_{cc}(t')X'_{cc}(t')\,dt' < \int_0^t \frac{8(1+c)}{6-c}\,dt'$$

$$\Longrightarrow \sqrt{\frac{4(1-c)}{6+c}t} \leq X_{cc} < \sqrt{\frac{16(1+c)}{6-c}t}. \qquad [9.3.10]$$

§9.4. Summary of estimates for the caustic and cocaustic.

In order to facilitate reference to the estimates for the caustic and co-caustic we list them in this section:

$$\frac{6-c}{1+c}X \le k_c' \le \frac{6+c}{1-c}X, \tag{9.2.2}$$

$$\frac{6-c}{2(1+c)}X^2 \le k_c \le \frac{6+c}{2(1-c)}X^2, \tag{9.2.3}$$

$$\frac{1-c}{6+c}\sqrt{\frac{6-c}{2(1+c)t}} \le X_c' \le \frac{1+c}{6-c}\sqrt{\frac{6+c}{2(1-c)t}}, \tag{9.2.5}$$

$$\sqrt{\frac{2(1-c)}{6+c}t} \le X_c \le \sqrt{\frac{2(1+c)}{6-c}t}, \tag{9.2.4}$$

$$\frac{6-c}{8(1+c)}X < k_{cc}' \le \frac{6+c}{2(1-c)}X, \tag{9.3.7}$$

$$\frac{6-c}{16(1+c)}X^2 < k_{cc} \le \frac{6+c}{4(1-c)}X^2, \tag{9.3.8}$$

$$\frac{2(1-c)}{(6+c)}\sqrt{\frac{6-c}{16(1+c)t}} \le X_{cc}' < \frac{8(1+c)}{(6-c)}\sqrt{\frac{6+c}{4(1-c)t}}, \tag{9.3.9}$$

$$\sqrt{\frac{4(1-c)}{6+c}t} \le X_{cc} < \sqrt{\frac{16(1+c)}{6-c}t}. \tag{9.3.10}$$

§9.5. Estimates for $[f'(X_1(t,X))]^{-1}$ for X near the cocaustic.

We begin our analysis on $[f'(X_i(t,X))]^{-1}$ by considering the case when X is close to the cocaustic and $i = 1$. By definition of the cocaustic we expect X_1 to be close to the caustic and so $[f'(X_1(X,t;f))]^{-1}$ becomes unbounded as X approaches X_{cc}. To illustrate this feature we use the mean value theorem three times to express this quantity in terms of the distance from the cocaustic, $|t - k_{cc}|$.

By definition,

$$f(X_1) = f(X)$$
$$\Longrightarrow f(X_{cc}) - |X - X_{cc}|f'(Y_a) = f(-X_c) + \frac{1}{2}|X_1 + X_c|^2 f''(Y_b),$$

where $X < Y_a < X_{cc}$ and $X_1 < Y_b < -X_c$. It follows that

$$|X_1 + X_c| = \frac{\sqrt{2}|f'(Y_a)|^{1/2}|X - X_{cc}|^{1/2}}{|f''(Y_b)|^{1/2}}.$$

Also,

$$f'(X_1) = f'(-X_c) - |X_1 + X_c|f''(Y_c),$$

where $X_1 < Y_c < -X_c$ and

$$|t - k_{cc}| = |X - X_{cc}||k'_{cc}(Y_d)|,$$

where $X < Y_d < X_{cc}$. Combining these three results gives the expression,

$$\begin{aligned}[f'(X_1)]^{-1} &= -\frac{|f''(Y_b)|^{1/2}|k'_{cc}(Y_d)|^{1/2}}{\sqrt{2}f''(Y_c)|f'(Y_a)|^{1/2}|t - k_{cc}|^{1/2}} \\ &= -\frac{|f''(Y_b)|^{1/2}|k'_{cc}(Y_d)|^{1/2}}{\sqrt{2}|f''(Y_c)||f'(Y_a)|^{1/2}|t - k_{cc}|^{1/2}}, \qquad [9.5.1]\end{aligned}$$

which we can use as a basis to begin the estimation process. The procedure is to simply take each term and use either

$$|f''(X) - q''(X)| \le c|X| \qquad [9.5.2]$$

or

$$|f'(X) - q'(X)| \le ct + \frac{1}{2}cX^2 \qquad [9.5.3]$$

to obtain upper and lower bounds.

Obtaining bounds for any term that is the second derivative of f is easy since $0 < c < 1$ restricts the 'tube' containing f'' so that it is linear and strictly negative for positive X. The first derivative, however, is much more difficult and this is chiefly because f' becomes zero at the caustic. This must be avoided if finite estimates for $[f'(X_1)]^{-1}$ are required. As we shall see, the way around this problem is to impose restrictions on X so that $|f'(Y_a)|$ is bounded away from zero. This of course begins to define in specific terms what we mean by the region close to the cocaustic.

Consider first of all the term $|f''(Y_b)|^{1/2}$. Now equation [9.5.2] implies that

$$(6 - c)|Y_b| \le |f''(Y_b)| \le (6 + c)|Y_b|,$$

but since $X_1 < Y_b < -X_c$ we can make the above estimates more restrictive, i.e.,

$$(6-c)X_c \leq |f''(Y_b)| \leq (6+c)X_{cc}. \quad [9.5.4]$$

In a similar fashion the term $|f''(Y_c)|$ can also be estimated; we obtain

$$(6-c)X_c \leq |f''(Y_c)| \leq (6+c)X_{cc}. \quad [9.5.5]$$

The term $|f'(Y_a)|^{1/2}$ is treated differently. Since this occurs in the denominator we must ensure that this term is bounded away from zero. Now from equation [9.5.3],

$$\begin{aligned} f'(Y_a) &< q'(Y_a) + ct + \frac{1}{2}cY_a^2 \\ &= (1+c)t - \frac{1}{2}(6-c)Y_a^2. \end{aligned} \quad [9.5.6]$$

For convenience we might like to choose, using [9.2.4],

$$\sqrt{\frac{2(1-c)t}{6+c}} < X_c \leq X < Y_a$$

as a lower bound for Y_a, however, with this choice,

$$f'(Y_a) < (1+c)t - \frac{(6-c)(1-c)t}{(6+c)}$$

and it is not obvious whether the right hand side is positive or negative. We must therefore choose a tighter bound on X (and consequently Y_a).

Suppose $a\sqrt{t} < X < Y_a$. Then

$$f'(Y_a) < \left\{(1+c) - \frac{1}{2}(6-c)a^2\right\}t$$

and hence if we wish to bound $f'(Y_a)$ away from zero a must satisfy the inequality,

$$a^2 > \frac{2(1+c)}{6-c}. \quad [9.5.7]$$

Clearly, as a increases the lower bound on X approaches the cocaustic and the region that is close to this boundary becomes smaller. We therefore chose a so that it only just satisfies [9.5.7]. This of course will correspond to a weaker estimate for $f'(Y_a)$ and hence ultimately a weaker estimate for $\partial^2\chi[f]/\partial^2 X$, but this does not matter for we simply consider a smaller neighbourhood of the origin to ensure a contraction mapping. If we choose $a^2 = 2(1+2c)/(6-c)$ then clearly [9.5.7] is satisfied and

$$\begin{aligned} f'(Y_a) &< \{(1+c)-(1+2c)\}\, t \\ &= -ct. \end{aligned}$$

A lower bound for $f'(Y_a)$ is easy. Using [9.5.3] and [9.3.10],

$$\begin{aligned} f'(Y_a) &> q'(Y_a) - ct - \frac{1}{2}cY_a^2 \\ &= (1-c)t - \frac{1}{2}(6+c)Y_a^2 \\ &> (1-c)t - \frac{1}{2}(6+c)X_{cc}^2 \\ &> (1-c)t - \frac{1}{2}(6+c)\cdot\frac{16(1+c)t}{6-c} \\ &> -\frac{(1+c)(6+c)t}{6-c} - (6+c)\cdot\frac{8(1+c)t}{6-c} \\ &= -\frac{9(6+c)(1+c)t}{6-c}. \end{aligned}$$

Thus finally,

$$ct < |f'(Y_a)| < \frac{9(6+c)(1+c)t}{6-c} \quad \forall X > \sqrt{\frac{2(1+2c)t}{6-c}}. \qquad [9.5.8]$$

The last term to consider is $|k'_{cc}(Y_d)|$. From previous discussions on estimates for the cocaustic and its derivative,

$$\frac{6-c}{8(1+c)}Y_d < k'_{cc}(Y_d) \leq \frac{6+c}{2(1-c)}Y_d$$

provided $c < 1/24$. Since

$$\sqrt{\frac{2(1+2c)t}{6-c}} < X < Y_d < X_{cc} < \sqrt{\frac{16(1+c)t}{6-c}},$$

it follows that

$$\frac{6-c}{8(1+c)}\sqrt{\frac{2(1+2c)t}{6-c}} < |k'_{cc}(Y_d)| < \frac{6+c}{2(1-c)}\sqrt{\frac{16(1+c)t}{6-c}}. \qquad [9.5.9]$$

With all these results ([9.5.4], [9.5.5], [9.5.8] and [9.5.9]) we can now construct estimates for $[f'(X_1)]^{-1}$ using [9.5.1]; we obtain

$$-\frac{\sqrt{2}(6+c)^{3/2}(1+c)^{1/2}}{(6-c)^{3/2}(1-c)c^{1/2}t^{1/2}|t-k_{cc}|^{1/2}} < [f'(X_1)]^{-1}$$
$$< -\frac{(6-c)^{7/4}(1-c)^{1/4}(1+2c)^{1/4}}{24\sqrt{2}(6+c)^{7/4}(1+c)^{3/2}t^{1/2}|t-k_{cc}|^{1/2}}$$

$$\Longrightarrow -\frac{\sqrt{2}(6+c)^{3/2}(1+c)^{3/2}}{(6-c)^{3/2}(1-c)^{3/2}c^{1/2}t^{1/2}|t-k_{cc}|^{1/2}} < [f'(X_1)]^{-1}$$
$$< -\frac{(6-c)^{7/4}(1-c)^{7/4}}{24\sqrt{2}(6+c)^{7/4}(1+c)^{7/4}t^{1/2}|t-k_{cc}|^{1/2}}.$$

Using [9.3.5] we can then remove most of the terms involving c; we obtain

$$-\frac{\sqrt{2}\cdot 10^{3/2}}{9^{3/2}c^{1/2}t^{1/2}|t-k_{cc}|^{1/2}} < [f'(X_1)]^{-1} < -\frac{9^{7/4}}{24\sqrt{2}\cdot 10^{7/4}t^{1/2}|t-k_{cc}|^{1/2}}$$

$$\Longrightarrow -\frac{5}{3c^{1/2}t^{1/2}|t-k_{cc}|^{1/2}} < [f'(X_1)]^{-1} < -\frac{1}{48t^{1/2}|t-k_{cc}|^{1/2}} \qquad [9.5.10]$$

for all

$$\sqrt{\frac{2(1+2c)t}{6-c}} < X < X_{cc} \quad \text{and} \quad c < \frac{1}{24}.$$

§9.6. Estimates for $[f'(X_2(t,X))]^{-1}$ for X near the cocaustic.

We follow similar arguments to the previous section since if X is close to the cocaustic, then we expect X_2 to be close to the caustic in the opposite sense to X_1 (i.e. $X_1 \leq -X_c \leq X_2$). Now by Taylor's theorem,

$$f(X_{cc}) - |X - X_{cc}|f'(Y_a) = f(-X_c) + \frac{1}{2}|X_2 + X_c|^2 f''(Y_b)$$

and

$$f'(X_2) = f'(-X_c) + |X_2 + X_c| f''(Y_c).$$

Here Y_a, Y_b and Y_c such that $X < Y_a < X_{cc}$, $-X_c < Y_b < X_2$ and $-X_c < Y_c < X_2$ are different to those defined in the previous section. These equations imply that

$$[f'(X_2)]^{-1} = \frac{|f''(Y_b)|^{1/2}}{\sqrt{2}f''(Y_c)|f'(Y_a)|^{1/2}|X - X_{cc}|^{1/2}}.$$

Again, by the mean value theorem, $|X - X_{cc}| = |t - k_{cc}||k'_{cc}(Y_d)|^{-1}$ where $X < Y_d < X_{cc}$ so that finally,

$$[f'(X_2)]^{-1} = \frac{|f''(Y_b)|^{1/2}|k'_{cc}(Y_d)|^{1/2}}{\sqrt{2}|f''(Y_c)||f'(Y_a)|^{1/2}|t - k_{cc}|^{1/2}}, \qquad [9.6.1]$$

noting that $f''(Y_c)$ is positive.

Consider first of all $|f''(Y_b)|$. Equation [9.5.2] implies that

$$(6 - c)|X_2| < (6 - c)|Y_b| \leq |f''(Y_b)| \leq (6 + c)|Y_b| < (6 + c)X_c. \qquad [9.6.2]$$

Similarly,

$$(6 - c)|X_2| < |f''(Y_c)| < (6 + c)X_c. \qquad [9.6.3]$$

The previous section calculated a bound for $|f'(Y_a)|$,

$$ct < |f'(Y_a)| < \frac{9(6 + c)(1 + c)t}{(6 - c)} \quad \forall X > \sqrt{\frac{2(1 + 2c)t}{6 - c}}. \qquad [9.5.8]$$

This is valid even though technically the Y_a are no longer the same quantity. The reason for this is that in both cases, when estimates for $|f'(Y_a)|$ are calculated, Y_a terms are introduced via equation [9.5.6] and then removed by using $X_c \leq X < Y_a < X_{cc}$. This makes the Y_a in this context equivalent. For a similar reason we again have

$$\frac{6 - c}{8(1 + c)}\sqrt{\frac{2(1 + 2c)t}{6 - c}} < |k'_{cc}(Y_d)| < \frac{6 + c}{2(1 - c)}\sqrt{\frac{16(1 + c)t}{6 - c}}. \qquad [9.5.9]$$

Finally, inserting estimates [9.6.2], [9.6.3], [9.5.8] and [9.5.9] into equation [9.6.1] gives

$$\begin{aligned} \frac{(6 - c)^{7/4}(1 + 2c)^{1/4}|X_2|^{1/2}}{12 \cdot 2^{1/4}(6 + c)^{3/2}(1 + c)^{3/2}t^{3/4}|t - k_{cc}|^{1/2}} &< [f'(X_2)]^{-1} \\ &< \frac{2^{1/4}(6 + c)(1 + c)^{1/2}}{c^{1/2}(6 - c)^{3/2}(1 - c)^{1/2}|X_2||t - k_{cc}|^{1/2}} \end{aligned} \qquad [9.6.4]$$

At this point the estimation process diverges from that of the previous section. Although this was not explicitly done, the quantity, $|X_1|$, which is the unknown function of X, was removed using $-X_{cc} \leq X_1 \leq -X_c$. These are of course the least rigorous bounds on X_1 there can be. In this case, however, since $-X_c \leq X_2 \leq X_c$, we must conclude that $0 \leq |X_2| \leq X_c$ meaning that $[f'(X_2)]^{-1}$ could potentially be infinite as illustrated by equation [9.6.4]. To solve this problem we need to find an upper bound, X_{UB} say, such that $-X_c < X_2 < X_{UB}$. The way that we shall do this is a two step process: firstly it will be shown that for fixed $t > 0$, f is strictly increasing on $[0, X_c)$, and strictly decreasing on $(X_c, X_{cc}]$. Then, using these results and the asymmetry of the problem, we shall find an X_{UB} such that for all $X > a\sqrt{t}$, $f(X_2) < f(X_{UB})$ and consequently $X_2 < X_{UB}$. Here a is an undetermined constant different to that of the previous section.

First of all let us argue that f is strictly increasing on $[0, X_c)$ and strictly decreasing on $(X_c, X_{cc}]$. By definition, X_c and $-X_c$ are the only solutions to $f'(X) = 0$. Thus for $X \in (0, X_c)$, either $f'(X) > 0$ or $f'(X) < 0$. Now $|f''(X) - q''(X)| \leq c|X|$ implies that for all $X > 0$, $f''(X) \leq q''(X) + c|X| = -6X + cX < 0$ since $c < 1/24$. Thus $f'(X)$ is strictly decreasing for all $X > 0$ and hence must be positive on $(0, X_c)$. By the continuity of f' this region can be extended to $[0, X_c)$. Thus f is strictly increasing on $(-X_c, X_c)$. Furthermore, since $f'(X)$ is strictly decreasing for all $X > 0$, it follows that $f'(X)$ is negative in $(X_c, X_{cc}]$ and that $f(X)$ is strictly decreasing in this region.

We can now begin to find the constant, a, that defines X_{UB}. Unfortunately there is no chronological argument that begins with an assumption and ends by defining X_{UB}. Instead we *suppose* that X_{UB} is given and then proceed to show that it is in fact an upper bound for X_2. We begin by simplifying the estimate for f which makes the following calculation easier. Previous analysis shows that integrating the second order derivatives of f with the appropriate boundary condition gives $|f(X) - q(X)| \leq ct|X| + c|X|^3/6$. Since we are working with a fixed t domain, we can define a new constant, k, and write for any point in the 3-dust region,

$$\begin{aligned} |f(X) - q(X)| &\leq ct|X| + \frac{1}{6}c|X|^3 \\ &\leq \left(ct + \frac{1}{6}cX_{cc}^2\right)|X| \\ &= kt|X|. \end{aligned}$$

We shall find it useful to relate k to c. Using [9.3.10],

$$k = c + \frac{cX_{cc}^2}{6t} < c\left(1 + \frac{8(1+c)}{3(6-c)}\right)$$

and since $c < 1/24$,

$$\frac{8(1+c)}{3(6-c)} < \frac{8 \cdot 25}{3 \cdot 143} < \frac{9 \cdot 25}{3 \cdot 125} = \frac{3}{5}$$

so that, without loss of generality, we can redefine $k = 2c$.

Next, since q is not invertible on $\left[-\sqrt{4t/3}, \sqrt{4t/3}\right]$, we define q_i $(i = 1, 2, 3)$ that do have inverses by $q_1 = q \ \forall X \in \left[-\sqrt{4t/3}, -\sqrt{t/3}\right]$, $q_2 = q \ \forall X \in \left(-\sqrt{t/3}, \sqrt{t/3}\right]$ and $q_3 = q \ \forall X \in \left(\sqrt{t/3}, \sqrt{4t/3}\right]$. Here the quantities $\sqrt{t/3}$ and $\sqrt{4t/3}$ represent the caustic and cocaustic corresponding to $x = q(t, X)$. Now, supposing that X_{UB} is defined by $X_{UB} = q_2^{-1}\left(q(a\sqrt{t}) + 4ct \cdot a\sqrt{t}\right)$, then

$$\begin{aligned}
f(X_2) &= f(X) \\
&\le f(a\sqrt{t}) \\
&\le q(a\sqrt{t}) + 2ct|a\sqrt{t}| \\
&= q(a\sqrt{t}) + 2ct \cdot a\sqrt{t} \\
&< q(a\sqrt{t}) + 4ct \cdot a\sqrt{t} - 2ct|X_{UB}| \\
&= q\left(q_2^{-1}\left(q(a\sqrt{t}) + 4ct \cdot a\sqrt{t}\right)\right) - 2ct|X_{UB}| \\
&= q\left(X_{UB}\right) - 2ct|X_{UB}| \\
&\le f\left(X_{UB}\right) \\
\Longrightarrow X_2 &< X_{UB}.
\end{aligned}$$

There are a number of conditions that X_{UB} and the constant a must satisfy for this argument to work. First of all we have

$$-\frac{2}{3}t\sqrt{\frac{t}{3}} < q(a\sqrt{t}) + 4ct \cdot a\sqrt{t} < \frac{2}{3}t\sqrt{\frac{t}{3}}. \qquad [9.6.5]$$

This ensures that $q(a\sqrt{t}) + 4ct \cdot a\sqrt{t}$ lies in the domain of q_2^{-1}. Clearly, without this condition X_{UB} cannot be defined. The quantities $2t/3\sqrt{t/3}$ and $-2t/3\sqrt{t/3}$ are simply the maximum and minimum values of q_2, namely $q\left(\sqrt{t/3}\right)$ and $q\left(-\sqrt{t/3}\right)$ respectively. The second condition is that $|X_{UB}| < a\sqrt{t}$. This arises from the above mathematics since without this statement we could not conclude that $f(X_2) < f(X_{UB})$. Thirdly we require that X_{UB} is negative. This last condition is crucial for estimating $[f'(X_2)]^{-1}$ as it ensures that this quantity remains bounded.

Now,

$$\begin{aligned} X_{UB} &= q_2^{-1}\left(q(a\sqrt{t}) + 4ct \cdot a\sqrt{t}\right) \\ &= q_2^{-1}\left(a\sqrt{t} \cdot t - \left(a\sqrt{t}\right)^3 + 4ct \cdot a\sqrt{t}\right). \end{aligned}$$

If one temporarily reinserts the time dependence of q then

$$\begin{aligned} X_{UB} &= q_2^{-1}\left(q\left((1+4c)t, a\sqrt{t}\right)\right) \\ &= \left\{\frac{-a + \sqrt{4(1+4c) - 3a^2}}{2}\right\}\sqrt{t} \end{aligned}$$

and hence in addition to the above three conditions we must have

$$a^2 < \frac{4}{3}(1+4c), \qquad [9.6.6]$$

which again ensures the existence of X_{UB}. This requirement is different to the one above which is also ensuring that X_{UB} can be defined because in this case the restriction is more on X than on X_2 or X_{UB}. The inequality, $4(1+4c) - 3a^2 > 0$, implies that $4(1+4c)t - 3X^2 > 0$ and can be interpreted as ensuring that the point, (t, X), lies within the 3-dust region defined by q at a later time of $(1+4c)t$. This means that the above procedure for showing that $f(X_2) < f(X_{UB})$ is, in essence, a procedure which finds a constant, c, and a new cubic, $q_\beta(X,t) = q(X,(1+4c)t) = (1+4c)Xt - X^3$, which is an upper bound for f for all $X > 0$. Since q has been made invertible by segmenting its domain, q_β can be made invertible in a similar manner enabling X_{UB}, the solution to $q_\beta(X_{UB}, t) = q_\beta(a\sqrt{t}, t)$, to be determined.

Let us begin with the constraints on X_{UB} that are the least complicated. For X_{UB} to be negative,

$$a > \sqrt{4(1+4c) - 3a^2}$$

$$\implies a^2 > (1+4c). \qquad [9.6.7]$$

In fact we find that this requirement means that the inequality, $|X_{UB}| < a\sqrt{t}$, is automatically satisfied since

$$|X_{UB}| = \left\{ \frac{a - \sqrt{4(1+4c) - 3a^2}}{2} \right\} \sqrt{t}$$
$$< \frac{1}{2} a\sqrt{t}.$$

Let us now consider the final requirement of equation [9.6.5] which can be written as

$$-\frac{2}{3}\sqrt{\frac{1}{3}} < a(1 - a^2 + 4c) < \frac{2}{3}\sqrt{\frac{1}{3}}. \qquad [9.6.8]$$

Now [9.6.6] and [9.6.7] imply that (remembering that $a > 0$)

$$-\frac{a}{3}(1+4c) < a(1 - a^2 + 4c) < 0$$

so [9.6.8] is satisfied if

$$a(1+4c) < 2\sqrt{\frac{1}{3}}.$$

In order to maximise the area of the region adjacent to the cocaustic, a must be chosen to be as small as possible. Equations [9.6.6] and [9.6.7] imply that $(1+4c) < a^2 < 4(1+4c)/3$. For this reason we therefore chose $a^2 = 1 + 6c$. This means that c is required to satisfy

$$(1+6c)(1+4c)^2 < \frac{4}{3}$$

$$\Longrightarrow 96c^3 + 64c^2 + 14c - \frac{1}{3} < 0$$

and this inequality holds if $c < 1/48$.

X_{UB} can now be determined as a function of t. We have

$$X_{UB} = \left\{ \frac{-\sqrt{1+6c} + \sqrt{4(1+4c) - 3(1+6c)}}{2} \right\} \sqrt{t}$$
$$= \left\{ \frac{-\sqrt{1+6c} + \sqrt{1-2c}}{2} \right\} \sqrt{t}.$$

Now,

$$\left(-\sqrt{1+6c}+\sqrt{1-2c}\right)^2 = 1+6c+1-2c-2\sqrt{1+6c}\sqrt{1-2c}$$
$$> 1+6c+1-2c-2(1+6c)$$
$$= 16c,$$

hence

$$-\sqrt{1+6c}+\sqrt{1-2c} < -4\sqrt{c} < -4c,$$

noting that $c < 1$. It follows that $X_{UB} < -2c\sqrt{t}$.

In conclusion we now have

$$-\sqrt{\frac{2(1-c)}{6+c}t} < -X_c < X_2 < X_{UB} < -2c\sqrt{t}$$

or

$$2c\sqrt{t} < |X_2| < \sqrt{\frac{2(1-c)}{6+c}t}.$$

As can be seen, the quantity $-2c\sqrt{t}$ is used rather than the messy but exact definition of X_{UB}. This is fine as the object of this exercise was purely to find an upper bound for X_2 which is negative and bounded away from zero; criterion which $-2c\sqrt{t}$ satisfies. X_{UB} was simply used as a tool to prove existence of an upper bound for X_2 for all $\sqrt{(1+6c)t} < X \leq X_c$.

Having determined bounds for X_2, we can return to the main discussion of this chapter and determine the bounds for $[f'(X_2)]^{-1}$. Using these results, [9.6.4] becomes

$$\frac{2^{1/4}c^{1/2}(6-c)^{7/4}(1+2c)^{1/4}t^{1/4}}{12(6+c)^{3/2}(1+c)^{3/2}t^{3/4}|t-k_{cc}|^{1/2}} < [f'(X_2)]^{-1}$$
$$< \frac{(6+c)(1+c)^{1/2}}{2^{3/4}c^{3/2}(6-c)^{3/2}(1-c)^{1/2}t^{1/2}|t-k_{cc}|^{1/2}}$$

$$\Longrightarrow \frac{2^{1/4}c^{1/2}(6-c)^{7/4}(1-c)^{7/4}}{12\cdot(6+c)^{7/4}(1+c)^{7/4}t^{1/2}|t-k_{cc}|^{1/2}} < [f'(X_2)]^{-1}$$
$$< \frac{(6+c)^{3/2}(1+c)^{3/2}}{2^{3/4}c^{3/2}(6-c)^{3/2}(1-c)^{3/2}t^{1/2}|t-k_{cc}|^{1/2}}$$

$$\Longrightarrow \frac{2^{1/4}\cdot 9^{7/4}c^{1/2}}{12\cdot 10^{7/4}t^{1/2}|t-k_{cc}|^{1/2}} < [f'(X_2)]^{-1} < \frac{10^{3/2}}{2^{3/4}\cdot 9^{3/2}c^{3/2}t^{1/2}|t-k_{cc}|^{1/2}}$$

$$\Longrightarrow \frac{c^{1/2}}{12\sqrt{2}t^{1/2}|t-k_{cc}|^{1/2}} < [f'(X_2)]^{-1} < \frac{1}{\sqrt{2}c^{3/2}t^{1/2}|t-k_{cc}|^{1/2}} \qquad [9.6.9]$$

for all

$$\sqrt{(1+6c)t} < X < X_{cc} \quad \text{and} \quad c < \frac{1}{48}.$$

§9.7. Summary of estimates for $[f(X_1(t,X))]^{-1}$ and $[f(X_2(t,X))]^{-1}$ for X near the cocaustic.

To clarify the restrictions placed on X and c, the estimates obtained in the previous two sections are summarised here. We have

$$-\frac{5}{3c^{1/2}t^{1/2}|t-k_{cc}|^{1/2}} < [f'(X_1)]^{-1} < -\frac{1}{48t^{1/2}|t-k_{cc}|^{1/2}} \qquad [9.5.10]$$

provided

$$\sqrt{\frac{2(1+2c)t}{6-c}} < X < X_{cc} \quad \text{and} \quad c < \frac{1}{24},$$

and

$$\frac{c^{1/2}}{12\sqrt{2}t^{1/2}|t-k_{cc}|^{1/2}} < [f'(X_2)]^{-1}] < \frac{1}{\sqrt{2}c^{3/2}t^{1/2}|t-k_{cc}|^{1/2}} \qquad [9.6.9]$$

provided

$$\sqrt{(1+6c)t} < X < X_{cc} \quad \text{and} \quad c < \frac{1}{48}.$$

Clearly, for $c < 1/48$,

$$\frac{2(1+2c)}{6-c} < \frac{100}{287}$$

and hence

$$1+6c > 1 > \frac{100}{287} > \frac{2(1+2c)}{6-c}.$$

So in order for the above estimates for $[f(X_1)]^{-1}$ and $[f(X_2)]^{-1}$ to be valid simultaneously in the region close to the cocaustic, we must define this region to be $\left\{(X,t) \mid \sqrt{(1+6c)t} < X \leq X_{cc}\right\}$.

§9.8. Estimates for $[f'(X_1(t,X))]^{-1}$ for $X > X_c$ near the caustic.

Because X is near to the caustic, X_1 must be close to the cocaustic by definition. Thus, provided X is sufficiently close to the caustic, we expect $f'(X_1)$ to be well behaved and bounds can be obtained without expanding in powers of $|t - k_c|$.

We have from [9.5.3],

$$(1-c)t - \frac{1}{2}(6+c)X_1^2 \leq f'(X_1) \leq (1+c)t - \frac{1}{2}(6-c)X_1^2 \quad \forall X_1. \qquad [9.8.1]$$

The least restrictive bounds on X_1 are $-X_{cc} \leq X_1 \leq -X_c$ and if these are used in conjunction with [9.8.1] then we obtain

$$\left\{1 - c - \frac{8(1+c)}{(6-c)}(6+c)\right\} t \leq f'(X_1) \leq \left\{1 - c - \frac{(1-c)}{(6+c)}(6-c)\right\} t,$$

using the bounds for X_c and X_{cc} as given by equations [9.2.4] and [9.3.10]. As can be seen this does not provide useable estimates because $f'(X_1)$ is required to be non-zero and the above does not enforce this. To proceed, therefore, we need to determine an upper bound, X_{UB} say (which is different to that of the previous section), which can ensure that $f'(X_1)$ is negative and non-zero so that $[f'(X_1)]^{-1}$ remains bounded.

The procedure for determining X_{UB} follows a similar argument to that of the previous section where an upper bound for X_2 was found. Using the fact that f is decreasing $\forall X < -X_c$ and $X > X_c$, it will be shown that $f(X_1) > f(X_{UB})$ for all $X < b\sqrt{t}$ and hence $X_1 < X_{UB}$.

To prove this inequality, let us suppose that $X_{UB} = q_1^{-1}\left(q(b\sqrt{t}) - 4ct \cdot b\sqrt{t}\right)$. Then, using $|f(X) - q(X)| \leq 2c|X|$,

$$\begin{aligned}
f(X_1) &= f(X) \\
&> f(b\sqrt{t}) \\
&> q(b\sqrt{t}) - 2ct|b\sqrt{t}| \\
&> q(b\sqrt{t}) - 4ct|b\sqrt{t}| + 2ct|X_{UB}| \\
&= q\left(q_1^{-1}\left(q(b\sqrt{t}) - 4ct \cdot b\sqrt{t}\right)\right) + 2ct|X_{UB}| \\
&> f(X_{UB}) \\
\Longrightarrow X_1 &< X_{UB}.
\end{aligned}$$

Once again this series of mathematical statements only works provided a number of conditions are met. Firstly, in order for X_{UB} to be defined, $q\left(b\sqrt{t}\right) - 4ctb\sqrt{t}$ must lie in the domain of q_1, i.e.,

$$-\frac{2t}{3}\sqrt{\frac{t}{3}} < q\left(b\sqrt{t}\right) - 4ctb\sqrt{t} < \frac{2t}{3}\sqrt{\frac{t}{3}}. \qquad [9.8.2]$$

Since q_1 is strictly decreasing the quantities $2t/3\sqrt{t/3}$ and $-2t/3\sqrt{t/3}$ can easily be identified as $q\left(-\sqrt{4t/3}\right)$, the value of q at the cocaustic and $q\left(-\sqrt{t/3}\right)$, the value of q at the caustic respectively. The other condition that enables X_{UB} to be defined is essentially a restriction on the possible values of X. To see this we have from its definition,

$$\begin{aligned} X_{UB} &= q_1^{-1}\left(q\left(b\sqrt{t}\right) - 4ct \cdot b\sqrt{t}\right) \\ &= q_1^{-1}\left(b\sqrt{t}\cdot t - (b\sqrt{t})^3 - 4ct \cdot b\sqrt{t}\right). \end{aligned}$$

If we temporarily reinsert the time dependence of q then

$$\begin{aligned} X_{UB} &= q_1^{-1}\left(q\left((1-4c)t, b\sqrt{t}\right)\right) \\ &= \left\{\frac{-b - \sqrt{4(1-4c) - 3b^2}}{2}\right\} \end{aligned} \qquad [9.8.3]$$

and so for the square root to be real,

$$b^2 < \frac{4}{3}(1-4c). \qquad [9.8.4]$$

This inequality has a similar interpretation to that of [9.6.6] for it ensures that the point (t, X) lies within the 3-dust region defined by q at an earlier time of $(1-4c)t$. This means that the above argument showing that $f(X_1) > f(X_{UB})$ is a procedure defining a new cubic, $q_\alpha(X,t) = q(X,(1-4c)t) = (1-4c)Xt - X^3$, which is an upper bound for f for all $X < 0$ and that X_{UB} is simply the solution to $q_\alpha(X_{UB},t) = q_\alpha(b\sqrt{t},t)$.

The last two conditions are $|X_{UB}| < b\sqrt{t}$, which allows the proof of $f(X_1) > f(X_{UB})$ to go through, and

$$X_{UB}^2 > \frac{2(1+c)t}{(6-c)}, \qquad [9.8.5]$$

which ensures that $f'(X_1)$ is negative and bounded away from zero (the point to this exercise) and is derivable from [9.8.1]. Let us consider the condition that $|X_{UB}| < b\sqrt{t}$. From [9.8.3], X_{UB} is clearly negative and so this condition becomes

$$\frac{b+\sqrt{4(1-4c)-3b^2}}{2} < b$$

$$\Longrightarrow b^2 > (1-4c). \tag{9.8.6}$$

There now exist upper and lower bounds on b^2 as given by [9.8.4] and [9.8.6] and to proceed, b^2 should be defined. Now as b increases the range of admissible values for X_1 also increases as X_{UB} moves closer to $-X_c$. Clearly, it is our desire to have the region close to the caustic as large as possible. To this end b is defined by $b^2 = 4(1-6c)/3$ which only just satisfies the bounds imposed on b^2. Since c is at least less than $1/48$, this choice automatically satisfies [9.8.4] and [9.8.6] simultaneously. With X_{UB} defined explicitly by equation [9.8.3], [9.8.5] becomes

$$b^2 + 2b\sqrt{4(1-4c)-3b^2} + 4(1-4c) - 3b^2 > \frac{8(1+c)}{6-c}$$

$$\Longrightarrow \frac{4}{3}(1-6c) + 2\sqrt{\frac{4}{3}(1-6c)}\sqrt{8c} + 8c > \frac{8(1+c)}{6-c}$$

with the above choice for b^2. With the current value for c, $\sqrt{4(1-6c)/3}\sqrt{8c} > 8c$ and the above becomes

$$\frac{4}{3}(1+12c) > \frac{8(1+c)}{6-c}.$$

This inequality is equivalent to $12c^2 - 65c < 0$, which has solutions $0 < c < 65/12$. Since we already have $c < 1/48$, the choice that $X < \sqrt{4(1-6c)/3}$ means that X_{UB} satisfies [9.8.5].

The last constraint to consider is of course [9.8.2]. Now

$$\begin{aligned} q\left(b\sqrt{t}\right) - 4ct\cdot b\sqrt{t} &= bt\sqrt{t} - (b\sqrt{t})^3 - 4ct\cdot b\sqrt{t} \\ &= ((1-4c)b - b^3)t\sqrt{t} \\ &= \sqrt{\frac{4}{3}(1-6c)}\left\{-\frac{1}{3}(1-12c)\right\}t\sqrt{t}, \end{aligned}$$

which is negative since $1 - 12c > 0$ for $c < 1/48$. Thus if [9.8.2] is to be satisfied, we simply need to show that

$$-\frac{2}{3}\sqrt{\frac{1}{3}} < -\frac{1}{3}(1 - 12c)\sqrt{\frac{4}{3}(1 - 6c)}.$$

This, however, is trivial for the above becomes

$$1 > (1 - 12c)\sqrt{1 - 6c},$$

which is true for all $c > 0$.

With b given, X_{UB} can be defined explicitly, however, this produces a complicated bound for X_1. For this reason, X_{UB} is itself estimated by a quantity, $\overline{X}_{UB}$ say, such that $X_{UB} < \overline{X}_{UB} < 0$ satisfies

$$X_{UB}^2 > \overline{X}_{UB}^2 > \frac{2(1+c)}{6-c}t. \qquad [9.8.7]$$

This statement is important for it allows X_1 to be estimated by $\overline{X}_{UB}$ whilst insuring that $f'(X_1)$ is negative and bounded away from zero. Now since $c < 1/48$,

$$X_{UB} = \left\{\frac{-\sqrt{4(1-6c)/3} - \sqrt{8c}}{2}\right\}\sqrt{t}$$

implies that

$$\begin{aligned}
X_{UB}^2 &= \left\{\frac{4(1-6c)/3 + 8c + 2\sqrt{4(1-6c)/3}\sqrt{8c}}{4}\right\}t \\
&> \left\{\frac{1}{3} + \sqrt{\frac{1}{3}(48c - 6c)8c}\right\}t \\
&= \left\{\frac{1}{3} + c\sqrt{112}\right\}t \\
&= \frac{1}{3}(1 + c\sqrt{1008})t \\
&> \frac{1}{3}(1 + c\sqrt{961})t \\
&= \frac{1}{3}(1 + 31c)t \\
\Longrightarrow X_{UB}^2 &< -\sqrt{\frac{1}{3}(1 + 31c)t}.
\end{aligned}$$

If $\overline{X}_{UB} = -\sqrt{(1+31c)t/3}$ then [9.8.7] implies that

$$\frac{1}{3}(1+31c) > \frac{2(1+c)}{6-c}$$

$$\Longrightarrow 31c^2 - 179c < 0$$

and hence [9.8.7] holds provided $c < 179/31$.

In conclusion, therefore, using [9.3.10],

$$\sqrt{\frac{4(1-c)t}{6+c}} < X_1 < X_{UB} < -\sqrt{\frac{1}{3}(1+31c)t}$$

for all $X_c \leq X < \sqrt{4(1-6c)t/3}$ so that [9.8.1] implies that

$$(1-c)t - \frac{1}{2}(6+c)\frac{4(1-c)t}{6+c} < f'(X_1) < (1+c)t - \frac{1}{2}(6-c)\frac{1}{3}(1+31c)t$$

$$\Longrightarrow -(1-ct)t < f'(X_1) < \left(-30c + \frac{1}{6}c + \frac{31}{6}c^2\right)t$$

$$\Longrightarrow -(1-c)t < f'(X_1) < -25ct$$

$$\Longrightarrow -\frac{1}{25ct} < [f'(X_1)]^{-1} < -\frac{1}{(1-c)t}. \qquad [9.8.8]$$

§9.9. Estimates for $[f'(X_2(t,X))]^{-1}$ for $X > X_c$ near the caustic.

Since we are close to the point where X_3 is relabelled as X_2, $[f'(X_2)]^{-1}$ is expected to become unbounded as X approaches X_c, the point of transition. For this region we must expand $[f'(X_2)]^{-1}$ in powers of $|t-k_c|$. By Taylor's theorem,

$$\frac{1}{2}|X-X_c|^2 f''(Y_a) = -\frac{1}{2}|X_2-X_c|^2 f''(Y_b),$$

$$f'(X_2) = -|X_2-X_c| f''(Y_c) = |X_2-X_c||f''(Y_c)|$$

and

$$|t-k_c| = |X-X_c||k_c'(Y_d)|$$

where $X_c < Y_a < X$, $X_2 < Y_b < X_c$, $X_2 < Y_c < X_c$ and $X_c < Y_d < X$. These equations together imply that

$$[f'(X_2)]^{-1} = \frac{|f''(Y_b)|^{1/2}|k_c'(Y_d)|}{|f''(Y_c)||f''(Y_a)|^{1/2}|t-k_c|}. \qquad [9.9.1]$$

Let us consider first of all the terms that involve second derivatives of f and estimate these. Using $|f''(X) - q''(X)| \leq c|X|$, it follows that

$$(6-c)X_c < |f''(Y_a)| < (6+c)X,$$

$$(6-c)|X_2| < |f''(Y_b)| < (6+c)X_c$$

and

$$(6-c)|X_2| < |f''(Y_c)| < (6+c)X_c.$$

Next, from the estimates given by [9.2.2] and [9.2.4],

$$\frac{6-c}{1+c}X_c < \frac{6-c}{1+c}Y_d < |k_c'(Y_d)| < \frac{6+c}{1-c}Y_d < \frac{6+c}{1-c}X.$$

Hence equation [9.9.1] implies the following upper and lower bounds,

$$\frac{(6-c)^{3/2}|X_2|^{1/2}}{(6+c)^{3/2}(1+c)X^{1/2}|t-k_c|} < [f'(X_2)]^{-1} < \frac{(6+c)^{3/2}X}{(6-c)^{3/2}(1-c)|X_2||t-k_c|}. \qquad [9.9.2]$$

As can be seen, in order to proceed with estimating $[f'(X_2)]^{-1}$ the quantities $|X_2|$ and X need to be bounded by functions of t. The latter is the easiest for its bounds are determined coarsely by $X_c \leq X \leq X_{cc}$ and more accurately by $X_c \leq X \leq b\sqrt{t}$ where $b\sqrt{t}$ defines the region close to the caustic. Estimating $|X_2|$, however, involves a complicated process similar to that of the previous section.

Clearly, from [9.9.2], $|X_2|$ must be bounded away from zero. Since X is close to the caustic, X_2 must be close to the caustic in the opposite sense (i.e. $0 < X_2 < X_c < X$) and hence for X very close to X_c, X_2 will be positive. The way forward therefore, is to decrease the value of b so that X_2 is always positive and bounded below by $X_{LB} > 0$. The proof of the existence of X_{LB} is stated concisely in the following theorem.

Theorem. *Suppose that* $X_{LB} = q_2^{-1}\left(q\left(\sqrt{(1-6c)t}\right) - 4ct\sqrt{(1-6c)t}\right)$, $X_c \leq X \leq \sqrt{(1-6c)t}$ *and* $|f(X) - q(X)| < 2ct|X|$, *then* $X_{LB} < X_2 \leq X_c$.

Proof. Let $X < b\sqrt{t}$ and define $X_{LB} = q_2^{-1}\left(q\left(b\sqrt{t}\right) - 4ct \cdot b\sqrt{t}\right)$. Since f is increasing on $(-X_c, X_c)$ and decreasing on (X_c, ∞) we have

$$\begin{aligned}
f(X_2) &= f(X) \\
&\geq f(b\sqrt{t}) \\
&\geq q(b\sqrt{t}) - 2ct \cdot b\sqrt{t} \\
&> q(b\sqrt{t}) - 4ct \cdot b\sqrt{t} + 2ct|X_{LB}| \\
&= q\left(q_2^{-1}\left(q(b\sqrt{t}) - 4ct \cdot b\sqrt{t}\right)\right) + 2ct|X_{LB}| \\
&= q(X_{LB}) + 2ct|X_{LB}| \\
&> f(X_{LB}) \\
\Longrightarrow X_2 &> X_{LB}.
\end{aligned}$$

There are, as expected, a number of conditions that X_{LB} and consequently b must satisfy. Firstly, to insure that X_{UB} exists, $q\left(b\sqrt{t}\right) - 4ct \cdot b\sqrt{t}$ must lie in the domain of q_2. In other words,

$$-\frac{2}{3}t\sqrt{\frac{t}{3}} < q\left(b\sqrt{t}\right) - 4ct \cdot b\sqrt{t} < \frac{2}{3}t\sqrt{\frac{t}{3}} \qquad [9.9.3]$$

where the upper and lower bounds are simply the maximum and minimum values for $q_2(X)$. Secondly, since

$$\begin{aligned}
X_{LB} &= q_2^{-1}\left(q(b\sqrt{t}) - 4ct \cdot b\sqrt{t}\right) \\
&= \left\{\frac{-b + \sqrt{4(1-4c) - 3b^2}}{2}\right\}\sqrt{t},
\end{aligned} \qquad [9.9.4]$$

in order for X_{LB} to be real,

$$b^2 < \frac{4}{3}(1 - 4c). \qquad [9.9.5]$$

Additional constraints on X_{LB} other than those needed for its existence are $X_{LB} > 0$ and $|X_{LB}| < b\sqrt{t}$. The first of these two conditions ensures that $X_{LB} > 0$ which means that $[f'(X_2)]^{-1}$ for $X \geq X_c$ close to the caustic can be estimated. The last constraint allows the conclusion that $f(X_2) > f(X_{LB})$ to be made.

Now $X_{LB} > 0$ implies, from [9.9.4], that

$$\sqrt{4(1-4c) - 3b^2} > b$$

$$\Longrightarrow b^2 < (1 - 4c), \qquad [9.9.6]$$

which of course supersedes [9.9.5] in determining b.

The constraint, $|X_{LB}| < b\sqrt{t}$, determines a lower bound for b; since b is chosen to ensure that $X_{LB} > 0$,

$$|X_{LB}| < b\sqrt{t}$$

$$\Longrightarrow \frac{-b + \sqrt{4(1-4c) - 3b^2}}{2} < b$$

$$\Longrightarrow b^2 > \frac{1}{3}(1 - 4c). \qquad [9.9.7]$$

The final condition to consider is that of [9.9.3]. In actual fact this condition determines a range of admissible values for c. This range, which at the moment stands at $0 < c < 1/48$, is of course dependant on the choice of b. However, it is more important to allow the boundary for the current region in tX space to determine b rather than the upper value for c as otherwise it becomes impossible to estimate $[f'(X_i)]^{-1}$ over all of the tX plane. To this end we chose $b^2 = 1 - 6c$. This choice, which only just satisfies [9.9.6] and [9.9.7] for $c < 1/48$, clearly allows the region close to the caustic such that $X > X_c$ to be as large as possible.

With this choice of b,

$$\begin{aligned} q(b\sqrt{t}) - 4ct \cdot b\sqrt{t} &= b(1 - b^2 - 4c)t\sqrt{t} \\ &= 2cbt\sqrt{t} \\ &> 0 \end{aligned}$$

and hence [9.9.3] can be simplified to

$$2cb < \frac{2}{3}\sqrt{\frac{1}{3}}.$$

This in turn implies that

$$c^2(1 - 6c) < \frac{1}{27},$$

which is satisfied $\forall c > 0$.

Having proved that a lower bound for X_2 exists, it is useful to determine this bound as a function of t. Now from [9.9.4],

$$X_{LB} = \left\{\frac{-\sqrt{1-6c}+\sqrt{4(1-4c)-3(1-6c)}}{2}\right\}\sqrt{t}$$
$$= \left\{\frac{-\sqrt{1-6c}+\sqrt{1+2c}}{2}\right\}\sqrt{t}$$

$$\Longrightarrow X_{LB}^2 = \left\{\frac{1-6c+1+2c-2\sqrt{1-6c}\sqrt{1+2c}}{4}\right\}\sqrt{t}$$
$$> \left\{\frac{1-6c+1+2c-2(1-6c)}{4}\right\}\sqrt{t}$$
$$= 2c\sqrt{t}$$
$$\Longrightarrow X_{LB} > \sqrt{2ct}$$
$$> c\sqrt{2t}.$$

Hence in conclusion,

$$c\sqrt{2t} < X_{LB} < X_2 < X_c \quad \forall X_c \le X \le \sqrt{(1-6c)t} \quad \text{and} \quad c < \frac{1}{48}.$$

This means that from [9.9.2],

$$\frac{(6-c)^{3/2}\sqrt{c\sqrt{2t}}}{(6+c)^{3/2}(1+c)\sqrt{\sqrt{(1-6c)t}}|t-k_c|} < [f'(X_2)]^{-1} < \frac{(6+c)^{3/2}\sqrt{(1-6c)t}}{(6-c)^{3/2}(1-c)c\sqrt{2t}|t-k_c|}$$

$$\Longrightarrow \frac{2^{1/4}(6-c)^{3/2}(1-c)^{3/2}c^{1/4}}{(6+c)^{3/2}(1+c)^{3/2}|t-k_c|} < [f'(X_2)]^{-1} < \frac{(6+c)^{3/2}(1+c)^{3/2}}{\sqrt{2}(6-c)^{3/2}(1-c)^{3/2}c|t-k_c|}.$$

Using [9.3.5], this implies that

$$\frac{9^{3/2}\cdot 2^{1/4}c^{1/4}}{10^{3/2}|t-k_c|} < [f'(X_2)]^{-1} < \frac{10^{3/2}}{9^{3/2}\sqrt{2}c|t-k_c|}$$

$$\Longrightarrow \frac{c^{1/4}}{|t-k_c|} < [f'(X_2)]^{-1} < \frac{5}{6c|t-k_c|} \qquad [9.9.8]$$

for all

$$X_c \le X \le \sqrt{(1-6c)t} \quad \text{and} \quad c < \frac{1}{48}.$$

§9.10. Summary of estimates for $[f'(X_1(t,X))]^{-1}$ and $[f'(X_2(t,X))]^{-1}$ for $X > X_c$ near the caustic.

Since the estimates for these quantities require different constraints on the values for X, they will be listed here for clarity. We have

$$-\frac{1}{25ct} < [f'(X_1)]^{-1} < -\frac{1}{(1-c)t} \qquad [9.8.8]$$

for all

$$X_c \leq X < \sqrt{\frac{4}{3}(1-6c)t} \quad \text{and} \quad c < \frac{1}{48},$$

and

$$\frac{c^{1/4}}{|t-k_c|} < [f'(X_2)]^{-1} < \frac{5}{6c|t-k_c|} \qquad [9.9.8]$$

for all

$$X_c \leq X \leq \sqrt{(1-6c)t} \quad \text{and} \quad c < \frac{1}{48}.$$

Hence, so that the above estimates for $[f'(X_1)]^{-1}$ and $[f'(X_2)]^{-1}$ are valid simultaneously on the region close to the caustic with $X \geq X_c$, we must define this region by choosing the minimum upper bound on X, i.e. $\left\{(X,t) \mid X_c \geq X > \sqrt{(1-6c)t}\right\}$.

§9.11. Estimates for $[f'(X_1(t,X))]^{-1}$ and $[f'(X_2(t,X))]^{-1}$ for the region where X is bounded away from the caustic and cocaustic ($X_c < X < X_{cc}$).

For X finitely far from the caustic and cocaustic such that $X_c < X < X_{cc}$, both $[f'(X_1)]^{-1}$ and $[f'(X_2)]^{-1}$ are expected to be well behaved. The estimates for these quantities can therefore be obtained directly from $|f'(X)-q'(X)| \leq ct+cX^2/2$ rather than by expanding in powers of either $|t-k_c|$ or $|t-k_{cc}|$.

The primary goal is to estimate $f'(X_i)$ and to ensure that this quantity is bounded away from zero. Now for all X,

$$t(1-c) - \frac{1}{2}(6+c)X^2 \leq f'(X) \leq t(1+c) - \frac{1}{2}(6-c)X^2 \qquad [9.11.1]$$

and since $f'(X_1)$ is negative, this requirement becomes

$$X_1^2 > \frac{2(1+c)}{6-c}t. \qquad [9.11.2]$$

Similarly, the fact that $f'(X_2)$ should be positive and bounded away from zero means that

$$X_2^2 < \frac{2(1-c)}{6+c}t. \qquad [9.11.3]$$

As will be seen, [9.11.2] and [9.11.3] determine the boundaries to the region of which we are considering. Since f is decreasing for both $(-\infty, -X_c)$ and (X_c, ∞) and increasing for $(-X_c, X_c)$, it follows that the processes which make X_1 as large as possible and X_2 as small as possible, whilst still satisfying [9.11.2] and [9.11.3], must jointly determine the upper boundary for this region. Likewise making X_2 as large as possible will determine the lower boundary.

Let us first of all estimate $[f'(X_1)]^{-1}$. Clearly an upper bound, X_{UB} say, for X_1 is needed which satisfies [9.11.2] with X_1 replaced by X_{UB}. The existence of X_{UB} is proved in the following theorem.

Theorem. *Suppose* $X_{UB} = q_1^{-1}\left(q\left(\sqrt{4(1-6c)t/3}\right) - 4ct\sqrt{4(1-6c)t/3}\right)$, $\overline{X}_{UB} = -\sqrt{(1+30c)t/3}$, $X_c \le X < \sqrt{4(1-6c)t/3}$ *and* $|f(X) - q(X)| \le 2ct|X|$, *then* $0 > \overline{X}_{UB} > X_{UB} > X_1$ *such that* X_{UB}^2 *and* $\overline{X}_{UB}^2$ *are both greater than* $2(1+c)t/(6-c)$.

Proof. Suppose that $X < b\sqrt{t}$ and define $X_{UB} = q^{-1}\left(q\left(b\sqrt{t}\right) - 4ct \cdot b\sqrt{t}\right)$. Since f is decreasing on both $(-\infty, -X_c)$ and (X_c, ∞),

$$\begin{aligned}
f(X_1) &= f(X) \\
&\ge q\left(b\sqrt{t}\right) - 2ct \cdot b\sqrt{t} \\
&> q\left(b\sqrt{t}\right) - 4ct \cdot b\sqrt{t} + 2ct\,|X_{UB}| \\
&= q\left(q_1^{-1}\left(q\left(b\sqrt{t}\right) - 4ct \cdot b\sqrt{t}\right)\right) + 2ct\,|X_{UB}| \\
&= q\left(X_{UB}\right) + 2ct\,|X_{UB}| \\
&\ge f(X_{UB}) \\
\Longrightarrow X_1 &< X_{UB}
\end{aligned}$$

provided

$$-\frac{2}{3}t\sqrt{\frac{t}{3}} < q\left(b\sqrt{t}\right) - 4ct \cdot b\sqrt{t} < \frac{2}{3}t\sqrt{\frac{t}{3}} \qquad [9.11.4]$$

and

$$|X_{UB}| < b\sqrt{t}. \qquad [9.11.5]$$

In addition to these constraints we must have

$$X_{UB}^2 > \frac{2(1+c)}{6-c}t \qquad [9.11.6]$$

and for the existence of X_{UB}, which has the form

$$X_{UB} = \left\{\frac{-b-\sqrt{4(1-4c)-3b^2}}{2}\right\}\sqrt{t}, \qquad [9.11.7]$$

$$b^2 < \frac{4}{3}(1-4c). \qquad [9.11.8]$$

Let us consider [9.11.4]–[9.11.8] and ensure that b and consequently c satisfy these constraints. Firstly, [9.11.5] implies that

$$\frac{b+\sqrt{4(1-4c)-3b^2}}{2} < b$$

$$\Longrightarrow b^2 > 1-4c, \qquad [9.11.9]$$

and this together with [9.11.8] give upper and lower bounds for b^2 which are self-consistent if $c < 1/4$.

In order for this region to be as large as possible it is sensible to define $b^2 = 4(1-6c)/3$, a choice which is clearly compatible with both [9.11.8] and [9.11.9] for $c < 1/48$. With this choice, [9.11.6] and [9.11.7] imply that

$$\frac{4}{3}(1-6c) + 2\sqrt{\frac{4}{3}(1-6c)}\sqrt{4(1-4c)-4(1-6c)} + 4(1-4c) - 4(1-6c) > \frac{8(1+c)}{6-c},$$

and since c is at least less than $1/48$, the above inequality holds if

$$\frac{4}{3} + 2\sqrt{\frac{4}{3}(48c-6c)}\sqrt{8c} > \frac{8(1+c)}{6-c}.$$

Now $2\sqrt{4/3\cdot 42\cdot 8} > 40$ so again the above inequality holds if

$$\frac{4}{3} + 40c > \frac{8(1+c)}{6-c}.$$

This is equivalent to $30c^2 - 173c < 0$, which has solutions $0 < c < 173/30$. Therefore, the choice of $b^2 = 4(1-6c)/3$ with $c < 1/48$ means that X_{UB}^2 satisfies [9.11.6].

The last condition that b must satisfy is [9.11.4]. Now,

$$\begin{aligned} q(b\sqrt{t}) - 4ct \cdot b\sqrt{t} &= \left(b - b^3 - 4cb\right) t\sqrt{t} \\ &= -\frac{1}{3}b(1-12c)t\sqrt{t} \\ &< 0 \end{aligned}$$

for $c < 1/48$. Thus [9.11.4] simplifies to

$$\sqrt{1-6c}(1-12c) < 1,$$

which is identically satisfied for any c less than 1. This proves existence of an $X_{UB} > X_1$ that satisfies [9.11.6].

If we try to determine X_{UB} explicitly as a function of t, [9.11.7] with $b^2 = 4(1-6c)/3$ implies that

$$X_{UB} = \left\{\frac{-\sqrt{4(1-6c)/3} - \sqrt{8c}}{2}\right\}\sqrt{t},$$

giving a rather complicated bound for $[f'(X_1)]^{-1}$. In practice, therefore, a new quantity, $\overline{X}_{UB}$, such that $\overline{X}_{UB} > X_{UB} > X_1$ shall be defined and used to determine bounds for $[f'(X_1)]^{-1}$. Now,

$$\begin{aligned} X_{UB}^2 &= \left\{\frac{4(1-6c)/3 + 8c + 2\sqrt{8c}\sqrt{4(1-6c)/3}}{4}\right\}\sqrt{t} \\ &> \frac{1}{3}(1+30c)t. \end{aligned}$$

If $\overline{X}_{UB} = -\sqrt{(1+30c)t/3}$ then clearly by the above, $\overline{X}_{UB} > X_{UB}$. $\overline{X}_{UB}$ also satisfies [9.11.2] with X_1 replaced by $\overline{X}_{UB}$. To see this we note that in order for

$$\overline{X}_{UB}^2 > \frac{2(1+c)}{6-c}t$$

to be a true statement, we require c to satisfy

$$\frac{1}{3}(1+30c)t > \frac{2(1+c)}{6-c}t$$

$$\Longrightarrow 30c^2 - 173c < 0$$

$$\Longrightarrow 0 < c < \frac{173}{30},$$

which is compatible with current values for c and completes the proof.

With all of the above information gathered by the above theorem it becomes a relatively trivial task to calculate estimates for $[f'(X_1)]^{-1}$. For $c < 1/48$ and $X_c \leq X < \sqrt{4(1-6c)t/3}$, $-X_{cc} < X_1 < X_{UB} < \overline{X}_{UB}$ and [9.11.1] imply that

$$t(1+c) - \frac{1}{2}(6+c)\frac{16(1+c)}{6-c}t < f'(X_1) < t(1+c) - \frac{1}{2}(6-c)\cdot\frac{1}{3}(1+30c)t.$$

These bounds can be simplified by noting that

$$\begin{aligned} 1 + c - \frac{1}{6}(6-c)(1+30c) &= -\frac{173}{6}c + 3c^2 \\ &< -\frac{173}{6}c + 3c \\ &< -25c. \end{aligned}$$

Also, since $c < 1/48$, we have, by [9.3.5],

$$\begin{aligned} 1 + c - \frac{8(6+c)(1+c)}{6-c} &> 1 + c - 8\cdot\frac{10}{9}(1-c) \\ &= -\frac{71}{9} + \frac{89}{9}c \\ &> -7. \end{aligned}$$

Therefore in conclusion,

$$-\frac{1}{25ct} < [f'(X_1)]^{-1} < -\frac{1}{7t} \qquad [9.11.10]$$

for $c < 1/48$ and $X_c \leq X < \sqrt{4(1-6c)t/3}$.

Let us now consider an estimate for $[f'(X_2)]^{-1}$. For the region that is sandwiched between those regions close to the caustic and cocaustic, the range of $X_2(X)$ must include zero. Hence an upper bound *and* a lower bound for X_2 must be found to ensure that X_2 satisfies [9.11.3]. The following two theorems prove the existence of X_{LB} and X_{UB}, which are defined as the lower and upper bounds respectively of X_2. In addition, simpler estimates, $\overline{X}_{LB}$ and $\overline{X}_{UB}$, will be found such that $\overline{X}_{LB} < X_{LB} < X_2 < X_{UB} < \overline{X}_{UB}$. All of these quantities bounding X_2 are required to satisfy [9.11.3] with X_2 replaced and this will also be shown.

Theorem. *Suppose* $X_{UB} = q_2^{-1}\left(q\left(\sqrt{2(1-3c)t/3}\right) + 4ct\sqrt{2(1-3c)t/3}\right)$, $\overline{X}_{UB} = \sqrt{1(1+36c)t/6}$, $\sqrt{2(1-3c)t/3} < X \leq X_{cc}$ *and* $|f(X) - q(X)| < 2ct|X|$, *then* $X_2 < X_{UB} < \overline{X}_{UB}$ *such that* X_{UB}^2 *and* $\overline{X}_{UB}^2$ *are both less than* $2(1-c)t/(6+c)$.

Proof. Let a be such that $a\sqrt{t} < X$ and define $X_{UB} = q_2^{-1}\left(q\left(a\sqrt{t}\right) + 4ct \cdot a\sqrt{t}\right)$. Then since f is decreasing on (X_c, ∞) and increasing on $(-X_c, X_c)$,

$$\begin{aligned}
f(X_2) &= f(X) \\
&< f\left(a\sqrt{t}\right) \\
&\leq q\left(a\sqrt{t}\right) + 2ct \cdot a\sqrt{t} \\
&< q\left(a\sqrt{t}\right) + 4ct \cdot a\sqrt{t} - 2ct\,|X_{UB}| \\
&= q\left(q_2^{-1}\left(q\left(a\sqrt{t}\right) + 4ct \cdot a\sqrt{t}\right)\right) - 2ct\,|X_{UB}| \\
&< f(X_{UB}) \\
\implies X_2 &< X_{UB}
\end{aligned}$$

provided $|X_{UB}| < a\sqrt{t}$. The other constraints that X_{UB} must satisfy are that

$$X_{UB}^2 < \frac{2(1-c)}{6+c}t \qquad [9.11.11]$$

(required to ensure that all positive values of X_2 satisfy [9.11.3]) and

$$\left|q\left(a\sqrt{t}\right) + 4ct \cdot a\sqrt{t}\right| < \frac{2}{3}t\sqrt{\frac{t}{3}} \qquad [9.11.12]$$

(for existence of X_{UB}).

Now from the definition of X_{UB} we have

$$X_{UB} = \left\{\frac{-a + \sqrt{4(1+4c) - 3a^2}}{2}\right\},$$

so clearly a must satisfy $a^2 < 4(1+4c)/3$. Since X_2 has positive and negative values for this region of tX space, it makes sense to assume that X_{UB} is positive and as large as possible. This assumption implies that $a^2 < 1 + 4c$, which of course supersedes the above estimate. With this choice of a, $|X_{UB}| < a\sqrt{t}$ implies that $a^2 > (1+4c)/3$. Thus in summary,

$$\frac{1}{3}(1+4c) < a^2 < 1 + 4c. \qquad [9.11.13]$$

To ensure that this region of tX space is as large as possible we need to make a as small as possible. The obvious choice is to make a^2 only slightly larger than its lower bound, such as $a^2 = (1+6c)/3$ for example. However, it turns out that this choice of a means that $q\left(a\sqrt{t}\right) + 4ct \cdot a\sqrt{t}$ is no longer within the domain of q_2^{-1}. In fact any choice for a of the form, $a^2 = 1/3 + O(|c|)$, implies that X_{UB} cannot be defined. The details of this remark can be explained quite easily: if $a^2 = 1/3 + O(|c|)$, then $q\left(a\sqrt{t}\right) + 4ct \cdot a\sqrt{t} = a(1-a^2+4c)t\sqrt{t}$ evaluated at $c = 0$ equals $2t/3\sqrt{t/3}$. Moreover this cubic in c has a positive gradient in a neighbourhood of the origin. This means that in order for [9.11.12] to be satisfied we must prescribe a minimum value for c, a procedure which is rather incompatible with the contraction mapping proof one is trying to formulate.

For the above reason, a^2 is chosen to be of the form $a^2 = 2/3 + O(|c|)$. Indeed if $a^2 = 2(1-3c)/3$, for example, then the cubic, $a(1-a^2+4c)$, evaluated at $c = 0$ equals $1/3\sqrt{2/3}$ and hence even with a positive gradient, [9.11.12] can be satisfied for small enough c. To see this we have

$$q\left(a\sqrt{t}\right) + 4ct \cdot a\sqrt{t} = \frac{1}{3}(1+18c)\sqrt{\frac{2}{3}(1-3c)} \cdot t\sqrt{t},$$

which is positive for $c < 1/48$. [9.11.12] then becomes $(1-3c)(1+18c)^2 < 2$, which is equivalent to $972c^3 - 216c^2 - 33c + 1 > 0$. This inequality is satisfied for $c < 1/48$.

Finally X_{UB} must satisfy [9.11.11]. This requires that

$$\frac{2}{3}(1-3c) - 2\sqrt{\frac{2}{3}(1-3c)}\sqrt{2+20c} + 2 + 20c < \frac{8(1-c)}{6+c}. \qquad [9.11.14]$$

Now,

$$\begin{aligned}\sqrt{\frac{2}{3}(1-3c)}\sqrt{2+20c} &= \sqrt{\frac{2}{3}(1-3c)(2+20c)} \\ &> 1-3c\end{aligned}$$

meaning that [9.11.14] is satisfied if

$$\frac{2}{3}(1+36c) < \frac{8(1-c)}{6+c}$$

$$\implies 36c^2 + 229c - 6 < 0. \qquad [9.11.15]$$

The equation $36x^2+229x-6=0$ has solutions $x=(-229\pm\sqrt{53305})/72$. So [9.11.15] is satisfied if

$$\frac{-229\pm\sqrt{53305}}{72}=\frac{-458\pm\sqrt{213220}}{144}>\frac{-458\pm\sqrt{212521}}{144}=\frac{1}{48}>c.$$

In other words for $c<1/48$, X_{UB} satisfies [9.11.11] if a^2 is chosen to be $2(1-3c)/3$. Moreover, since

$$X_{UB}^2<\frac{1}{6}(1+36c)t<\frac{2(1-c)}{6+c}t,$$

we can define $\overline{X}_{UB}=\sqrt{(1+36c)t/6}$ and use this instead of X_{UB} when estimating $[f'(X_2)]^{-1}$.

Theorem. *Supposing* $X_{LB}=q_2^{-1}\left(q\left(\sqrt{4(1-6c)t/3}\right)-4ct\sqrt{4(1-6c)t/3}\right)$, $\overline{X}_{LB}=-\sqrt{1(1-30c)t/3}$, $X_c\leq X<\sqrt{4(1-6c)t/3}$ *and* $|f(X)-q(X)|<2ct|X|$, *then* $\overline{X}_{LB}<X_{LB}<X_2$ *such that* X_{LB}^2 *and* $\overline{X}_{LB}^2$ *are both less than* $2(1-c)t/(6+c)$.

Proof. Let b be such that $X<b\sqrt{t}$ and define $X_{LB}=q_2^{-1}\left(q\left(b\sqrt{t}\right)-4ct\cdot b\sqrt{t}\right)$. Then since f is decreasing on (X_c,∞) and increasing on $(-X_c,X_c)$,

$$\begin{aligned}
f(X_2)&=f(X)\\
&>f\left(b\sqrt{t}\right)\\
&>q\left(b\sqrt{t}\right)-2ct\cdot b\sqrt{t}\\
&>q\left(b\sqrt{t}\right)-4ct\cdot b\sqrt{t}+2ct|X_{LB}|\\
&=q\left(q_2^{-1}\left(q\left(b\sqrt{t}\right)-4ct\cdot b\sqrt{t}\right)\right)+2ct|X_{LB}|\\
&>f\left(X_{LB}\right)\\
\implies X_2&>X_{LB}
\end{aligned}$$

provided $|X_{LB}|<b\sqrt{t}$. In addition, X_{LB} must satisfy

$$X_{LB}^2<\frac{2(1-c)}{6+c}t \qquad [9.11.16]$$

and for the existence of X_{LB},

$$\left|q\left(b\sqrt{t}\right)-4ct\cdot b\sqrt{t}\right|<\frac{2}{3}t\sqrt{\frac{t}{3}}. \qquad [9.11.17]$$

Now the functional form of X_{LB} is

$$X_{LB} = \left\{ \frac{-b + \sqrt{4(1-4c) - 3b^2}}{2} \right\} \sqrt{t}, \qquad [9.11.18]$$

so clearly $b^2 < 4(1-4c)/3$ in order for the square root to be real. Again, since $X_2(X)$ can be positive or negative, X_{LB} is assumed to be less than zero requiring that $b^2 > 1 - 4c$. This then means that

$$|X_{LB}| = -X_{LB} < b\sqrt{t}$$

$$\Longrightarrow b^2 > \frac{1}{3}(1-4c).$$

In summary then,

$$\frac{1}{3}(1-4c) < b^2 < \frac{4}{3}(1-4c).$$

Choosing $b^2 = 4(1-6c)/3$ is valid for $c < 1/48$ and means that the region of tX space between X_c and X_{cc} can be made as large as possible. With this choice, [9.11.17] is identically satisfied for

$$q(b\sqrt{t}) - 4ct \cdot b\sqrt{t} = -\frac{1}{3}(1-12c)\sqrt{\frac{4}{3}(1-6c)}t\sqrt{t}$$

$$\Longrightarrow \left|q\left(b\sqrt{t}\right) - 4ct \cdot b\sqrt{t}\right| < \frac{1}{3}\sqrt{\frac{4}{3}}t\sqrt{t}$$

$$= \frac{2}{3}t\sqrt{\frac{t}{3}}.$$

The final condition to consider is [9.11.16]. Using [9.11.18],

$$X_{LB}^2 < \frac{1}{3}(1-30c)t$$

and [9.11.16] becomes

$$\frac{1}{3}(1-30c) < \frac{2(1-c)}{6+c}$$

$$\Longrightarrow 30c^2 + 173c > 0.$$

This inequality holds for all positive c meaning that X_{LB} must satisfy [9.11.16]. In addition to this, since

$$X_{LB}^2 < \frac{1}{3}(1-30c)t < \frac{2(1-c)}{6+c}t,$$

we can define $\overline{X}_{LB} = -\sqrt{(1-30c)t/3}$ such that $\overline{X}_{LB} < X_{LB} < X_2$ and $\overline{X}_{LB}^2$ is less than $2(1-c)/(6+c)$, thus proving the theorem.

Pulling together the information that is contained in the above two theorems we have, for $c < 1/48$ and $\sqrt{2(1-3c)t/3} < X < \sqrt{4(1-6c)t/3}$, $-\sqrt{(1-30c)t/3} < X_2 < \sqrt{(1+36c)t/6}$. Depending on the range of values for c we are allowed, there is a choice of upper bound for X_2^2: if $1/96 < c < 1/48$ then we must choose $X_2^2 < (1+36c)t/6$ and if $c < 1/96$ we must choose $X_2^2 < (1-30c)t/3$. For flexibility, however, a lower bound for c is not desirable. Hence the range of c is reduced further to $c < 1/96$ so that $0 < X_2^2 < (1-30c)t/3$. Finally then, using [9.11.1], we obtain

$$\left(\frac{163}{6}c + 5c^2\right)t < f'(X_2) < (1+c)t$$

$$\Longrightarrow 33ct < f'(X_2) < (1+c)t$$

$$\Longrightarrow \frac{1}{(1+c)t} < [f'(X_2)]^{-1} < \frac{1}{33ct}. \qquad [9.11.19]$$

§9.12. Summary of estimates for $[f'(X_1(t,X))]^{-1}$ and $[f'(X_2(t,X))]^{-1}$ for the region where X is bounded away from the caustic and cocaustic ($X_c < X < X_{cc}$).

We have

$$-\frac{1}{25ct} < [f'(X_1)]^{-1} < -\frac{1}{7t} \qquad [9.11.10]$$

for $c < 1/48$ and $X_c \le X < \sqrt{4(1-6c)t/3}$, and

$$\frac{1}{(1+c)t} < [f'(X_2)]^{-1} < \frac{1}{33ct} \qquad [9.11.19]$$

for $c < 1/96$ and $\sqrt{2(1-3c)t/3} < X < \sqrt{4(1-6c)t/3}$. Hence, in order for the above estimates to be valid simultaneously on the region between both the caustic and cocaustic, we define this region to be $\left\{(X,t) \mid \sqrt{2(1-3c)t/3} < X < \sqrt{4(1-6c)t/3}\right\}$.

§9.13. Estimates for $[f'(X_1(t,X))]^{-1}$ for $X < X_c$ near the caustic.

In this section estimates for $[f'(X_1)]^{-1}$ for $X \le X_c$ are calculated. Since X is close to the caustic, $[f'(X_1)]^{-1}$ is expected to be well behaved and thus expanding this quantity in terms of $|t - k_c|$ is not necessary. Having said this, however, we will be forced to modify our approach slightly to account for the fact that since $-X_c \le X \le X_c$, we can never have $|X_{UB}| < a\sqrt{t}$, a requirement that we have previously had to satisfy. The following paragraph explains.

Since $f'(X_1)$ is expected to be non-zero for these values of X we have, using [9.5.3],

$$(1-c)t - \frac{1}{2}(6+c)X_1^2 \le f'(X_1) \le (1+c)t - \frac{1}{2}(6-c)X_1^2.$$

With this approach we immediately have the condition that

$$X_1^2 > \frac{2(1+c)}{6-c}t, \qquad [9.13.1]$$

which ensures that $[f'(X_1)]^{-1}$ is negative. If this assumption was not made $[f'(X_1)]^{-1}$ could become unbounded. Clearly an upper bound, X_{UB} say, for X_1 as a function of X is needed such that [9.13.1] is satisfied with X_{UB}^2 replacing X_1^2. Continuing with the theme from previous sections, we have a method that proves existence of this upper bound, however, if we recall, this process depends on the fact that a, which defines the lower bound for the region containing X (i.e. $a\sqrt{t} < X < b\sqrt{t}$), satisfies $|X_{UB}| < a\sqrt{t}$. Clearly, since X is in the region where it could be relabelled as X_2 and that $X_1 \le -X_c < a\sqrt{t} < X_2 \le X_c$, $|X_{UB}|$ is *necessarily* greater than $a\sqrt{t}$. This requires the existence proof for X_{UB} to be modified. Instead we will require that $|X_{UB}|$ is less than n multiples of $a\sqrt{t}$ ($|X_{UB}| < na\sqrt{t}$) which is clearly true for large enough n. This amounts to redefining the curves q_α and q_β such that $q_\alpha(X,t) = q(X,(1-2c(1+n))t)$ and $q_\beta(X,t) = q(X,(1+2c(1+n))t)$. The following theorem determines possible values for n, compatible with current restrictions on c, that prove existence of X_{UB}.

Theorem. *If* $X_{UB} = q_1^{-1}\left(q\left(\sqrt{t}/8\right) - t\sqrt{t}/12\right)$, $\overline{X}_{UB} = -\sqrt{3t/8}$, $1/8\sqrt{t} < X \le X_c$ *and* $|f(X) - q(X)| < 2ct|X|$, *then* $X_1 < X_{UB} < \overline{X}_{UB}$ *such that* X_{UB}^2 *and* $\overline{X}_{UB}^2$ *are both greater than* $2(1+c)t/(6-c)$.

Proof. Suppose $X > a\sqrt{t}$ and define $X_{UB} = q_1^{-1}\left(q\left(a\sqrt{t}\right) - 2c(1+n)t \cdot a\sqrt{t}\right)$ where $n > 0$. Then, since f is decreasing on $(-X_c, X_c)$ and increasing on $(-X_{cc}, -X_c)$, and providing $na\sqrt{t} > |X_{UB}|$,

$$\begin{aligned} f(X_1) &= f(X) \\ &> f\left(a\sqrt{t}\right) \\ &> q\left(a\sqrt{t}\right) - 2ct \cdot a\sqrt{t} \\ &> q\left(a\sqrt{t}\right) - 2ct \cdot a\sqrt{t} - 2nct \cdot a\sqrt{t} + 2ct|X_{UB}| \\ &= q\left(q_1^{-1}\left(q\left(a\sqrt{t}\right) - 2c(1+n)t \cdot a\sqrt{t}\right)\right) + 2ct|X_{UB}| \\ &= q\left(X_{UB}\right) + 2ct|X_{UB}| \\ &> f(X_{UB}) \\ \implies X_1 &< X_{UB}. \end{aligned}$$

The similarities between this and the previous sections can be seen. The only difference is that values for n must be determined that allow the above procedure to go through. The same arguments as those in the previous section give rise to the following constraints:

$$\left|q\left(a\sqrt{t}\right) - 2c(1+n)t \cdot a\sqrt{t}\right| < \frac{2}{3}\sqrt{\frac{t}{3}}, \qquad [9.13.2]$$

$$X_{UB} = \left\{\frac{-a - \sqrt{4(1-2c(1+n)) - 3a^2}}{2}\right\}\sqrt{t}, \qquad [9.13.3]$$

$$na\sqrt{t} > |X_{UB}| \qquad [9.13.4]$$

and

$$X_{UB}^2 > \frac{2(1+c)}{6-c}t. \qquad [9.13.5]$$

In addition to these equations, [9.13.3] implies that

$$a^2 < \frac{4}{3}(1 - 2c(1+n)) \qquad [9.13.6]$$

and

$$n < \frac{1}{2c} - 1. \qquad [9.13.7]$$

The conditions [9.13.2]–[9.13.6] are of course expressing the criteria which have been discussed before. [9.13.7], however, is new. Clearly the above procedure is constructing a curve $q_\alpha(X,t) = q(X,(1-2c(1+n))t)$ which is a lower or upper estimate for f when X is positive or negative respectively. Moreover this curve *looks like* $q = Xt - X^3$. If on the other hand [9.13.7] is not satisfied then $q_\alpha(X,t) = q(X, -|1-2c(1+n)|t)$, which would look like $q = -Xt - X^3$. Although this is compatible with the initial conditions one might wish to prescribe, it does not allow for a caustic set near the origin for positive t.

Considering first of all [9.13.4], which should give a lower bound for a^2, we have from [9.13.3],

$$2na > a + \sqrt{4(1-2c(1+n)) - 3a^2}$$

$$\Longrightarrow (2n-1)^2 a^2 > 4(1-2c(1+n)) - 3a^2$$

$$\Longrightarrow (n^2 - n + 1)a^2 > 1 - 2c(1+n). \qquad [9.13.8]$$

To proceed we need to fix a^2 and n as functions of c. Firstly, choose $a^2 = 1/64$; this is motivated by the fact that $a\sqrt{t}$ must be significantly less than X_c. In other words for small enough c we have the inequality, $a\sqrt{t} < 2(1-c)/(6+c) < X_c$, ensuring that the region close to the caustic such that $X \le X_c$ is not an empty set. Secondly, choose $n = 1/3c - 1$. This is the largest value for n permissible by [9.13.7]. By initially choosing such a large value, [9.13.4] is more likely to be satisfied without further adjusting the upper bound on c.

With these values, [9.13.8] becomes $165c^2 + 9c - 1 < 0$ implying that $0 < c < 1/19$. With current values of c being less that $1/96$ we conclude that [9.13.8], and hence [9.13.4], is satisfied with these choices of a^2 and n.

Let us now consider [9.13.5]. By [9.13.3] we have

$$a^2 + 2a\sqrt{4\left(1-2c(1+n)\right) - 3a^2} + 4\left(1-2c(1+n)\right) - 3a^2 > \frac{8(1+c)}{6-c} \qquad [9.13.9]$$

$$\Longrightarrow \frac{1}{64} + \frac{1}{4}\sqrt{\frac{247}{192}} + \frac{247}{192} > \frac{8(1+c)}{6-c}.$$

This inequality is clearly satisfied if

$$\frac{1}{64} + \frac{1}{4}\sqrt{\frac{5}{4}} + \frac{5}{4} > \frac{8(1+c)}{6-c},$$

and again if

$$\frac{1}{64} + \frac{1}{4} + \frac{5}{4} > \frac{8(1+c)}{6-c}$$

$$\Longrightarrow \frac{1}{64} + \frac{1}{4} \cdot \frac{5}{4} + \frac{5}{4} > \frac{8(1+c)}{6-c}$$

$$\Longrightarrow \frac{3}{2} > \frac{8(1+c)}{6-c} \qquad [9.13.10]$$

$$\Longrightarrow c < \frac{2}{19}.$$

Thus for $c < 1/96$, $a^2 = 1/64$ and $n = 1/3c - 1$, [9.13.5] is satisfied.

The last condition to consider is [9.13.2]. Now,

$$q(a\sqrt{t}) - 2c(1+n)t \cdot a\sqrt{t} = at\sqrt{t} - \left(a\sqrt{t}\right)^3 - 2c(1+n)t \cdot a\sqrt{t}$$

and so [9.13.2] is equivalent to

$$|a(1 - a^2 - 2c(1+n))| < \frac{2}{3}\sqrt{\frac{1}{3}}$$

$$\Longrightarrow \frac{61}{1536} < \frac{2}{3}\sqrt{\frac{1}{3}},$$

which is a true statement.

Having proved existence of X_{UB} a more elegant estimate, $\overline{X}_{UB}$, such that $\overline{X}^2_{UB}$ is greater than $2(1+c)/(6-c)$ can be established. Now [9.13.9] and [9.13.10] together imply that

$$X^2_{UB} > \frac{3}{8}t > \frac{2(1+c)}{6-c}t$$

for all $c < 1/96$. This establishes the quantity $\overline{X}_{UB} = -\sqrt{3t/8}$ and proves the theorem.

We are now in a position to estimate $[f'(X_1)]^{-1}$. Since $-X_{cc} \le X < \overline{X}_{UB}$, [9.5.3] implies that

$$(1-c)t - \frac{1}{2}\frac{16(1+c)(6+c)}{6-c}t < f'(X_1) < (1+c) - \frac{1}{2}(6-c)\frac{3}{8}t$$

using [9.3.10]. [9.3.5] gives, for $c < 1/24$,

$$\frac{(1+c)(6+c)}{6-c} < \frac{10}{9}(1-c)$$

$$\Longrightarrow \left\{(1-c) - \frac{8\cdot 10}{9}(1-c)\right\}t < f'(X_1) < \left\{(1+c) - \frac{1}{2}(6-c)\frac{3}{8}\right\}t$$

$$\Longrightarrow -8(1-c)t < f'(X_1) < -\frac{1}{8}(1-10c)t$$

$$\Longrightarrow -\frac{8}{(1-10c)t} < [f'(X_1)]^{-1} < -\frac{1}{8(1-c)t}. \qquad [9.13.11]$$

This inequality holds for $1/8\sqrt{t} < X \le X_c$ and $c < 1/96$.

§9.14. Estimates for $[f'(X_3(t,X)))]^{-1}$ for $X < X_c$ near the caustic.

In this region $[f'(X_3)]^{-1}$ is expected to become unbounded as X approaches X_c. We therefore estimate this quantity in terms of the distance from the caustic. Now the mean value theorem gives

$$f'(X_3) = |X_3 - X_c| f''(Y_a)$$

with $X_c < Y_a < X_3$. Similarly, $f(X) = f(X_3)$ implies that

$$|X_3 - X_c| = |X - X_c| \frac{|f''(Y_b)|^{1/2}}{|f''(Y_c)|^{1/2}}$$

where $X < Y_b < X_c$ and $X_c < Y_c < X_3$. Also,

$$|t - k_c| = |X - X_c| \left|k_c'(Y_d)\right|$$

where $X < Y_d < X_c$. Combining these results the expression for $[f'(X_3)]^{-1}$ becomes

$$[f'(X_3)]^{-1} = -\frac{|f''(Y_c)|^{1/2}\,|k_c'(Y_d)|}{|f''(Y_a)||f''(Y_b)|^{1/2}|t-k_c|}. \qquad [9.14.1]$$

As usual equation [9.5.2] implies that

$$(6-c)X_c < (6-c)|Y_a| \leq |f'(Y_a)| \leq (6+c)|Y_a| < (6+c)X_3.$$

Similarly,

$$(6-c)X < |f'(Y_b)| < (6+c)X_c$$

and

$$(6-c)X_c < |f'(Y_c)| < (6+c)X_3.$$

In addition, [9.2.2] implies that

$$\frac{6-c}{1+c}X < \frac{6-c}{1+c}|Y_d| \leq \left|k_c'(Y_d)\right| < \frac{6+c}{1-c}X_c.$$

If we insert these estimates into [9.14.1] then we obtain

$$-\frac{(6+c)^{3/2}X_3^{1/2}}{(6-c)^{3/2}(1+c)X^{1/2}|t-k_c|} < [f'(X_3)]^{-1} < -\frac{(6-c)^{3/2}X}{(6+c)^{3/2}(1-c)X_3|t-k_c|},$$

which, using [9.3.5], implies that

$$-\frac{10^{3/2}X_3^{1/2}}{9^{3/2}(1+c)^{3/2}X^{1/2}|t-k_c|} < [f'(X_3)]^{-1} < -\frac{9^{3/2}X}{10^{3/2}(1-c)X_3|t-k_c|}$$

$$\Longrightarrow -\frac{3X_3^{1/2}}{2(1+c)X^{1/2}|t-k_c|} < [f'(X_3)]^{-1} < -\frac{2X}{3(1-c)X_3|t-k_c|}. \qquad [9.14.2]$$

To simplify these estimates further upper and lower bounds for X and X_3 are needed. In this case, for estimates of X_3 the straight forward choice of $X_c \leq X_3 \leq X_{cc}$ is sufficient. Moreover, we only have the condition that X is bounded away from zero and less than X_c so there is no reason why we cannot use the lower bound, $1/8\sqrt{t}$, determined by the previous section. It follows then that [9.14.2] becomes

$$-\frac{3\sqrt{2}X_{cc}^{1/2}}{(1+c)t^{1/4}|t-k_c|} < [f'(X_3)]^{-1} < -\frac{\sqrt{t}}{12(1-c)X_{cc}|t-k_c|},$$

and using [9.3.10] this becomes

$$- \frac{3\sqrt{2}\sqrt{\sqrt{16(1+c)(6-c)^{-1}t}}}{(1+c)t^{1/4}|t-k_c|} < [f'(X_3)]^{-1}$$
$$< -\frac{\sqrt{t}}{12(1-c)\sqrt{16(1+c)(6-c)^{-1}t}|t-k_c|}$$

$$\implies -\frac{6\sqrt{2}}{(1+c)^{3/4}(6-c)^{1/4}|t-k_c|} < [f'(X_3)]^{-1} < -\frac{(6-c)^{1/2}}{48(1+c)^{3/2}|t-k_c|} \qquad [9.14.3]$$

for all $1/8\sqrt{t} \le X \le X_c$ and $c < 1/24$.

§9.15. Summary of estimates for $[f'(X_1(t,X))]^{-1}$ and $[f'(X_2(t,X))]^{-1}$ for the region where X is close to the caustic ($X \le X_c$).

We have

$$-\frac{8}{(1-10c)t} < [f'(X_1)]^{-1} < -\frac{1}{8(1-c)t} \qquad [9.13.11]$$

and

$$-\frac{6\sqrt{2}}{(1+c)^{3/4}(6-c)^{1/4}|t-k_c|} < [f'(X_3)]^{-1} < -\frac{(6-c)^{1/2}}{48(1+c)^{3/2}|t-k_c|} \qquad [9.14.3]$$

for all $1/8\sqrt{t} \le X \le X_c$ and $c < 1/96$.

§9.16. Estimates for $[f'(X_1(t,X))]^{-1}$ and $[f'(X_3(t,X))]^{-1}$ for X close to the origin.

Let us consider estimates for $[f'(X_3)]^{-1}$ first of all as by the asymmetry of f, these will automatically imply bounds on $[f'(X_1)]^{-1}$. Since in this region X is close to the origin, it should be possible to bound X_3 away from X_c by reducing the upper bound on X. This means that finite estimates for $[f'(X_3)]^{-1}$ can be obtained because we can appeal directly to [9.5.7] rather than expanding in terms of $|t-k_c|$.

Now, [9.5.7] implies that

$$(1-c)t - \frac{1}{2}(6+c)X_3^2 \le f'(X_3) \le (1+c)t - \frac{1}{2}(6-c)X_3^2.$$

Clearly, if X_3 is sufficiently far from X_c then $f'(X_3)$ will always be negative. We shall therefore find an upper bound on X which defines a lower bound for X_3, denoted by X_{LB} say, which is greater than $2(1+c)t/(6-c)$. These conditions are proved in the following theorem.

Theorem. *If* $X_{LB} = q_3^{-1}\left(q\left(1/2\sqrt{t}\right) + 3ct\sqrt{t}\right)$, $\overline{X}_{LB} = \sqrt{7t/16}$, $0 \leq X < 1/2\sqrt{t}$ *and* $|f(X) - q(X)| < 2ct|X|$, *then* $X_3 > X_{LB} > \overline{X}_{LB}$ *such that* X_{LB}^2 *and* $\overline{X}_{LB}^2$ *are both greater than* $2(1+c)t/(6-c)$.

Proof. Define $X_{LB} = q_3^{-1}\left(q\left(b\sqrt{t}\right) + 2c(1+n)t \cdot b\sqrt{t}\right)$ where b is a constant chosen so that $X < b\sqrt{t}$. Then, since f is increasing on $(-X_c, X_c)$ and decreasing on (X_c, X_{cc}),

$$\begin{aligned} f(X_3) &= f(X) \\ &< f(b\sqrt{t}) \\ &< q(b\sqrt{t}) + 2ct \cdot b\sqrt{t} \\ &< q(b\sqrt{t}) + 2ct \cdot b\sqrt{t} + 2nct \cdot b\sqrt{t} - 2ct|X_{LB}| \\ &= q\left(q_3^{-1}\left(q\left(b\sqrt{t}\right) + 2c(1+n)t \cdot b\sqrt{t}\right)\right) 2ct|X_{LB}| \\ &= q\left(X_{LB}\right) - 2ct|X_{LB}| \\ &< f\left(X_{LB}\right) \\ \Longrightarrow X_3 &> X_{LB} \end{aligned}$$

if $nb\sqrt{t} > |X_{LB}|$. The conditions that must be satisfied are:

$$\left|q\left(b\sqrt{t}\right) + 2c(1+n)t \cdot b\sqrt{t}\right| < \frac{2}{3}\sqrt{\frac{t}{3}}, \qquad [9.16.1]$$

$$X_{LB} = \left\{\frac{-b + \sqrt{4(1+2c(1+n)) - 3b^2}}{2}\right\}\sqrt{t} \qquad [9.16.2]$$

$$\Longrightarrow b^2 < 1 + 2c(1+n), \qquad [9.16.3]$$

which ensures that X_{LB} is positive,

$$nb\sqrt{t} > |X_{LB}| \qquad [9.16.4]$$

and

$$X_{LB}^2 > \frac{2(1+c)}{6-c}t. \qquad [9.16.5]$$

Consider [9.16.4] first of all. We have

$$(2n+1)^2b^2 > 4(1+2c(1+n)) - 3b^2.$$

Define $n = 2$ and $b^2 = 1/4$. Then with these choices,

$$(n^2+n+1)b^2 > 1 + 2c(1+n)$$

$$\Longrightarrow \frac{7}{4} > 1 + 6c,$$

which is true for all $c < 1/8$.

Now consider [9.16.5]. With [9.16.2] this becomes

$$b^2 - 2b\sqrt{4(1+2c(1+n)) - 3b^2} + 4(1+2c(1+n)) - 3b^2 > \frac{8(1+c)}{6-c}$$

$$\Longrightarrow \frac{1}{4} - \sqrt{4(1+6c) - \frac{3}{4}} + 4(1+6c) - \frac{3}{4} > \frac{8(1+c)}{6-c}.$$

If $c < 1/96$ then this inequality is satisfied if

$$\frac{1}{4} - \sqrt{\frac{7}{2}} + \frac{7}{2} > \frac{8(1+c)}{6-c},$$

which in turn is satisfied if

$$\frac{1}{4} - 2 + \frac{7}{2} > \frac{8(1+c)}{6-c}$$

$$\Longrightarrow \frac{7}{4} > \frac{8(1+c)}{6-c} \qquad [9.16.6]$$

$$\Longrightarrow c < \frac{1}{4},$$

which current restrictions on c satisfy.

Finally consider [9.16.1]. This inequality becomes, with the above choices for n and b,

$$\frac{3}{8} + 3c < \frac{2}{3}\sqrt{\frac{1}{3}}$$

$$\Longrightarrow \frac{3}{8} + 3c < \frac{1}{3}$$

$$\Longrightarrow c < \frac{1}{5},$$

which is again satisfied by $c < 1/96$.

Having proved existence of X_{LB}, $\overline{X}_{LB}$ can now be defined. From [9.16.6],

$$X_{LB}^2 > \frac{7}{16}t > \frac{2(1+c)}{6-c}t$$

implying that $\overline{X}_{LB} = \sqrt{7t/16}$ is a valid choice.

With these estimates for X_3, we can now establish estimates for $[f'(X_3)]^{-1}$. Equation [9.5.7] implies that

$$(1-c)t - \frac{1}{2}(6+c)X_{cc}^2 \leq f'(X_3) < (1+c)t - \frac{1}{2}(6-c)X_{LB}^2$$

$$\Longrightarrow (1-c)t - \frac{1}{2}(6+c)\frac{16(1+c)}{6-c}t < f'(X_3) < (1+c)t - \frac{1}{2}(6-c)\frac{7}{16}t.$$

Using [9.3.5] this becomes

$$(1-c)t - \frac{8\cdot 10}{9}(1-c)t < f'(X_3) < -\frac{5}{32}(2-5c)t$$

$$\Longrightarrow -\frac{32}{5(2-5c)t} < [f'(X_3)]^{-1} < -\frac{1}{8(1-c)t} \qquad [9.16.7]$$

for all $0 \leq X < 1/2\sqrt{t}$ and $c < 1/96$.

This result can be used to immediately estimate $[f'(X_1)]^{-1}$. Since $X \geq 0$, it follows that X_3 must be closer to X_c than X_1 is to $-X_c$. This means that if $X_{UB} = -X_{LB} = -\sqrt{7t/16}$, then for all $0 \leq X < 1/2\sqrt{t}$ and $c < 1/96$, $X_1 < X_{UB}$ and

$$-\frac{32}{5(2-5c)t} < [f'(X_1)]^{-1} < -\frac{1}{8(1-c)t}. \qquad [9.16.8]$$

§9.17. Summary of estimates for $[f'(X_1(t,X))]^{-1}$ and $[f'(X_3(t,X))]^{-1}$ for X near the origin.

We have

$$-\frac{32}{5(2-5c)t} < [f'(X_i)]^{-1} < -\frac{1}{8(1-c)t} \quad i = 1,3 \qquad \text{[9.16.7] and [9.16.8]}$$

for all $0 \leq X < 1/2\sqrt{t}$ and $c < 1/96$.

APPENDIX 1.

§A1.1. Presentation of formulae for transforming differential equations in *tx* space to differential equations in *tv* space.

In this appendix, four relations regarding the partial derivatives of functions with two independent variables are derived. These shall be used in Chapter 4 to transform equations [4.1.1]-[4.1.3], which are differential equations with respect to t and x, to equivalent equations with t and v as the independent variables.

If $x = x(v(t,v),t)$ then

$$\left(\frac{\partial x}{\partial x}\right)_t = \left(\frac{\partial x}{\partial v}\right)_t \left(\frac{\partial v}{\partial x}\right)_t + \left(\frac{\partial x}{\partial t}\right)_v \left(\frac{\partial t}{\partial x}\right)_t .$$

Now we have $(\partial t/\partial x)_t = 0$ and $(\partial x/\partial x)_t = 1$, hence

$$\left(\frac{\partial v}{\partial x}\right)_t = \frac{1}{(\partial x/\partial v)_t}. \qquad [A1.1.1]$$

Similarly,

$$\left(\frac{\partial x}{\partial t}\right)_x = \left(\frac{\partial x}{\partial v}\right)_t \left(\frac{\partial v}{\partial t}\right)_x + \left(\frac{\partial x}{\partial t}\right)_v \left(\frac{\partial t}{\partial t}\right)_x .$$

But $(\partial x/\partial t)_x = 0$ and $(\partial t/\partial t)_x = 1$ so that

$$\left(\frac{\partial x}{\partial t}\right)_v = -\left(\frac{\partial x}{\partial v}\right)_t \left(\frac{\partial v}{\partial t}\right)_x . \qquad [A1.1.2]$$

If $f = f(v(t,v),t)$ then

$$\left(\frac{\partial f}{\partial x}\right)_t = \left(\frac{\partial f}{\partial v}\right)_t \left(\frac{\partial v}{\partial x}\right)_t + \left(\frac{\partial f}{\partial t}\right)_v \left(\frac{\partial t}{\partial x}\right)_t$$

$$\Longrightarrow \left(\frac{\partial f}{\partial x}\right)_t = \left(\frac{\partial f}{\partial v}\right)_t \left(\frac{\partial v}{\partial x}\right)_t , \qquad [A1.1.3]$$

and similarly,

$$\left(\frac{\partial f}{\partial t}\right)_x = \left(\frac{\partial f}{\partial v}\right)_t \left(\frac{\partial v}{\partial t}\right)_x + \left(\frac{\partial f}{\partial t}\right)_v \left(\frac{\partial t}{\partial t}\right)_x$$

$$\Longrightarrow \left(\frac{\partial f}{\partial t}\right)_x = \left(\frac{\partial f}{\partial v}\right)_t \left(\frac{\partial v}{\partial t}\right)_x + \left(\frac{\partial f}{\partial t}\right)_v . \qquad [A1.1.4]$$

APPENDIX 2.

§A2.1. Metric spaces and contraction mappings [TA].

In chapters 8 and 9 we describe the setting up of an existence proof for the solution to the differential equations defining the Newtonian formulation of caustics in a spacetime. The method which we base this existence proof on is to use the fixed point theorems pertaining to contraction mappings on metric spaces. This will be explained in detail within the main text; the point of this appendix, however, is to briefly introduce the reader who is unfamiliar with such tools, to a few relevant definitions and an example that neatly illustrates what we are trying to do with the main equations. We begin by stating the following

Definition. *A non-empty set, S, of objects together with a function, $d_S: S \times S \longrightarrow \mathbb{R}$, satisfying:*

1. $d_S(x,x) = 0$,

2. $d_S(x,y) > 0$ *if* $x \neq y$,

3. $d_S(x,y) = d_S(y,x)$,

4. $d_S(x,y) \leq d_S(x,z) + d_S(z,y)$,

$\forall x, y, z \in S$ *is called a metric space and denoted by* (S, d_S). *The function,* d_S, *is called the metric.*

Two possible examples of a metric space are now given. The first corresponds to the case where $S = \mathbb{R}$ and $d_S(x,y) = |x - y|$. This illustrates the fact that d_S is to be thought of as the distance from x to y; properties 1 through 3. Another example, more relevant to the case that we consider in the main text, is that where $S = C[a,b]$, the set of continuous functions on $[a,b]$. Here the metric is given by

$$d_S(\phi_1, \phi_2) = \max_{a \leq x \leq b} |\phi_1(x) - \phi_2(x)|. \qquad [A2.1.1]$$

In both of these examples we can easily show that items 1 through 4 are satisfied.

Definition. *A sequence, $\{x_n\}$, in a metric space, (S, d_S), is called a Cauchy sequence if, for every $\varepsilon > 0$, there exists an integer, N, such that $d_S(x_n, x_m) < \varepsilon$ whenever $n \geq N$ and $m \geq N$.*

It is possible, from the definition of a convergent sequence, to prove that all convergent sequences are Cauchy sequences. The point of introducing this, however, is that we can now give the idea of a complete metric space.

Definition. *A metric space is complete if every Cauchy sequence in S converges in S.*

For the above two examples we can show that both $(\mathbb{R}, d_S)$ and $(C[a,b], d_S)$ are complete.

With these ideas the concept of a contraction mapping on (S, d_S) can be defined. Furthermore, we now have sufficient information to allow us to state Banach's fixed-point theorem.

Definition. *Let $J: S \longrightarrow S$ be a map of (S, d_S) onto itself. Then J is called a contraction of S if $d_S(J[x], J[y]) < k d_S(x, y)$ $\forall x, y \in S$ and $k < 1$.*

Theorem (Banach's fixed-point theorem). *A contraction, J, of a complete metric space, S, has a unique fixed point, i.e. there exists a point $x \in S$ such that $J[x] = x$.*

§A2.2. Fixed-point theorems and existence proofs.

In this section we shall illustrate the power of fixed-point theorems and contraction mappings by proving an existence theorem for the solution to a particular class of ordinary differential equations. That is to say, let $f(x, y)$ be a real valued function defined on an open set, Ω, of $\mathbb{R}^2$ such that it satisfies a *Lipschitz condition* of the form,

$$|f(x, y_1) - f(x, y_2)| \leq K|y_1 - y_2|, \qquad [A2.2.1]$$

for all (x, y_1) and (x, y_2) in Ω. Then we shall prove the existence of a unique solution to the following:

$$\frac{dy}{dx} = f(x, y) \qquad [A2.2.2]$$

with initial condition,

$$y(x_0) = y_0. \qquad [A2.2.3]$$

Theorem. *Assume that $f(x, y)$ is a continuous bounded function in an open set, Ω, of $\mathbb{R}^2$ satisfying equation $[A2.2.1]$. Then there exists a unique solution to $[A2.2.2]$ and $[A2.2.3]$ on $[a, b] \in \Omega$ provided K in $[A2.2.1]$ satisfies $K < 1/\delta(b - a)$.*

We shall make this proof self contained by introducing the concept of continuous functions on a metric space. In other words, we give the condition that functions must satisfy in order for them to be members of $C[a, b]$.

Definition. *Let (S, d_S) and (T, d_T) be metric spaces and $f: S \longrightarrow T$ a function from S to T. Then f is said to be continuous at a point $s \in S$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $d_T(f(x), f(s)) < \varepsilon$ whenever $d_S(x, s) < \delta$.*

Proof of theorem. We first of all show that equations $[A2.2.2]$ and $[A2.2.3]$ are equivalent to an integral equation, which we use as an indicator to construct a mapping, J, from $C[a, b]$ onto itself. We will see that the solution to equations $[A2.2.2]$ and $[A2.2.3]$ corresponds to a fixed point of J. Since we know that $C[a, b]$ equipped with the metric defined in equation $[A2.1.1]$ is complete, the proof automatically follows from our fixed-point theorem if we can show that J is a contraction on $C[a, b]$.

So, we begin by integrating equation $[A2.2.2]$ between x and x_0. We have

$$\int_{x_0}^{x} \frac{dy}{dx} dx = \int_{x_0}^{x} f(t, y(t)) dt$$

$$\Longrightarrow \int_{y(x_0)}^{y(x)} dy = \int_{x_0}^{x} f(t, y(t)) dt$$

$$\Longrightarrow y(x) = y(x_0) + \int_{x_0}^{x} f(t, y(t))dt.$$

Now suppose that given $x, x_0 \in [a, b]$ and $\phi \in C[a, b]$, $J[\phi]$ is defined by

$$J[\phi](x) = y(x_0) + \int_{x_0}^{x} f(t, \phi(t))dt.$$

We show that $J[\phi]$ is also continuous: let θ and ϕ be continuous functions on $[a, b]$, then

$$\begin{aligned} d_S\left(J[\phi], J[\theta]\right) &= \max_{a \le x \le b} \left| \int_{x_0}^{x} f(t, \phi(t)) - f(t, \theta(t))dt \right| \\ &\le \max_{a \le x \le b} \int_{x_0}^{x} \left| f(t, \phi(t)) - f(t, \theta(t)) \right| dt \\ &\le \max_{a \le x \le b} K \int_{x_0}^{x} \left| \phi(t) - \theta(t) \right| dt \\ &\le K(b-a) \max_{a \le x \le b} \left| \phi(t) - \theta(t) \right|. \end{aligned}$$

So if $\varepsilon > K(b-a)\delta$ we have $d_S\left(J[\phi], J[\theta]\right) < \varepsilon$ whenever $d_S\left(\phi, \theta\right) < \delta$ and consequently $J[\phi]$ is continuous and therefore maps $C[a, b]$ onto itself.

Finally we show that J is a contraction. By the above argument,

$$d_S\left(J[\phi], J[\theta]\right) \le K(b-a) d_S\left(\phi, \theta\right) < d_S\left(\phi, \theta\right)$$

as required, which thus completes the proof.

This example is a nice way of introducing the reader to the method of using contraction mappings to prove existence of solutions to certain types of differential equation. We can see that the procedure essentially defines an iterative scheme in which the solution exists at the end of an infinite number of successive applications of J on an initial *guessed* solution, ϕ_0. That is to say, $y(x) = \lim_{n \to \infty} J^n[\phi_0](x)$ where $J^2[\phi_0] = J\left[J[\phi_0]\right]$, $J^3[\phi_0] = J\left[J\left[J[\phi_0]\right]\right]$ and so on. We measure the 'closeness' of the nth iteration at any stage using the metric function. Thus if $\phi_n = J^n[\phi_0]$ then

$$d_S\left(y, \phi_n\right) = d_S\left(J[y], J[\phi_{n-1}]\right) < d_S\left(y, \phi_{n-1}\right)$$

and the nth approximation is nearer to the true solution than the $(n-1)$th approximation.

The procedure that is developed in the main text, although is fixed upon a very specific class of differential equation, has many similarities with the above. For this reason we shall now highlight the three crucial steps in the above that appear in the development of an existence proof for the Newtonian equations of motion. The first step was simply to realise that the differential equation and the initial condition that the solution had to satisfy could be written in terms of an integral equation. This suggested a possible candidate for the contraction mapping. The second step, and this, in general, is usually the most difficult, was to define the metric space in which one expects to find the solution to this integral equation: in the above, we supposed that the solution lay in the set of continuous functions on $[a,b]$. These two stages together allowed us to define a map, J, from this metric space onto itself such that the solution corresponded to a fixed point of J. The third and final stage was to show that this fixed point is unique. For this we proved that J was a contraction on $C[a,b]$ and hence, by the fixed-point theorem, has a unique fixed point. Note that here we assumed that as a metric space, $C[a,b]$ is complete. The uniqueness of this fixed point then implies uniqueness of the solution to $[A2.2.2]$ and $[A2.2.3]$.

REFERENCES.

[YMS] Yodzis P, Mueller zum Hagen H, Seifert H-J 1973, 1974 *Commun. Math. Phys.* **34** 135, **37** 29.

[CO] Clarke C J S. and O'Donnell N 1992 Dynamical extension through a space-time singularity *Rend. Sem. Mat. Univers. Polotecn. Torino* **50** (1) 39–60.

[C] Clarke C J S 1993 The cosmic censorship hypothesis *The Renaissance of General Relativity and Cosmology* eds. G Ellis et al. (Cambridge: Cambridge University Press).

[SC] Swatton D J R and Clarke C J S 1998 The existence of Newtonian dust caustics *Class. Quantum Grav.* **15** 2891–2899.

[A] Arnol'd V I, Gusein-Zade S M, Varchenko A N 1985 Singularities of Differentiable Maps 1 (Boston: Birkhauser).

[R] Rendall A D 1992 On the choice of matter model in general relativity *Approaches to numerical relativity*, ed. R.A. d'Inverno (Cambridge: Cambridge University Press).

[L] Logan J D 1987 Applied mathematics: a contemporary approach (New York: Wiley).

[S] Synge J L 1960 Relativity: the general theory (Amsterdam: North–Holland Publishing company).

[N] Nakahara N 1990 Geometry, Topology and Physics (Bristol and New York: Adam Hilger).

[W] Wald R M 1984 General Relativity (Chicago: The University of Chicago Press).

[TA] Apostol T M 1974 Mathematical analysis 2nd ed. (Reading, MA: Addison–Wesley Publishing Company).