

City Research Online

City St George's, University of London

Citation: Osborne, A. D. (1979). Gravitation and dynamical systems: An investigation into the feasibility of modelling gravitation as a dynamical system. (Unpublished Doctoral thesis, The City University)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/37098/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

To [redacted], all of whom contributed to this thesis in their own special way.

GRAVITATION AND DYNAMICAL SYSTEMS

An investigation into the feasibility of modelling gravitation as a dynamical system

by

Anthony Derrick OSBORNE B.Sc. (City University)

A Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The Department of Mathematics
The City University, London, England.

October, 1979.

THE CITY UNIVERSITY LIBRARY,
ST. JOHN STREET, LONDON, E.C.1.

Contents

		Page
	Acknowledgements	4
	Declaration	5
	Abstract	6
	Chapter One:- Introduction	8
1.1)	Preliminary Comments	8
1.2)	Geometry: Truth or Convenience?	9
1.3)	Singularities: Fact or Fiction?	10
1.4)	Nonuniqueness in the Weak Field	13
1.5)	Alternative Strong-Field Models	13
1.6)	Spacetime Deformation via Dynamical Systems	15
	Chapter Two:- Gravitation and the Singularity Problem	19
2.1)	Gravitation and the Structure of Space and Time	19
2.2)	Differential Geometry	21
	2.2.1) Manifolds	21
	2.2.2) Vectors, One-Forms and Tensors	23
	2.2.3) Maps on Manifolds and Submanifolds	29
	2.2.4) Linear Connections, the Curvature Tensor and Geodesics	31
	2.2.5) Surfaces in a Manifold	36
	2.2.6) Variations of Curves and Conjugate Points	38
2.3)	The Local Structure of Space-time	44
2.4)	The Global Structure of Space-time: - Singularities	46
	2.4.1) Space-time Singularities	46
	2.4.2) The Schwarzchild Metric	47
	2.4.3) The Definition of Space-time Singularities	48
	2.4.4) The Singularity Theorems	49
2.5)	Resumé and Critique.	56
	Chapter Three:- Dynamical Systems Theory.	60
3.1)	Dynamical Systems	60
	3.1.1) Definition and Examples	60
	3.1.2) Systems of Autonomous Ordinary Differential Equations	62
3.2)	Vector Fields and Dynamical Systems	66
3.3)	Phase Portraits and Critical Elements	70
3.4)	Systems of non-autonomous Equations	73
3.5)	Properties of Orbits in the Neighbourhood of a Simple Critical Point	76
3.6)	Multiple Critical Points	85
	3.6.1) Multiple Critical Points of the First and Second Kind.	85
	3.6.2) Multiple Critical Points of the First Kind	86
	3.6.3) Multiple Critical Points of the Second Kind.	88
	3.6.4) Example.	91
3.7)	Limit and Minimal Sets:- Poincaré-Bendixson Theory.	92
3.8)	The Stability of Orbits of Vector Fields	94
3.9)	Structural Stability	102
3.10)	Bifurcation Theory	106
3.11)	Concluding Remarks.	113

Chapter Four:- The Singularity Theorems in the Language of the Qualitative Theory of Differential Equations 115

4.1) The Causal Structure of Space-time and Space-time Singularities. 115
4.2) Regular and Singular Conjugate Points. 119
4.3) The Singularity Theorems. 126

Chapter Five:- Models of Gravity Approximating to the Newtonian Equations 135

5.1) The One-Dimensional Newtonian Equation 135
5.2) Feasible Amendments of the One-Dimensional Newtonian Equation. 136
5.2.1) Equations of the Form $\ddot{r} = f(r)$ 136
5.2.2) Dynamical Amendments to the One-Dimensional Newtonian Equation of the type $\ddot{r} = f(r)$. 142
5.2.3) Equations of the Form $\ddot{r} = f(r) + \alpha g(r,\dot{r})$ 146
5.2.4) Dynamical Amendments to the One-Dimensional Newtonian Equation of the type $\ddot{r} = f(r) + \alpha r g(r)$ 154
5.3) Properties of the Amended One-Dimensional Newtonian Equations 154
5.3.1) Critical Points 154
5.3.2) Stability of Orbits 162
5.3.3) Closed Paths of $\ddot{r} = f(r) + \alpha g(r)\dot{r}$ 165
5.3.4) Structural Stability and Bifurcations 171
5.4) The Two-Dimensional Newtonian Equations 175
5.5) Feasible Amendments of the Two-Dimensional Newtonian Equations 176
5.5.1) Equations of the Form $\dot{r} = F(r)$, $\dot{\theta} = G(r)$ 177
5.5.2) Feasible Amendments of the Two-Dimensional Newtonian Equations of the Form (5.5 -1). 178
5.6) Properties of the Amended Two-Dimensional Newtonian Equations 183
5.6.1) Critical Points 184
5.6.2) Existence of Closed Paths 190
5.6.3) Structural Stability 198
5.7) Various Numerical Solutions of Specific Models. 199
5.8) Concluding Remarks. 214

Chapter Six:- Models of Gravity Approximating to the Schwarzchild Equations. 215

6.1) The Scwarzchild Equations of Motion 215
6.2) Feasible Amendments of the Schwarzchild Equations 219
6.2.1) A More General Set of Geodesic Equations 219
6.2.2) Feasible Amendments to the Schwarzchild Geodesic Equations of the Form (6.2 -4). 222
6.3) Properties of the Amended Schwarzchild Equations 225
6.3.1) Critical Points 227
6.3.2) Existence of Periodic and Almost Periodic Orbits 230
6.3.3) Structural Stability. 235
6.4) Properties of the Amended Metric 235
6.4.1) Generalized Kruskal-Szekeres Co-ordinates. 235
6.4.2) Corrected Weak and Strong Energy Conditions 240
6.5) Numerical Solutions of Specific Models 243
6.6) Concluding Remarks. 250

Chapter Seven:- Conclusions and Speculations. 251

References 254

Acknowledgements

I would like to thank all those people who made the production of this thesis possible. In particular, I am grateful to Dr. Allan E.G. Stuart for his many helpful comments and suggestions during the course of the three and a half years it has taken to produce. I also thank The City University library and computer centre, without whose services I would have been lost. Gratitude goes to the Science Research Council, whose grant supported this research and a special thanks goes to [redacted] for doing the typing.

Declaration

I hereby declare that for a period of one year from the submission of this thesis, it should not be copied in whole or in part.

At the end of the one year period, I grant powers of discretion to the University Librarian to allow the thesis to be copied in whole or in part without further reference to the author.



A. D. Osborne

Abstract

Theoretical objections to the concept of a spacetime singularity and the scanty, and sometimes controversial, nature of the available experimental data concerning the existence of black holes, has heightened interest in the search for viable, non-singular alternatives to Einstein's theory of gravitation. One way of developing such alternatives is to perturb or deform the solution spacetimes of the Einstein model in such a manner that essential singularities are either eliminated or avoided, but useful features, such as the weak-field predictions, are retained. The aim of the present work is to study the feasibility of such a programme, and to analyse and classify some of the possible perturbations.

Our approach to the problem is novel for this field of work, the objective being to associate dynamical systems with the flows of the geodesic equations. Since dynamical systems are defined for all values of the flow parameter, the proper-time in our case, the resulting geodesic flows will be complete. This means that any singularities on the original manifolds will either be changed into regular or critical points, or isolated by critical sets in which case they are effectively disconnected from the spacetimes of all observers.

The thesis commences with a general introduction in which we discuss the pros and cons of the singularity argument and motivate our approach. In chapter 2 we review the Einstein model, paying particular attention to the definitions and proofs of the singularity theorems, while chapter 3 is devoted to a presentation of those aspects of dynamical systems theory which are needed for our purposes. However, since the latter is not a subject which is usually associated with conventional gravitation theory, we have attempted to make this presentation more or less self-contained.

The next three chapters contain our new results. Thus, in chapter 4 we 'translate' the singularity theorems into the language of dynamical systems theory and examine their possible breakdown. Then, in chapters 5 and 6, we construct and analyse several amended versions of Newtonian and Schwarzschild gravitation respectively; the Newtonian amendments being necessary to preserve the use of the Newtonian limit as a boundary condition. Finally, in chapter 7, we discuss the physical interpretation of our results and assess the viability of the method used.

Chapter One: Introduction

1.1) Preliminary comments

Some of the more important and interesting problems of classical gravitation today are those concerning the singularities and other pathological features of Einstein's theory. Because of the very restricted nature of the experimental data, the acceptability or unacceptability of the Einstein model depends, to a large extent on whether its theoretical structure and predictions seem reasonable or not.

It must be admitted, that of the many metric theories that agree with the experimental data in the immediate post-Newtonian domain, the Einstein model is the most compelling. If we accept the basic philosophy that general relativity is a geometrical symmetry of nature and hence, via the equivalence principle, that gravitation is a manifestation of a curved spacetime[1,2,3], then the vacuum field equations $R_{\mu\nu} = 0$ are the simplest nontrivial structure that one can impose on a psuedo-Riemannian manifold. In addition, if we restrict ourselves once again to these vacuum equations, the Einstein model is "pure". There are no free parameters or potentials to specify and, in this sense, it shares with Maxwell's free-field equations the distinction of being one of the only two strictly nonphenomenological theories of physics that we have today.

Now the interpretation of gravitational dynamics in terms of Riemannian geometry together with the simplicity of Einstein's equations, does lead to an elegant and, some would say, beautiful theory. But physical theories are not judged by beauty alone and work done over the last three decades has shown that the very features which are responsible for the beauty and elegance of the Einstein model, may also be responsible for significant "deficiencies" in its theoretical and predictive structure. Thus, in the first place, it is an isolated theory. All attempts at unifying the principles of general relativity with those of classical electromagnetism or quantum theory have resulted in formulations which are either ambiguous, highly contrived or both[4,5]. Secondly, it is an essentially singular description and

leads inevitably to gravitational collapse with all its attendant pathologies. For example, the bizarre extrapolations of the consequences of the singularity theorems are in sharp contrast to the physically reasonable assumptions needed to prove them[6,7]. Let us examine these deficiencies of the Einstein theory in more detail.

1.2) Geometry: Truth or Convenience?

The isolation of gravitation is due to its interpretation, by Einstein and his followers, as being entirely geometric in origin, i.e., a consequence of the curvature of spacetime. Thus, the Newtonian concept of motion under the "force" of gravity is replaced by "free" motion along the geodesics of a curved, four-dimensional, differentiable manifold which has the symmetry of general relativity and is locally Minkowskian. There are two points to notice about this description. Firstly, a particular geometry has been elevated to the status of the real or true geometry of spacetime. Thus, this geometry is no longer a convenience, but a necessity. Secondly the covariance group of general relativity, even in specially selected asymptotic situations[8,9], is very much larger than the symmetry groups which arise in other areas of physics. As a result, it appears to be impossible to define an unambiguous, irreducible and nonredundant set of canonical observables, and this, of course, has serious consequences in any quantum treatment of the gravitational field and its interactions with matter[10].

Now the debate on the geometrisation of physics goes back to Riemann and Clifford in the last century[11], who began to suspect that at least some branches of physics could be expressed in terms of the non-Euclidean geometrics that were a matter of current interest at the time. It is, more or less, axiomatized developments of these ideas that have found their expression in general relativity, the various unified theories of electromagnetism and gravitation and the already unified theory of Raivich and Misner[12]. According to this view, there is one geometry of spacetime such that matter moves along geodesics in this geometry. The job of the physicist is to compute this geometry in terms of the sources such as mass and charge etc., and general relativity tells us how to do this. Once we know this, we know all there is to know about the universe. Einstein himself believed that there exists a generalisation of his theory which will be sufficient for a description of all physical phenomena,

including quantum phenomena, and this philosophy, followed with a religious fervour by the geometrodynamicists, is stated most clearly by John Wheeler, prophet of geometrodynamics, who says[13]: "There is nothing in the world except empty curved space. Matter, charge, electromagnetism and other fields are only manifestations of the curvature of space".

Fortunately, for those with more secular inclinations, there is an opposing point of view which was developed by the conventionalist philosophers of science of whom Henri Poincaré was the leading advocate. Their point of view is that geometry is a product of mathematics, there is a wide variety of geometrics that the mathematician has invented and the physicist can choose any of these geometrics against which to describe the behaviour of matter. Furthermore, the physicist always deals with a combination of geometry and matter, never with the two separately, and he may choose to express a theory in one geometry or another just as he can choose to use English, French or Chinese with which to express his ideas. It may, on occasions, be easier to express a theory in a particular geometry, but this does not make this geometry the "real" geometry. This philosophy is summed up in Poincaré's own words[14]: "One geometry cannot be more true than another; it can only be more convenient".

The conclusions to be drawn from the above discussion can be summed up as follows. At the classical level, if the Einstein model is an accurate description of gravitation it does not really matter whether it is expressed in the geometrodynamicists' language or not. However, at the quantum level it may be necessary to recognise the possible separation of geometry and gravitational fields before further progress can be made towards producing a quantised theory of gravitation.

1.3) Singularities: Fact or Fiction?

Since the original work of Penrose and Hawking on the existence of geometrical singularities in the time development of physically reasonable solution spacetimes of the Einstein model, much attention has been paid to these specific predictions of the theory. For example, the problems of the definition and structure of the singularities and further theorems

on their existence have been the subjects of an enormous number of research papers over the last decade or so.

Singularities are predicted to occur in two areas. The first is in the past at the beginning of the present expansion of the universe. This is thought to be the "big bang" and is generally regarded as the origin of the universe. The second area in which singularities are predicted is in the collapse of isolated regions of high mass concentration such as burnt-out stars. Now these predictions follow from an extrapolation of the Einstein model by many orders of magnitude beyond the range for which its validity has been established, i.e., they are predictions in the domain of very high curvatures and very strong fields, and, as a consequence, there are two opposing views regarding their relevance. On the one hand, the geometrodynamicists, and other firm believers in Einstein's theory, regard these singularities as useful, valid and physically meaningful predictions of relativistic cosmology, i.e., the so-called "black holes". On the other hand, the conventionalists take the point of view that singularities in any theory are, to say the least, physically and mathematically ambiguous "objects", and that their occurence is an indication of the limitations of the theory. If this is the case, then it follows that the Einstein model is incorrect or, at best, approximate in regions of high curvature and strong fields.

Now which of these opinions turns out to be the correct one should, of course, be decided by experiment, but one must distinguish between consistency and uniqueness in assessing experimental verification. In the case of the big bang, the Hubble expansion and the isotropic $3^{o}K$ radiation are consistent with an initial singularity, but the experiment is not repeatable and to compute a unique path back to a singular origin, without any other "fossil" guide seems to be more in the realm of a theological speculation rather than a scientific explanation. As far as the black holes are concerned, the geometrodynamicists "holy-grail" lies in Cygnus X-1. Preliminary x-ray data indicate that there is a massive object in this binary system which they claim to be a black hole. However, the evidence is scanty and controversial[15] and, besides, there are other possible explanations

of the data which do not involve a black hole[16]. Another feature which complicates the interpretation of this, or any other data, and indicates the dangers inherent in treating singularities as physically meaningful objects, is that the theoretical picture of the black hole keeps changing. Having started out life as evergrowing, cosmic vacuum cleaners, it now appears that they can merely evaporate away[7].

Quite apart from the experimental status of black holes, there also exists the theoretical problem of defining precisely what is meant by a "singular spacetime", when the whole spacetime manifold is not specified a priori at present singularities are associated with incomplete null and timelike geodesics. Thus, a spacetime is said to be singular if it contains inextendible null and (or) timelike geodesic segments with finite affine parameter. (Note that completeness has to be defined via the tangent bundle rather than on the manifold itself, since the metric on the latter is not positive-definite[17,18].) However, incompleteness does not necessarily involve an infinity of a physical variable nor does it necessarily prevent the extension of a cosmological model towards the past. For example, Taub-NUT spacetime has incomplete geodesics in a region which is locally nonsingular[19,20] and if a cosmological model has an infinite density in the past, this infinity (which undoubtedly represents a singularity) and the incompleteness of geodesics need not have any relation[21]. Thus it seems that completeness or incompleteness of a spacetime is not the whole answer and the search for a more satisfactory definition of singularities continues.

The conclusions to be drawn from the above arguments are threefold. Firstly, singularities, in the accepted technical sense, seem to be a common feature of a large number of otherwise physically reasonable solutions of the Einstein equations. Secondly, the definition of a singular spacetime is far from being physically intuitive and does not always correspond to infinities in physical variables. Thirdly, the interpretation of a mathematical singularity as a physical object leads to an "anything-goes" situation in which even the most fanciful and bizarre speculations are permitted. Thus, it seems very probable that the Einstein model is not a useful description of gravitation at the

strong-field end of the spectrum.

1.4) Nonuniqueness in the Weak Field

The flaws in the Einstein model, discussed above, lead us to take a closer look at its experimental successes. The first conclusion we come to is that the three "classical" measurements of red shift, light deflection and perihelion advance together with the more recently proposed radar reflection experiment, merely test the weak field behaviour of a diagonal Ricci tensor. In other words, if we assume that the spacetime of the solar system is static, spherically symmetric and asymptotically flat, then its Ricci tensor, $R_{\mu\nu}$, is diagonal and obviously goes to zero in the limit of zero field. The experiments then tell us that these diagonal components are still zero in the limit of weak fields. The vacuum equations of the Einstein model, i.e., $R_{\mu\nu}=0$, satisfy this condition and, hence, are consistent with the observational data. However, they are not unique in this respect. For, if we probe further and assume that the diagonal components of $R_{\mu\nu}$ can be represented by power series expansions in m/r, then the only restrictions that the experiments place on these expansions are that R_{00}, R_{22} and R_{33} have to be at most of $O(m^3/r^3)$, while R_{11} can be of $O(m^2/r^2)$. One can take advantage of this fact to construct other models of gravitation which also satisfy the weak field tests and, indeed, many authors have done so. For example, the scalar-tensor theories of Brans and Dicke[22], of Yilmaz[23] and of Tupper[24] fall into this category.

No doubt, as techniques improve and experiments become more precise, tighter limits will be placed on the diagonal components of the Ricci tensor. However, it is unlikely that any degree of accuracy in weak-field measurements will ever eliminate all but one theory of gravitation. Thus, the superiority of the Einstein model in the weak-field domain is based more on its simplicity and elegance rather than on its agreement with the experimental data.

1.5) Alternative Strong-Field Models

The singularity theorems and the chronic problems of quantizing a

noncanonical theory, indicate that the weaknesses of Einstein's model lie at the strong-field rather than the weak-field end of the gravitational spectrum. Thus, there is a need for alternative strong-field models and any indication in this direction is "a priori" interesting. The hope is that such models, if they can be devised, would be free from singularities, except in obviously artificial or pathological situation, and at the same time be sufficiently flexible so that they can be "adjusted" to fit the experimental data at the weak-field end. The nonuniqueness of the Einstein model in this latter domain gives us confidence that such an adjustment should be possible.

If we accept the need for alternative strong-field models, then the question arises as to which is the best possible approach to the construction of such models and here, of course, opinion divides. Many authors[25] believe that singularities and quantization are closely related and that a strong-field model must inevitably be a quantum model. They[5,25] generally argue that a quantized version of the Einstein theory would lead to a better understanding of the singularities and perhaps even prevent them from occurring; much in the same way that Planck's quantization of blackbody radiation eliminated the Rayleigh-Jeans catastrophe and Bohr's quantization of the atom prevented its classically predicted collapse. However, the problem of developing a quantum formalism for Einstein gravity has been under study for at least the last twenty-five years and, as yet, there exists no convincing mathematical evidence either to support or deny these speculations. The many attempts at constructing quantum models within the Einstein framework, have generally led to nonrenormalizable theories[25,26] which, as a consequence, have given little, if any, insight into the nature and significance of singular spacetimes. Since the outlook for further progress in this direction is not very encouraging, we have no other option, but to return to the classical domain.

Now within the classical domain, and if we restrict outselves to essentially metric theories, there are three possible avenues of approach. The first is to develop alternative field equations using curvature alone. For example, models which make use of the derivatives of the Ricci tensor, rather than the Ricci tensor itself, would presumably

give more information about the high curvature or strong-field regime. Thus, we could consider free-field equations of the form

$$R_{\mu\gamma;\varepsilon} = R_{\mu\varepsilon;\gamma} \quad ,$$

which were first studied by Stephenson[27] and by Kilmister[28], and more recently revived by Yang[29]. These equations are a generalisation, to third differential order, of Einstein's free-field equation $R_{\mu\nu} = 0$ and have a large variety of interesting solution spacetimes. These include both "dressed" and "undressed" Schwarzschild solutions[30] as well as spacetimes in which gravity is repulsive and, in some cases, periodic[31]. Another class of models which falls into this first category and which we might mention, are those based on nonlinear Lagrangians in which the curvature scalar has a finite upper bound. The hope is that by limiting the curvature one might avoid gravitational collapse and a specific example in this class has been studied by Muller et al.[32]

The second approach is finding nonsingular classical models is to introduce additional geometrical features into the theory. For example, Trautman[33] has proposed that spin and torsion may avert gravitational collapse when treated in the framework of Einstein-Cartan theory. This seems a promising line of attack and some interesting results have been obtained by Prasanna[34].

The third approach is perhaps the least ambitious of the three, but, in some ways, is the best defined. The idea is to consider specific singular spacetimes of the Einstein model and to perturb or deform them in such a manner, that essential singularities and other pathologies are either eliminated or avoided, but useful features, such as weak-field predictions, are retained. In this thesis we study the feasibility of such a programme by analysing and classifying some of the possible deformations.

1.6) Spacetime Deformation via Dynamical Systems

As we mentioned earlier, singularities of the Einstein model are predicted to occur in two areas. The first is in the past of cosmological solutions which represent spacetimes of an expanding

universe, while the second is in more specific situations representing the collapse of localized concentrations of large mass. Now cosmological models have additional assumptions built into them and are not suitable candidates for a preliminary investigation into the possibility of eliminating singularities. Thus, in the present work, we concentrate on the second area, i.e., that of gravitational collapse. In this case, the final or black hole states are represented by the Schwarzschild and Kerr solutions[6] (or their charged counterparts), and to these must be added the Newtonian model, since the latter is an essential boundary condition for the former. Furthermore, since the Kerr solution can be obtained from the Schwarzschild solution by a transformation involving the replacement of the real mass m by the complex quantity m + ia, where a is the angular momentum[6], it is reasonable to suppose that the same thing can be done with an amended Schwarzschild spacetime. Thus, we restrict our analysis to the Schwarzschild and Newtonian solutions.

Our approach to the problem is novel for this field of work, the objective being to associate dynamical systems with the flows of the geodesic equations. Since dynamical systems are defined for all values of the flow parameter, the proper-time in our case, the resulting geodesic flows will be complete. This means that any singularities on the original manifolds will either be changed into regular or critical points, or isolated by critical sets in which case they are effectively disconnected from the spacetimes of all observers. As a point of interest, we might add that dynamical systems, or the qualitative theory of differential equations, is a geometric approach to the study of differential equations on a manifold initiated by Poincaré. Thus our work is also an attempt to bring together the ideas of Einstein and Poincaré.

The main body of the thesis consists of five chapters. In chapter 2 we give a survey of gravitation and the structure of spacetime with particular reference to Einstein's theory. The need for looking at the large scale structure of the universe as opposed to its local properties is brought out and the reason for assuming spacetime to be a differentiable manifold is also discussed. We then turn our attention to the problems of singularities in this structure, i.e., the problem of their definition

and their occurrence. The possible breakdown of the present singularity theorems is then discussed. Lastly, we present a case for looking at the equations of motion from the point of view of dynamical systems.

Chapter 3 is devoted to a rigorous presentation of those aspects of dynamical systems theory which are needed for our purposes. However, since this subject is not usually associated with conventional gravitation theory, we have attempted to make this presentation more or less self-contained. Thus, we give examples of the abstract concept of a dynamical system before showing, again with examples, how vector fields acting on manifolds can give rise to dynamical systems. The idea of the phase portrait is then introduced together with relevant definitions, theorems and further examples. The unimportant notions of stability of orbits and of structural stability are also discussed.

The next three chapters contain our new results. Thus, in chapter 4 we 'translate' the singularity theorems into the language of dynamical systems theory. The aims of this exercise are, firstly to explore the possibility of using dynamical systems theory to give more transparent proofs of the singularity theorems and, secondly, to identify the 'simplest' amendments which would lead to a breakdown of these theorems. Then, in chapters 5 and 6, we use some of these ideas to construct and analyse several amended versions of Newtonian and Schwarzschild gravitation respectively. The theoretical analyses in these chapters are supported, in many cases, by detailed computer calculations of both radial and nonradial orbits to ensure that the weak-field conditions can be met. However, for reasons of space we have not included lengthy programme descriptions.

To summarise our main results, we find that singularities can be avoided in three general situations. The first is if the gravitational force remains attractive, but becomes weaker at small distances leading to an "asymptotically free" theory. The second is if the effective gravitational potential develops a repulsive core implying that gravitation changes from being attractive to being repulsive at small distances. While the third is if there is a dissipative mechanism

which slows down the collapse to such an extent that it takes an infinite time to become singular. A discussion of the physical interpretation and implications of these results, together with an assessment of the inability of the method and suggestions for improvements and further research, is given in the final chapter, i.e., chapter 7.

Chapter Two:
Gravitation and the Singularity Problem

In this chapter, a review of Einstein's geometric theory of gravity is presented. The theory divides basically into two parts, that of local structure of the universe and that of global structure. After the mathematical preliminaries needed in such a theory are given, the local theory is discussed. In the next section, a review of space-time singularities and the part they play in the large scale structure of the universe is given and the chapter continues by investigating the singularity theorems, which predict the occurrence of singularities based on fairly general assumptions.

In the last section, the difficulties involved in defining singularities is made explicit and the possible breakdowns of Einstein's theory and the global techniques which predict the occurrence of singularities are looked into.

2.1). Gravitation and the Structure of Space and Time

Gravity plays the dominant role in the shaping of the structure of the universe, although it is the weakest of the four fundamental interactions (the other three being the strong and weak nuclear interactions and electro magnetism). This is because gravity has a long range and is a force which affects every particle in the same way.

Einstein's theory of General Relativity is concerned with the structure of space and time from the geometrical point of view and what effects gravity has on this geometry. Newton based his theory of gravitation on the existence of absolute, uniform, space and time, both of which could not be affected by anything external. The Einstein theory of Relativity, in contrast to this, assumes that space and time are inseparable from each other and divides into the theory of uniform, pseudo-Euclidean space-time which possesses the Minkowski metric[1] and the theory of non-uniform pseudo-Riemannian space.

The uniformity of pseudo-Euclidean space-time gives rise to the existence of a group of transformations, the most general being the Lorentz transformation, which leaves invariant the four-dimensional metric. The theory of Special Relativity is based on these transformations.[2]

It is now known that universal gravitation does not fit into the framework of pseudo-Euclidean space. Attempts to modify Newton's equations of gravity to four-dimensional geometric form within the framework of Special Relativity lead to serious shortcomings - all fail to agree with present experimental evidence, or the modifications are inconsistent. Straightforward steps to repair any inconsistency lead ultimately to General Relativity theory and to the existence of a pseudo-Riemannian (Lorentz) space-time. In fact, Schild[3,4,5] has shown that the existence of the gravitational redshift proves that a consistent theory of gravity cannot be constructed within the framework of Special Relativity.

Einstein found it possible to base a theory of universal gravitation on a non-uniform space-time which is locally pseudo-Euclidean, that is, the theory assumes that space-time has the structure of a manifold, which is curved in some sense. It is assumed that the curvature of space-time interacts with the various matter fields which describe the matter content of space-time, so that the existence of matter produces the curvature and, in turn, the curvature influences the motion of the matter. Thus, gravity is thought to be a manifestation of the curvature.

Until the early 1960's interest in gravitation theory centred mainly on local techniques,that is, the study of local laws satisfied by the various physical fields, expressed in the form of differential equations. In order to build up an understanding of the global structure of the universe, local computations were performed near each event and these local results were patched together. It was realized, however, that in order to apply any

theory of gravitation to physical problems, the large scale structure of the universe had to be looked into. It was the proof of Penrose's first theorem on the occurrence of singularities, based on the Einstein theory, that gave rise to global techniques for studying space-time.

Before the mathematical formulation of the local and global structure of the universe can be given, some preliminaries are needed and these are given in the next section.

2.2). Differential Geometry

In this section, the areas of differential geometry needed for General Relativity theory are presented. The definition of a manifold is first given, followed by such notions as the metric, vectors and tensors and curvature.

2.2.1). Manifolds

A manifold is essentially a space which is locally similar to Euclidean space. The structure permits differentiation to be defined, but does not distinguish between different coordinate systems.

Definition 2.2-1.

Let M be a set. A local n-dimensional chart on M is a bijection ψ from a subset $\mathcal{U}$ of M to an open subset of $\mathbb{R}^n$. ψ is usually denoted by $(\mathcal{U},\psi)$ to indicate the domain $\mathcal{U}$ of ψ.

A C^r n-dimensional atlas on M is a family of n-dimensional charts $\{(\mathcal{U}_i,\psi_i)\}$ such that,

(i) $M = \bigcup_i \{\mathcal{U}_i\}$

(ii) For two charts $(\mathcal{U}_i,\psi_i)$ and $(\mathcal{U}_j,\psi_j)$ with $\mathcal{U}_i \cap \mathcal{U}_j \neq \emptyset$, $\psi_j \circ \psi_i^{-1} / \psi_i(\mathcal{U}_i \cap \mathcal{U}_j)$ is a C^r map on $\mathbb{R}^n$.

A set M with a given C^r atlas of dimension n is known as a C^r n-dimensional differentiable manifold.

The concept of an atlas is needed as it is usually not possible to construct a suitable chart for a set M whose domain is M. It is assumed in the definition that the C^r atlas is complete, in the sense that it is not contained in any other C^r atlas of M. Notice that a set M may admit more than one C^r atlas into $\mathbb{R}^n$.

Definition 2.2-2

Let x be a chart of a differentiable manifold M, of dimension n. Such a chart defines on its domain $\mathcal{U}$ a set of coordinate functions x_i such that $x \equiv (x_1,\ldots,x_n)$. If $m \in \mathcal{U}$, $x(m) = (x_1(m),\ldots,x_n(m)) \in \mathbb{R}^n$, is the set of coordinates of m in the chart x.

$\mathcal{U}$ is a coordinate domain of M and also a coordinate neighbourhood of $m \in M$.

Thus, choosing a new chart (of an equivalent atlas) in a neighbourhood of m corresponds to the classical technique of transforming coordinates. The coordinate neighbourhoods on M define a topology and so any differentiable manifold is a topological space. Thus, all the usual concepts on topological spaces hold for differentiable manifolds.

This work is concerned with differentiable manifolds with reasonable local behaviour and by a 'manifold' will usually be meant a Hausdorff, second countable, C^∞ differentiable manifold.

Examples of manifolds abound, such as the set of all real r x s matrices, the n-dimensional sphere S^n, the Möbius strip and the figure of eight.[6] $\mathbb{R}^n$ is trivially an n-dimensional manifold with the identity chart.

The manifold model for space-time possesses a few extra

characteristics, one of which is orientability:-

Definition 2.2-3

A manifold is said to be orientable if it possesses an atlas such that in every non-empty intersection of coordinate domains $\mathcal{U}$, V, the Jacobian $\left| \frac{\partial x_i}{\partial y_j} \right|_m > 0$, where $(x_1,\ldots,x_n)$ and $(y_1,\ldots,y_n)$ are coordinate functions in $\mathcal{U}$,V respectively.

The Möbius strip is an example of a non-orientable manifold.[6]

2.2.2). Vectors, One-Forms and Tensors

A generalization of partial differentiation on Euclidean space is introduced. This leads to the concept of a tangent vector and tangent space, which has the structure of a vector space, at each point of the manifold. One-forms and tensors are defined in terms of maps acting on these vectors. The metric defined on a manifold is a special type of tensor.

Definition 2.2-4

Let ψ be a chart of a manifold M with domain $\mathcal{U}$ and let ψ' be a chart with domain V of a manifold N. Let f: $M \to N$. Let m be a point in the domain of f and assume $m \in \mathcal{U}$, $f(m) \in V$. Then $F = \psi' \circ f \circ \psi^{-1}$: $\mathbb{R}^n \to \mathbb{R}^n$ is a coordinate representative for f.

f: $M \to N$ is differentiable at m if F is of class C^∞ at $\psi(m)$. f is differentiable if it is differentiable at every point of its domain. If, in addition, f is an injection and f^{-1} is also differentiable, f is a diffeomorphism.

It can be shown that the definitions are independent of the choice of coordinate representative. Relative to a chart ψ with domain $\mathcal{U}$ on a manifold M and the identity chart on $\mathbb{R}$, a differentiable function f: $M \to \mathbb{R}$ with domain V has coordinate representative F so

that on $U \cap V$, $f = F \circ \psi$. Then $\frac{\partial f}{\partial \psi_i}$ may be defined as $F_{,i} \circ \psi : M \to \mathbb{R}$. These are differentiable functions with domain $U \cap V$ and clearly, in general, behave just like partial derivatives.

Definition 2.2-5

If $\mathcal{M}$ is a manifold, the set of differentiable functions $f: M \to \mathbb{R}$ whose domain includes a given point $m \in M$, will be denoted by $\mathcal{F}(m)$. A derivation on $\mathcal{F}(m)$ is a function $\Lambda: \mathcal{F}(m) \to \mathbb{R}$ such that,

$$\Lambda(fg) = (f(m))(\Lambda g) + (\Lambda f)(g(m)); \quad f,g \in \mathcal{F}(m).$$

It is fairly easy to show that the set of all derivations on $\mathcal{F}(m)$ has the structure of a real vector space which is called the tangent space at m and is denoted by $T_m M$. An element of $T_m M$ is a tangent vector at m.

Proposition 2.2-1[7]

If ψ is any chart of M whose domain includes a given point m, the vectors $\left(\frac{\partial}{\partial \psi_i}\right)_m$ $(i = 1,\ldots,n)$ form a basis for $T_m M$, called the canonical basis (coordinate basis) for $T_m M$ associated with ψ. □

The union of all tangent spaces $T_m M$ as m varies over $\mathcal{M}$ can be shown to have the structure of a vector bundle and of a 2n-dimensional manifold TM, called the tangent bundle.[8]

Having defined vectors at a point $m \in M$, one-forms and tensors may be defined in terms of these vectors.

Definition 2.2-6

A one-form ω at $m \in M$ is a real valued linear function on the space $T_m M$. $\omega : T_m M \to \mathbb{R}$ is given by $\omega(v) = \langle \omega, v \rangle \in R$ for each $v \in T_m M$.

It can be shown that the space of all one-forms at $m \in M$ is an n-dimensional vector space, called the dual space T_m^*M of T_mM.

Definition 2.2-7

$f: M \to \mathbb{R}$ defines a one-form df at $m \in M$, by the rule $\langle df, v\rangle = v(f)$ for each vector $v \in T_mM$. df is called the differential of f.

If $\{E_i\}$ is a basis for T_mM, the components of $v \in T_mM$ with respect to this basis, denoted by v^i are given by $v = v^iE_i$. A unique basis $\{E^i\}$ for the set of one-forms may be defined by $\langle E^i, Ej\rangle = \delta^ij$, where δ^ij is the Kronecker-delta. Then if $\omega \in T_m M$, $\omega = \omega_iE^i$, and for any $v \in T_mM$, $\langle\omega,v\rangle = \omega_iv^i$. In particular, if $\frac{\partial}{\partial x_i}$ is the coordinate basis for T_mM associated with a chart x, the set of differentials $(dx_1,\ldots,dx_n)$ at m from a basis for T_m^*M. In terms of this basis, the differential of $f: M \to \mathbb{R}$ is given by $df = (\frac{\partial f}{\partial x_i})dx_i$ at each point of the coordinate domain, which is just the classical 'total differential' of a function of n variables. If $df \neq 0$, the surfaces $\{f = \text{constant}\}$ in M are (n-1)-dimensional manifolds and df can be thought of as a normal to the surface $\{f = \text{constant}\}$. These concepts will be used at a later stage, when hypersurfaces are defined.

Definition 2.2-8

From the spaces T_mM and T_m^*M can be formed the cartesian product,

$$C^p_q{}_m M = \underbrace{T_m^*M \times\ldots\times T_m^*M}_{p \text{ factors}} \times \underbrace{T_mM \times\ldots\times T_mM}_{q \text{ factors}},$$ that is, the

ordered set of one-forms and vectors $(\omega^1,\ldots,\omega^p,v_1,\ldots,v_q)$ at m.

A tensor of type (p,q) at $m \in M$ is a multi-linear function $C^p_q{}_m M$, that is, a function which is linear in each argument. The space of all such tensors is denoted by $T^p_q{}_m M$. Thus, $T^1_o{}_m \equiv T_mM$

and $T^{o}_{1\ m} \equiv T^{*}_{m} M$.

Addition of tensors and multiplication of a tensor by scalars are defined as for arbitrary multilinear functions. With these rules, the space $T^{p}_{q\ m} M$ is a vector space of dimension n^{p+q} over $\mathbb{R}$.

If $t \in T^{p}_{q\ m} M$, the components of t with respect to the dual bases $\{E_i\}$ and $\{E^i\}$ for $T_m M$, $T^{*}_{m} M$ respectively are given by,

$$t^{i_1 \ldots i_p}_{j_1 \ldots j_q} = t(E^{i_1}, \ldots, E^{i_p}, E_{j_1}, \ldots, E_{j_q}) \text{ where each } i_a \text{ and } j_b$$

run from 1 to n. For simplicity, tensor relations are usually given in terms of the components with respect to some chosen basis.

Definition 2.2-9

The contraction of a tensor t of type (p,q), with components with respect to some given basis $t^{ab\ldots d}_{ef\ldots g}$, on the a and e indices, is defined to be the tensor of type (p-1, q-1) whose components with respect to the same basis are $t^{ab\ldots d}_{af\ldots g}$.

Similarly, a tensor can be contracted over any pair of upper and lower indices and it can be shown that the result is independent of the basis used, in its definition.[9] (If the tensor is contracted over two upper or two lower indices, the resultant tensor would depend on the basis used).

Definition 2.2-10

The components of the symmetric part of a tensor on the set of lower indices are denoted and defined by,

$$t^{b_1 \ldots b_q}_{(a_1 \ldots a_p)} = \frac{1}{p!} \left\{ \text{sum over all permutations of the indices } a_1, \ldots, a_p \left(t^{b_1 \ldots b_q}_{a_1 \ldots a_p} \right) \right\}.$$

The components of the antisymmetric part of a tensor on the set of lower indices are denoted and defined by,

$$t_{[a_1 \ldots a_p]}^{b_1 \ldots b_q} = \frac{1}{p!} \left\{ \begin{array}{l} \text{alternating sum over all permutations of} \\ \text{the indices } a_1, \ldots, a_p \left(t_{a_1 \ldots a_p}^{b_1 \ldots b_q} \right) \end{array} \right\}$$

There are exactly similar definitions for the upper indices. A tensor is symmetric in a given set of indices if it is equal to its symmetrized part on these indices and is antisymmetric if it is equal to its antisymmetrized part.

There are several tensors having symmetric properties, which play an important role in determining the properties of a manifold. The first of these is the metric.

Definition 2.2-11

A metric tensor at $m \in M$, denoted by g_m, is a symmetric tensor of type (0,2) at m. A metric, g, on M is a symmetric field which assigns to each point $m \in M$ a metric tensor.

At any point $m \in M$, $g_m(u,v) = g_m(E_i, E_j)\, u^i v^j = g_{ij} u^i v^j$, where $u, v \in T_m M$ and $\{E_i\}$ is a given basic for $T_m M$. u and v are orthogonal if $g_m(u,v) = 0$. g is said to be non-degenerate at m if there is no non-zero vector $u \in T_m M$ such that $g_m(u,v) = 0$, for all vectors $v \in T_m M$. Thus, if a metric is degerate, u is orthogonal to every vector.

It can be seen that degeneracy is not a desirable property and it will be assumed from now on that the given metric tensor is non-degenerate. Then the matrix of components $[g_{ij}]$ is non-singular at every point $m \in M$. In this case, there exists a unique symmetric tensor of type (2,0) with components g^{ij} at each point, where $g^{ij} g_{jk} = \delta^i{}_k$, so that the matrix $[g^{ij}]$ of components is the inverse of the matrix $[g_{ij}]$. The tensor g^{ij} and g_{ij} can be used to give an isomorphism between any upper tensor argument and any lower one, that is, they can be used to "raise and lower indices":-

$$t^{b\ldots}_{acd\ldots} = g_{ck}\, t^{bk\ldots}_{ad\ldots} \quad ; \quad t^{bc\ldots}_{ad\ldots} = g^{ck}\, t^{b\ldots}_{akd\ldots} \qquad (2.2\text{-}1)$$

Definition 2.2-12

The signature of g at $m \in M$ is the number of positive eigenvalues of the matrix $\left[g_{ij} \right]$ at m, minus the number of negative ones. A metric whose signature is n at each point is positive definite. A metric whose signature is (n - 2) is called a Lorentz metric.

If g is non-degenerate and continuous, the signature is constant on M. By choosing an orthonormal basis for $T_m M$, the metric components at m can be brought into the form,

$$g_{ab} = \text{diag} \left\{ \underbrace{+1, +1, \ldots, +1}_{\frac{1}{2}(n+s) \text{ terms}}, \underbrace{-1, \ldots, -1}_{\frac{1}{2}(n-s) \text{ terms}} \right\},$$

where s is the signature of g.

If g is positive definite, $g_m(v,v) \geq 0$ with $g_m(v,v) = 0$ only if $v = 0$, where $v \in T_m M$. If M possesses a positive definite metric, it possesses a metric in the topological sense[10]. This is not true if M possesses a Lorentz metric.

Definition 2.2-13

Let g be a Lorentz metric. A vector $v \in T_m M$ is called timelike, null, or spacelike, according to whether $g_m(v,v) < 0$, $g_m(v,v) = 0$ or $g_m(v,v) > 0$ respectively.

It can be shown that any non-compact manifold admits a Lorentz metric.[11] This is not always true of compact manifolds. For example, the sphere, S^2, does not admit a Lorentz metric.

Definition 2.2-14

A manifold admitting a positive definite metric is called Riemannian. A manifold admitting a Lorentz metric is called Lorentzian.

2.2.3) Maps on Manifolds and Submanifolds

In this subsection, certain special maps on manifolds are considered. The derived linear function, which maps tangent spaces to tangent spaces is defined and this leads to the definition of a submanifold. A curve on a manifold is defined in terms of a mapping.

Definition 2.2-15

Let m be a point in the domain of a differentiable function $\phi : M \to N$. (M,N do not necessarily have the same dimension). Let $n = \phi(m) \in N$. If $f \in \mathfrak{F}(n)$ then $f \circ \phi \in \mathfrak{F}(m)$ and so a vector $v \in T_m M$ determines a function $\phi_{*_m} v(f) = v(f \circ \phi)$, which is a derivation on $\mathfrak{F}(n)$. Thus, $\phi_{*_m} v \in T_n N$. The function $\phi_{*_m} : T_m M \to T_n N$ is known as the derived linear function on $T_m M$.

Clearly, if x,y are charts of M,N respectively, with the domain of x including m and the domain of y including n, ϕ_{*_m} is determined by its action,

$$\phi_{*_m} \left(\frac{\partial}{\partial x_i}\right)_m = \Sigma_\alpha \left(\frac{\partial}{\partial x_i} (y_\alpha \circ \phi)\right)_m \left(\frac{\partial}{\partial y_\alpha}\right)_n \qquad (2.2\text{-}2)$$

on the basis for $T_m M$ associated with x.

ϕ_{*_m} is a linear function on the vector space $T_m M$. Thus, by standard vector space theory, its range $\phi_{*_m}(T_m M)$ and kernel $(\phi_{*_m})^{-1} 0$ are vector subspaces of $T_n N$, and $T_m M$ respectively.

Definition 2.2-16

The dimension of the range of ϕ_{*_m} is called the rank of ϕ_{*_m}.

The rank of a differentiable function ϕ: $M \to N$ at a point m of its domain is the rank of ϕ_{*_m}.

Notice that the rank of ϕ: $M \to N$ is equal to the dimension of M if and only if ϕ_{*_m} is an injection.

Definition 2.2-17

The tangent bundle TM admits a projection π: $TM \to M$ defined by $\pi(v) = m$ if and only if $v \in T_m M$. A differentiable function ϕ: $M \to N$ determines a function ϕ_* : $TM \to TN$ defined by $\phi_*(v) = \phi_{*_m} V$, where $m = \pi(v)$. This is the differential of ϕ. (ϕ_* is sometimes called the tangent of ϕ. It is easily shown that $*$ is a covariant functor in the categorical sense).

If ϕ: $M \to N$, it is also possible to defined functions which map $T^*_m M$ to $T^*_n N$, where $m \in M$ and $n \in N$ and also functions which map $T^p_{q_m} M$ to $T^p_{q_n} N$, in terms of the derived linear function. It also plays an important role in the study of dynamical systems on manifolds (see chapter three).

Definition 2.2-18

A differentiable function ϕ: $M \to N$ is an immersion if its rank is equal to the dimension of M, at each point of its domain. Let i: $M \to N$ be the natural injection where $M \subset N$. (This is the function which maps an element in the subset M to itself, regarded as an element of the whole set N).

A manifold M is a submanifold of a manifold N if it is a subset of N and if the natural injection i: $M \to N$ is an immersion.

For example, the set of real symmetric n x n matrices is a submanifold of the set of all n x n real matrices.[12]

Definition 2.2-19

A curve c in a manifold M is defined to be the differentiable

function c: $\mathbb{R} \to M$ whose domain D is an open interval of $\mathbb{R}$.

A curve c is said to be timelike, null or spacelike if a tangent vector at each point is timelike, null or spacelike respectively.

Notice a tangent vector to a curve γ can be expressed as a multiple of $\left(\frac{\partial}{\partial t}\right)_\gamma$, in terms of the identity chart t for $\mathbb{R}$ - the tangent vector can be taken to be $\left(\frac{\partial}{\partial t}\right)_\gamma$, without loss of generality.

Freely falling particles on a space-time manifold will follow timelike geodesic curves and light rays will follow null geodesic curves. The concept of a geodesic curve is defined in terms of a linear connection, in the next subsection.

2.2.4) Linear Connections, the Curvature Tensor and Geodesics

In this subsection, an extra structure on a manifold is introduced - a linear connection. The Riemannian tensor, which gives a 'measure' of the curvature of the manifold can be defined in terms of this connection. Also, a generalized derivative of tensor fields, the covariant derivative, can be defined with this extra structure. By considering the covariant derivative of the tangent vector along a curve, a geodesic curve may be defined.

Definition 2.2-20

A C^k tensor field T of type (p,q) on a set $\mathcal{U} \subset M$, a manifold, is a map which assigns an element of $T^p_q{}_m M$ to each point $m \in \mathcal{U}$, such that the components of T with respect to any coordinate basis defined on an open subset of $\mathcal{U}$ are C^k functions. A C^k tensor field of type (1,0) is a C^k vector field and a C^∞ tensor field is simply called a vector field.

If f: $M \to \mathbb{R}$ is differentiable, and X is a vector field whose domain intersects the domain of f, Xf: $M \to \mathbb{R}$ is defined by $Xf(m) = (X(m))f$, $m \in M$. fX is a vector field with $(fX)(m) = f(m)X(m)$.

In the following definitions, by a 'function' will be meant a function mapping M to the real line.

Definition 2.2-21

A linear connection ∇ at a point $m \in M$ is a rule which assigns to each vector field X at m a differential operator ∇_X which maps an arbitrary C^r vector field Y into a vector field $\nabla_X Y$ where,

(i) $\nabla_{fX+gY} Z = f \nabla_X Z + g \nabla_Y Z$

(ii) $\nabla_X(\alpha Y + \beta Z) = \alpha \nabla_X Y + \beta \nabla_X Z$

(iii) $\nabla_X(fY) = X(f)Y + f \nabla_X Y.$

for any functions f,g; $\alpha, \beta \in \mathbb{R}$ and at least C' vector fields X,Y,Z.

Definition 2.2-22

Let X,Y be vector fields on a manifold M whose domains meet on an open set U. Let $f: M \to \mathbb{R}$, whose domain meets U.

Then $[X,Y]f = X(Yf) - Y(Xf)$ is a differential operator, called the Lie bracket of X and Y.

The set of all vector fields whose domain meet on U, with multiplication taken as the Lie bracket can be shown to be a Lie algebra[13]. Two important tensors on M can be defined in terms of the Lie bracket and the connection:-

Definition 2.2-23

The tensor field of type (1,2) given by the relation,

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$$

for arbitrary C^r vector fields X,Y is called the torsion tensor field.

The tensor field of type (1,3) given by the relation,

$$R(X,Y)Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z$$

for arbitrary C^r vector fields X,Y,Z is called the Riemann (curvature) tensor field.

In general, the torsion tensor field will not vanish for arbitrary vector fields. If T = 0, the connection is torsion-free.

As usual, these tensors are manipulated in terms of local components. Let $\{E_i\}$, $\{E^i\}$ be dual bases for the tangent and one-form space on a neighbourhood U_0 of M. The linear connection on U_0 is completely determined by n^3 functions Γ^i_{jk} where $\Gamma^i_{jk} = \langle E^i, \nabla_{E_j} E_k \rangle$. Note that Γ^i_{jk} do not transform as the components of a tensor.

Using coordinate bases for the tangent and one-form spaces, the components of the torsion tensor and Riemann tensor can be shown to be respectively,

$$T^i_{jk} = \Gamma^i_{jk} - \Gamma^i_{kj} \quad .$$

$$R^i_{jk\ell} = \frac{\partial}{\partial x_k} \Gamma^i_{\ell j} - \frac{\partial}{\partial x_\ell} \Gamma^i_{kj} + \Gamma^i_{kp} \Gamma^p_{\ell j} - \Gamma^i_{\ell p} \Gamma^p_{kj} \qquad (2.2\text{-}3)$$

where the Riemann tensor components with respect to arbitrary dual bases $\{E_i\}$, $\{E^i\}$ are defined by $R^i_{jk\ell} = \langle E^i, R(E_k, E_\ell)E_j \rangle$ [14].

Thus, the connection is torsion-free if $\Gamma^i_{jk} = \Gamma^i_{kj}$. Only this type of connection will be considered.

It can be verified that $R^i_{j(k\ell)} = R^i_{[jk\ell]} = 0$.

Definition 2.2-24

If ∇ is a linear connection and X,Y are vector fields, the field $\nabla_X Y$ is called the covariant derivative (with respect to ∇) of Y in the direction X at a point. The covariant derivative of Y, denoted by ∇Y, can be defined as the tensor field of type (1,1) which, when contracted with X, produces the vector field $\nabla_X Y$.

The components of ∇Y are written as $Y^a{}_{;b} = (\nabla_{E_j} Y)^a$ in terms of a local basis $\{E_i\}$. If $f: M \to \mathbb{R}$, $\nabla f = df$.

The definition of a covariant derivative can be extended to any C^r tensor field and, in terms of components with respect to a coordinate basis becomes,

$$\Gamma^{a\ldots d}_{e\ldots g;h} = \frac{\partial}{\partial x_h} T^{a\ldots d}_{e\ldots g} + \Gamma^a_{hj} T^{jb\ldots d}_{ef\ldots g} + (\text{all upper indices})$$

$$- \Gamma^j_{he} T^{ab\ldots d}_{jf\ldots g} - (\text{all lower indices})^{15}.$$

Definition 2.2-25

If T is a C^r tensor field defined along a curve $\gamma(t)$, the covariant derivative of T along $\lambda(t)$ is defined as $\frac{DT}{\partial t} = \nabla_{\frac{\partial}{\partial t}} \bar{T}$ where $\bar{T}$ is any C^r tensor field extending T onto an open neighbourhood of γ. T is parallelly propagated along γ if $\frac{DT}{\partial t} = 0$. ($\frac{DT}{\partial t}$ is independent of the extension $\bar{T}$).

Parallel transfer along a curve γ is a linear map from $T^p_{q_m} M$ to $T^p_{q_n} M$ which preserves all tensor products and contractions.

Definition 2.2-26

Let γ be a curve with tangent vector $\frac{\partial}{\partial t}$ at each point of the curve (with respect to the identity chart t on $\mathbb{R}$). The curve γ

is said to be a geodesic if $\frac{D}{\partial t}\left(\frac{\partial}{\partial t}\right)_\gamma$ is parallel to $\left(\frac{\partial}{\partial t}\right)_\gamma$,

i.e. if $\frac{\partial}{\partial t} = X$, $X^a{}_{;b}X^b = fX^a$ for some function f.

In this case there exists a new parameter $\lambda(t)$ along γ such that $\frac{D}{\partial v}\left(\frac{\partial}{\partial \lambda}\right)_\gamma = 0$. In terms of local coordinate bases this relation becomes

$$\frac{d^2x_a}{d\lambda^2} + \Gamma^a_{bc}\frac{dx_b}{d\lambda}\frac{dx_c}{d\lambda} = 0. \qquad (2.2\text{-}4)$$

λ is an affine parameter and is determined up to an additive and multiplicative constant.

Definition 2.2-27

If the domain of a geodesic, when parameterized by an affine parameter, is the whole of $\mathbb{R}$, the geodesic is complete. The manifold M is geodesically complete if all the geodesics on M are complete.

The concept of geodesic completeness plays an essential role in defining singularities, as will be shown later.

If a vector is parallelly transported along a curve γ, starting from $m \in M$ and ending at m, in general, a different vector will be obtained. The vectors will be the same for arbitrary closed curves, for each $m \in M$ only if $R^i_{jk\ell} = 0$ at all points of M[16]. In this case, the space is flat in the intuitive sense.

Definition 2.2-28

If $R^i_{jk\ell} = 0$ at all points of a manifold M, the defining connection is said to be flat.

In metric theories of gravitation, the unique torsion-free linear connection is determined by requiring $\nabla g = 0$, that is, in terms of local components, $g_{ij;k} = 0$. Then, in terms of a basis $\{E_i\}$ for the tangent space in a coordinate neighbourhood, the connection coefficients are given by $\Gamma_{abc} = g(E_a, \nabla_{E_b} E_c)$, and in

terms of a coordinate basis $\left\{ \frac{\partial}{\partial x_a} \right\}$ these become the usual Christoffel relations,

$$\Gamma_{abc} = g_{ad} \Gamma^{d}_{bc} = \tfrac{1}{2} \left\{ \frac{\partial}{\partial x_c} g_{ab} + \frac{\partial}{\partial x_b} g_{ac} - \frac{\partial}{\partial x_a} g_{bc} \right\} \qquad (2.2\text{-}5)^{17}$$

The Riemann tensor of the linear connection defined by the metric has the additional symmetry $R_{(ab)cd} = 0$, and as a consequence, $R_{abcd} = R_{cdab}$. This implies that the Ricci tensor, given in the following definition, is symmetric.

Definitions 2.2.29

The Ricci tensor is defined to be the contraction of the Riemann tensor, with components given by, $R_{bd} = R^{a}{}_{bad}$. The curvature scalar R is defined by $R = R^{a}{}_{a} = R^{a}{}_{bad} g^{bd}$.

Taking into account all the symmetries on the Riemann tensor, $R_{abcd} = 0$ on a one-dimensional manifold. If the dimension of the manifold is two, there is one independent component of the Riemann tensor which is essentially the curvature scalar and if the manifold is three-dimensional, the curvature tensor is completely determined by the Ricci tensor.

2.2.5) Surfaces in a Manifold

The concepts defined in this section play an important role in the theory of the large scale structure of space-time and, in particular, are needed in the following subsection on conjugate points. It is assumed throughout the subsection that the manifold in question is Lorentzian.

Definition 2.2-30

Let M be an n-dimensional manifold. If ζ is an (n-1)-dimensional manifold and $f: \zeta \to M$ is an injective immersion, $f(\zeta)$ is a hypersurface of M. Then there exists $n \in T^{*}_{f(m)}M$, $m \in \zeta$, such that for any vector

$v \in T_m M$, $\langle n, f_{*m} v \rangle = 0$. $f(\zeta)$ is respectively a spacelike, null, timelike, hypersurface if $g^{ab} n_a n_b < 0, = 0, > 0$.

Definition 2.2-31

A partial Cauchy surface is a spacelike hypersurface which no non-spacelike curve intersects more than once. A Cauchy surface is a spacelike hypersurface which every non-spacelike curve intersects exactly once.

Notice a hypersurface is a special case of a submanifold of M. Also needed for further discussion are the concepts of spacelike two-and three-surfaces, the three-surface being a special case of a spacelike hypersurface where the one-form n is derived from a function mapping $M \to \mathbb{R}$.

Definition 2.2-32

A spacelike two-surface ζ is a two-dimensional submanifold of M, defined locally by $f_1 = 0$, $f_2 = 0$ where f_1, f_2 are C^2 functions mapping $M \to \mathbb{R}$ such that when $f_1 = 0$, $f_2 = 0$ then $f_{1;a}$ and $f_{2;a}$ are non-vanishing and not parallel and

$$(f_{1;a} + \mu f_{2;a})(f_{1;b} + \mu f_{2;b}) g^{ab} = 0$$

for some real non-zero value of μ.

There exist two real values μ_1, μ_2 which satisfy the above equation. N_1^a and N_2^a, the two null vectors normal to ζ are defined as proportional to $g^{ab}(f_{1;b} + \mu_2 g_{2;b})$ and $g^{ab}(f_{1;b} + \mu_2 f_{2;b})$ respectively, and normalized by $N_1^a N_2^a g_{ab} = -1$. A pseudo-orthonormal basis can be completed by introducing two spacelike unit vectors Y_1^a, Y_2^a orthogonal to each other and to N_1^a and N_2^a. The two null second fundamental tensors of ζ are defined by,

$$ {}^{n}\chi_{ab} = -N_{nc;d}\,(Y_1{}^{c}Y_{1a} + Y_2{}^{c}Y_{2a})(Y_1{}^{d}Y_{1b} + Y_2{}^{d}Y_{2b});\quad n = 1,2. $$

An example of a spacelike two-surface in M is the sphere S^2. The three-dimensional analogue is the spacelike three-surface:-

Definition 2.2-33

A spacelike three-surface, ζ, is a three-dimensional submanifold defined locally by $f = 0$ where f is a C^2 function mapping $M \to \mathbb{R}$ and $g^{ab}f_{;a}f_{;b} < 0$ when $f = 0$.

The unit normal vector, N, to the surface is given by

$N^a = (-g^{bc}f_{;b}f_{;c})^{-\frac{1}{2}}\, g^{ad}f_{;d}$ and the second fundamental tensor χ of ζ by $\chi_{ab} = h_a{}^{c}h_b{}^{d}N_{c;d}$ where $h_{ab} = g_{ab} + N_aN_b$.

Notice any curve lying in a spacelike two-or three-surface must be spacelike. The second fundamental forms give information about the behaviour of the normals to the surface.

2.2.6) Variations of Curves and Conjugate Points

Conjugate points are basic to the proof of the singularity theorems. (A point p conjugate to a point q can be thought of as a point where infinitesimally neighbouring geodesics through q intersect.) They are defined in this subsection in terms of a vector field arising from a certain kind of variation of a curve. It is assumed that the manifold M in question possesses a metric, which is not necessarily Lorentzian.

Definition 2.2-34

Consider a curve $\gamma : \mathbb{R} \to M$ and let $\gamma : [\lambda_p, \lambda_q] \to M$ denote a segment of the curve (assuming that the parameter can be extended to λ_q). Let $p,q \in M$ denote the end points of the segment, that is, $p = \gamma(\lambda_p)$, $q = \gamma(\lambda_q)$.

Let $\mathbb{R}_s^{n+1}$ denote the subset of $\mathbb{R}^{n+1}$ consisting of all points

$(\tau_1, \tau_2, \ldots, \tau_n \lambda) \in R^{n+1}$ at which,

$$\lambda_p(\tau_1, \ldots, \tau_n) \leq \lambda \leq \lambda_q(\tau_1, \ldots, \tau_n)$$

$$-\varepsilon_i < \tau_i < \varepsilon_i \quad i = 1, \ldots, n$$

where each $\varepsilon_i > 0$.

ϕ: $R_s^{n+1} \to M$ is a n-parameter variation of $\gamma(\lambda)$ if the curve,

$$\phi(0, \ldots, 0, \lambda) = \gamma(\lambda), \text{ and}$$

$$\lambda_p(0, \ldots, 0) = \lambda_p, \ \lambda_q(0, \ldots, 0) = \lambda_q.$$

ϕ is an n-parameter variation, with fixed end points, if

$\phi(\tau_1, \ldots, \tau_n, \lambda_p) = p$, $\phi(\tau_1, \ldots, \tau_n, \lambda_q) = q$. for all values of τ_i, $i = 1, \ldots, n$.

The length of a curve $\gamma(\lambda)$ from $p = \gamma(\lambda_p)$ to $q = \gamma(\lambda_q)$ may be defined as

$$L = \int_{\lambda_p}^{\lambda_q} \left(\left| g\left(\frac{\partial}{\partial \lambda}, \frac{\partial}{\partial \lambda} \right) \right| \right)^{\frac{1}{2}} d\lambda.$$

(The above definition is usually expressed as

$$ds^2 = g_{ij} \, dx_i \, dx_j$$

where g_{ij} are the components of g with respect to the co-ordinate basis $\left\{ \frac{\partial}{\partial x_i} \right\}$. ds^2 represents the infinitesimal arc determined by the co-ordinate displacement $x_i \to x_i + dx_i$)

Variations can be used to determine whether or not a particular curve is of stationary length, that is, one of maximal or minimal length, between two fixed points. In a Riemannian manifold, an arbitrary geodesic is a curve of stationary length.[18] If the manifold is a Lorentz manifold, a timelike geodesic curve is of stationary length (assuming that each variation is also timelike). In both cases, the result of whether or not a geodesic is maximal or minimal depends on the existence of conjugate points. For a Lorentz manifold, it can be shown that a timelike geodesic curve γ from p to q is maximal if and only if there is no point conjugate to q along γ in (p,q).[19] Also, if a null geodesic contains conjugate points, it can be varied to give a timelike curve.[19] These results are used in the proof of the

singularity theorems.

Variations of curves from a surface ζ in M to a point $q \in M$ may also be considered. An n-parameter variation is defined as before except that, with fixed end points,

$\phi(\tau_1,\ldots,\tau_n,\lambda_p) = p$, $\phi(\tau_1,\ldots,\tau_n,\lambda_q) = q$ is replaced by,

$\phi(\tau_1,\ldots,\tau_n,\lambda_p)$ lies on ζ, $\phi(\tau_1,\ldots,\tau_n,\lambda_q) = q$.

Definition 2.2-35

Consider a one-parameter variation ϕ of a geodesic curve $\gamma(\lambda)$ (without fixed end points). The variation is a Jacobi variation if, for arbitrary $\tau \in (-\varepsilon,\varepsilon)$, the curve $\phi(\tau,\lambda)$ is a geodesic.

Thus, in an interval in which λ is defined, the set of Jacobi variations constitute a congruence of geodesics, of which the original curve is a member, (a congruence of curves meaning a family of curves, one through each point of some subset of M). Again if M is Lorentz ϕ is a Jacobi variation of a timelike, null,or spacelike geodesic if, for arbitrary $\tau \in (-\varepsilon,\varepsilon)$, the curve $\phi(\tau,\lambda)$ is respectively a timelike, null, or spacelike geodesic.

Definition 2.2-36

A vector field Z on a geodesic $\gamma(\lambda)$ is a Jacobi field if there exists a Jacobi variation ϕ of the geodesic $\gamma(\lambda)$ such that $Z(m) = \frac{\partial\phi}{\partial\tau}(0,\lambda_m)$ at each point $m = \gamma(\lambda_m)$ of the geodesic.

It can be shown that, in general, given a geodesic $\gamma(\lambda)$ there exists such a field[20]. The Jacobi field Z can be thought of as giving the separation of points equal distances from some arbitrary initial points along two neighbouring curves in a congruence of geodesics. It is fairly clear that the Jacobi field is related to the curvature tensor associated with the manifold in question.

Theorem 2.2-2[20]

A Jacobi field Z, on γ, satisfies $\frac{D^2Z}{\partial\lambda^2} = R(V,Z)V$ (2.2-6)

where V is the tangent vector field to γ.

□

In general, the Jacobi field will have n components if M is n-dimensional and the Jacobi equation (2.2-6) represents n equations. For the purpose of this work, it is sufficient to consider (2.2-6) expressed in terms of particular bases.

From now on it will be assumed that M is pseudo-Riemannian and four-dimensional. Only Jacobi fields on timelike and null geodesics need be considered.

Case One:- Timelike Geodesics

Consider a timelike geodesic $\gamma(\lambda)$ with unit tangent vector field V, that is, $g_m(V(m),V(m)) = -1$ at each point $m \in \gamma(\lambda_m)$. Let $\{E_i\}$ $\{E^i\}$ be dual orthonormal bases of T_mM, T_m^*M respectively, with $E_4 = V(m)$. If these are parallelly propogated along $\gamma(\lambda)$, E_4 remains equal to V(m) and the Jacobi equation, in terms of components becomes,

$$\frac{d^2}{d\lambda^2} Z^a = - R_{a4b4}Z^b \qquad (2.2\text{-}7)$$

(λ taken to be the arc-length s).

In (2.2-7) a,b = 1,...,4. However, Z(m) can be assumed to belong to the subspace of vectors orthogonal to V(m), and in this case (2.2-7) reduce to three equations, that is, a,b = 1,..., 3.

Case Two:- Null Geodesics

Consider now a congruence of null geodesics with tangent vector field K, so that $g_m(K(m),K(m)) = 0$ at each point m of a particular geodesic. Let $\{E_i^m\}$, $\{E^i\}$ be a set of dual pseudo-orthonormal

bases for $T_m M$, $T_m^* M$. ($E_4 = K(m)$, $E_3 = L(m)$ another null vector and $g(E_3,E_4) = -1$ and E_1,E_2 are unit spacelike vectors orthogonal to each other and to E_3,E_4). These dual bases can be parallelly propagated along the geodesic with $E_4 = K(m)$ and, again, the Jacobi equation in terms of components is (2.2-7). (λ taken to be the affine parameter). Projections of $Z(m)$ into a subspace of $T_m M$ consisting of equivalence classes of vectors orthogonal to $K(m)$, which differ only by a multiple of K are considered without loss of generality and in this case (2.2-7) reduces to two equations, that is, $a,b = 1,2$. Equation (2.2-7) will be considered again in Chapter four.

Definition 2.2-37

Two points on a geodesic γ are said to be conjugate if there exists on γ a non-zero Jacobi field $Z(\lambda)$ that vanishes at these two points.

Since the Jacobi field gives the separation of points equal distances from some arbitrary initial point along two neighbouring geodesics in a congruence, a point q conjugate to a point p is a point where infinitesimally neighbouring geodesics through p intersect.

The Jacobi field $Z(\lambda)$ vanishing at $p \in M$ satisfies (2.2-7) with initial conditions given by $Z(p) = 0$. In the timelike case there are three independent fields satisfying this condition and in the null case, there are two. The Jacobi fields are determined completely by specifying the values of $\dot{Z}(\lambda)$ at p. If any of these fields vanish again at q, then q must be conjugate to p.

It is also possible to define a point q conjugate to a surface $\mathcal{J}$ the point q can be thought of as a place where neighbouring geodesics normal to ζ intersect.

Definition 2.2-38

Let ζ represent a spacelike three-surface in the timelike geodesic case and a spacelike two-surface in the null geodesic case.

A point $q \in M$ along the geodesic $\gamma(\lambda)$ normal to ζ, is conjugate to ζ if there exists on $\gamma(\lambda)$ a non-zero Jacobi field Z which vanishes at q and which is related to a second fundamental tensor of ζ at $p \in \zeta$ by $\dot{Z}^a(\lambda) = \chi_{ab} Z^b$. (In the case of null geodesics, there are two families of null geodesics normal to the spacelike two-surface. If the tangent vector $K = N_i$ at ζ χ_{ab} is taken to be ${}_i\chi_{ab}$, $i = 1,2$).

The Jacobi fields which satisfy (2.2-7) are determined completely by specifying the values of $Z(\lambda)$ at $p \in \zeta$ in this case.

Definition 2.2-39

Let Z be a Jacobi field and let $Z(m) = A(m)B$ for each point $m = \gamma(\lambda_m)$ along a geodesic γ, where B is some constant column vector having the same number of components as Z(m). A is a Jacobi matrix field.

It is easily seen that any Jacobi matrix field must satisfy,

$$\frac{d^2}{d\lambda^2} A_{ab}(\lambda) = -R_{a4c4} A_{cb}(\lambda) \qquad (2.2\text{-}8)$$

$\theta = (\det A)^{-1} \frac{d}{d\lambda} (\det A)$ is called the expansion of the Jacobi matrix field. In the case of definition 2.2-37, let a Jacobi matrix be defined by $Z(\lambda) = A(\lambda)\dot{Z}(\lambda_0)$ and in the case of definition 2.2-38, let a Jacobi matrix be defined by $Z(\lambda) = A(\lambda)Z(\lambda_p)$. Clearly, in both cases, the expansion θ becomes infinite at the conjugate point q. This fact forms the basis of proof of the following theorems on the existence of conjugate points.

Proposition 2.2-3[21]

(i) If at some point $\gamma(\lambda_1)$ $(\lambda_1 > 0)$, the expansion θ has negative value $\theta_1 < 0$ and if $R_{ab} V^a V^b \geq 0$ everywhere where V is the tangent

vector field to the congruence of timelike or null geodesics, then there will be a point conjugate to p along $\gamma(\lambda)$ between $\gamma(\lambda_1)$ and $\gamma(\lambda_1 + (n/-\theta_1))$, n = 3 or 2 respectively, provided that $\gamma(\lambda)$ can be extended to this parameter value.

(ii) If $R_{ab}V^aV^b \geq 0$ and $\chi_{ab}g^{ab} < 0$, there will be a point conjugate to ζ along a timelike or null geodesic $\gamma(\lambda)$ within a distance $n\,|(-\chi_{ab}g^{ab})$, n = 3 or 2 respectively, from ζ, provided that $\gamma(\lambda)$ can be extended that far. □

(The second part of the proposition follows easily from the first as the initial value of θ is $\chi_{ab}g^{ab}$).

Proposition 2.2-4[21]

If $R_{ab}V^aV^b \geq 0$ everywhere where V is the tangent vector field to the congruence of timelike or null geodesics and if at some point $p = \gamma(\lambda_p)$, $V^cV^dV_{[a}R_{b]cd[e}V_{f]} \neq 0$ there will be λ_q, λ_r such that $q = \gamma(\lambda_q)$, $r = \gamma(\lambda_r)$ are conjugate along γ, provided γ can be extended to these values. □

2.3) The Local Structure of Space-time

This section looks at Einstein's theory of gravitation from a mathematical viewpoint. The relationship between the various matter fields and space-time curvature is given in tensor form and the effect of curvature on matter is given in terms of the geodesic equations.

The theory assumes that space and time are inseparable and so any event in space-time is represented by three space coordinates and a time coordinate. The Einstein theory is of the metric type and so the model for space-time is a connected, four-dimensional manifold, together with a Lorentz metric. (The manifold is taken to be connected since clearly, there would be no knowledge of any disconnected component). A Lorentz metric exists because of the assumptions of the structure of space-time. Its canonical form is the Minkowski metric of Special Relativity. (It is assumed to be at least C^2, so that the field equations can be defined at every point).

In this work, it is also assumed that the manifold is orientable and so it is possible to define continuously a division of non-separable vectors into two classes, which are labelled future- and past-directed.[22]

There are various matter fields on the space-time manifold which obey equations which can be expressed locally as relations between tensor components. One of the assumptions of the theory concerns these equations:-

Postulate One: Local Conservation of Energy and Momentum

There exists a symmetric tensor $_{\wedge}T^{ab}$ which depends on the matter fields and the metric and which satisfies,

(i) $T^{ab} = 0$ is a neighbourhood of the manifold if and only if all the matter fields vanish there.

(ii) $T^{ab}{}_{;b} = 0.$

T^{ab} is called the energy-momentum tensor.

The relationship between the curvature and the matter fields can clearly be expressed as a tensor relation between the curvature tensor and the energy-momentum tensor. In terms of local coordinate components, these are the Einstein field equations.

Postulate Two: The Einstein field equations

The field equations,

$$R_{ab} - \tfrac{1}{2}R\, g_{ab} + \Lambda g_{ab} = 8\pi\, T_{ab}, \quad \Lambda \in \mathbb{R} \qquad (2.3\text{-}1)$$

hold on the space-time manifold.

Given the energy-momentum tensor, the equations (2.3-1) determine the metric for the space-time manifold completely. Thus, any Lorentz metric is a space-time metric if the energy-momentum tensor is defined as the right hand side of (2.3-1). Naturally, most of these space-times will be physically unreasonable. Any solution of (2.3-1) will only lead to a realistic metric, if the given energy-momentum tensor has physically reasonable properties. In general, given the components

T_{ab}, (2.3-1) are impossible to solve exactly, even for the free space case, $T_{ab} = 0$.

Having obtained a suitable metric related to the energy-momentum tensor, the fact that curvature has an effect on matter is expressed mathematically by the fact that test particles must follow the timelike geodesics defined from the metric.

Obviously, solving the field and geodesic equations locally and patching together the solutions is not a satisfactory way of looking at the global structure of space-time, hence the need for global topological methods as used to prove the singularity theorems.

2.4) The Global Structure of Space-time:- Singularities

The aspect of the global structure of space-time with which this work is concerned is space-time singularities. This section is concerned with their definition and with the theorems which predict their occurrence.

2.4.1) Space-time Singularities

One of the most alarming facets of the Einstein theory of gravitation is the possibility of the existence of singularities in the structure of space-time. In simple terms, a space-time singularity can be thought of as a part of space-time where our accepted laws of physics breakdown. It could be a region of infinite gravitational forces or a region in which null or timelike geodesics terminate. In short, it can be thought of as a part of the 'edge' of space-time, which is at a finite distance instead of at infinity.

By a 'singularity' is meant a singularity of the manifold structure itself - the so-called 'apparent singularities' do not come under this heading. These are apparent singularities which occur in the space-time due to the fact that the metric has been expressed in the wrong choice of coordinate system in that particular region - in general, a manifold has to be defined in terms of more than one coordinate system, that is, in terms of more than one chart.

2.4.2) The Schwarzchild Metric

Outside a spherical and static star, the space-time geometry is obtained from the Schwarzchild metric,

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} + r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2\right) \qquad (2.4\text{-}1)$$

in terms of Schwarzchild coordinates and M is the mass inside radius of the star, which is constant outside the star.[23]

The metric (2.4-1) appears to behave badly near $r = 2M$, the Schwarzchild radius, as g_{rr} becomes infinite there and g_{tt} vanishes. However, it is not certain, without careful study, whether this pathology in the metric is due to the presence of a singularity in the space-time manifold itself, or merely to a pathology in the Schwarzchild coordinate system near $r = 2M$.

If the curvature scalar is calculated from the metric (2.4-1) in terms of a travelling frame of reference, it can be seen that it is a perfectly well-behaved function near $r = 2M$. Thus, there are no huge gravitational forces at $r = 2M$. The non-singular nature of the region $r = 2M$ is supported by the fact that other coordinate systems can be constructed from which the pathology is absent:-

For example, consider the coordinates defined by,

$$\left.\begin{aligned} u &= (r/2M - 1)^{\frac{1}{2}} e^{r/4M} \cosh(t/4M) \\ v &= (r/2M - 1)^{\frac{1}{2}} e^{r/4M} \sinh(t/4M) \end{aligned}\right\} \quad r > 2M$$

$$\left.\begin{aligned} u &= (1 - r/2M)^{\frac{1}{2}} e^{r/4M} \sinh(t/4M) \\ v &= (1 - r/2M)^{\frac{1}{2}} e^{r/4M} \cosh(t/4M) \end{aligned}\right\} \quad r < 2M \qquad (2.4\text{-}2)$$

These are the Kruskal-Szekeres coordinates. Under the transformation (2.4-2), the metric (2.4-1) becomes,

$$ds^2 = (32M^3/r)e^{-r/2M} \left(-dv^2 + du^2\right) + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right) \qquad (2.4\text{-}3)$$

where $u^2 - v^2 = (r/2M - 1)e^{r/2M}$.

Another well-behaved coordinate system is the Novikov system.[24]

By contrast, at r = 0, there exist infinite gravitational forces - the curvature scalar becomes infinite there. This region is a singularity of the space-time itself and no transformation of the coordinate system will bring about its absence. For example, it is still present in the transformed metric (2.4-3).

2.4.3) The Definition of Space-time Singularities

One of the most troublesome facets of the study of singularities is their very definition. Many hold the view that there still does not exist a fully satisfactory definition of a singularity.

It must be supposed that a singularity cannot be part of the space-time manifold itself, because of its very intuitive definition. The manifold is defined as the region in which the laws of physics hold - the singularities are the regions where these laws break down. The problem of defining whether the given space-time has a singularity now becomes one of determining whether or not any singular points have been 'cut out'. This should be recognized by the fact that the space-time is incomplete in some sense. Unfortunately, as mentioned in section 2.2), a Lorentz metric, unlike a positive-definite Riemannian metric does not define a metric in the topological sense and so singularities of a space-time cannot be defined in terms of metric incompleteness in the topological sense. However, in the case of Riemannian manifolds, it can be shown that metric completeness implies geodesic completeness. In fact, it can be shown that the two concepts are equivalent.[25] Thus, with a Lorentz metric, singularities may be defined in terms of geodesic incompleteness - most of the more successful definitions of singularities involve geodesic incompleteness. The significance of spacelike geodesic incompleteness is not clear, if the possibility of the existence of objects which follow spacelike curves, such as tachyons, is excluded. Thus, we have,

Definition 2.4-1

A space-time is said to be singular if it is either timelike or null geodesically incomplete.

(Thus timelike and null geodesic completeness are minimum conditions for space-time to be considered singularity-free). Note, however, that timelike and null geodesic completeness do not imply that the space-time contains no singularities.

The singularity theorems predict null or timelike geodesic incompleteness, and so predict the occurrence of a singularity in the sense of definition 2.4-1.

2.4.4) The Singularity Theorems

Many of the exact solutions of the Einstein field equations, which are realistic, contain singular points. However, all these solutions assume a high degree of symmetry and it was originally suggested by a number of people that singularities were simply the result of the assumed symmetries in the given space-time and that they would not occur in general solutions. This view was supported by Lifshitz and Khalatnikov[26]. With the appearance of the singulairty theorems, which predict the occurrence of singularities under quite general assumptions, the claim has been withdrawn.

The first theorem concerning singularities which did not involve any assumption of symmetry was given by Penrose, using his idea of a closed trapped surface. It was designed to prove the occurrence of a singularity in a star which collapses inside its Schwarzchild radius.

Definition 2.4-2

A closed trapped surface is a C^2 compact spacelike two-surface ζ (normally taken to be S^2), without boundary, such that the two families of null geodesics orthogonal to ζ are converging at ζ (that is, ${}_1\chi_{ab}g^{ab}$ and ${}_2\chi_{ab}g^{ab} < 0$).

Theorem 2.4-1[27]

Space-time M cannot be null geodesically complete if :-

(i) $R_{ab} K^a K^b \geq 0$ for all null vectors K^a at each point of M;

(ii) there is a non-compact Cauchy surface $\mathcal{C}$ in M

(iii) there is a closed trapped surface ζ in M.

□

It can be shown that assuming the Cauchy surface $\mathcal{C}$ to be non-compact is not a serious restriction - the theorem still holds if $\mathcal{C}$ is compact, provided there is an observer who can avoid falling into the collapsing star.[28] If the following reasonable condition is assumed for the energy-momentum tensor, then the Einstein equations can be used to show that condition (i), the null convergence condition, must hold[28]:-

The Weak Energy Condition

The energy-momentum tensor at each point $m \in M$ obeys the inequality $T_{ab} w^a w^b \geq 0$ for any timelike vector $w \in T_m M$. By continuity, this will then also be true for any null vector $w_1 \in T_m M$. (This holds for all experimentally detected fields).

Notice if the Einstein equations are assumed and the collapse of the star is exactly spherical, the equations may be integrated to give an exact solution - this will always be a singular metric. The theorem states that if exact spherical symmetry is not assumed, as long as a closed trapped surface is present, the singularity must still occur. It can be shown that a closed trapped surface will still occur providing the departures from spherical symmetry are not too great.[29]

Since the null convergence condition is assumed and even 'outgoing' rays are dragged back to ζ and are converging, something is bound to go wrong. The conditions (i) and (iii) imply the existence of conjugate points, if it is assumed that M is null geodesically complete and this leads to a contradiction and the results of the theorem, using condition (ii). (It is assumed in all the theorems that local causality holds, which basically states that a signal can be sent between two points of space-time which can be joined by a unique geodesic if and only if they can be joined by a non-spacelike curve. This rejects the

possibility of signals being sent along spacelike curves - this is physically reasonable, since there is no practical evidence of the existence of concepts such as tachyons).

The weakness of the theorem lies in the fact that a Cauchy surface has been assumed to exist, whereas there is no reason to suppose such an artificial construction exists in the real universe. Thus, the theorem may not be useful in physically realistic situations. The second singularity theorem does not assume the existence of such a surface and has wider applications. It is also concerned with timelike geodesic incompleteness as well as null geodesic incompleteness. Before the theorem can be quoted, a few more definitions are needed.

Definition 2.4-3

The generic condition is satisfied if every non-spacelike geodesic contains a point at which $K_{[a}R_{b]cd[e}K_{f]}K^{c}K^{d} \neq 0$, where K^{a} is the tangent vector to the geodesic at any point.

If the generic condition holds on a manifold M, then M is sufficiently general, that is, not too highly symmetric. The physically reasonable condition requires M to be curved - this excludes the possibility of Minkowski space-time being a model for the whole universe. It can easily be seen that Minkowski space-time is geodesically complete.

Another reasonable assumption for a space-time is that an observer cannot travel away from a point in the manifold and arrive back just as he was leaving, as this leads to all sorts of paradoxes. In this case, the observer can traverse a closed timelike curve. A slightly stronger condition is that there should be no non-spacelike curves and it is also reaonable to exclude situations in which there were non-spacelike curves which passed arbitrarily close to other non-spacelike curves which then passed arbitrarily close to the origin of the first curve - and so on. Thus, we have the following definitions:-

Definition 2.4-4

The chronology condition holds on M if there are no closed timelike

curves and the causality condition holds on M if there are no closed non-spacelike curves.

Definition 2.4-5

The strong causality condition holds at $m \in M$ if every neighbourhood of m contains a neighbourhood of m which no non-spacelike curve intersects more than once. The strong causality condition holds on M if it holds at every point of M.

It can be shown that if the null convergence and generic conditions hold, then the causality and chronology conditions are equivalent. Also, if M is assumed to be compact, then there are regions of M where the chronology condition does not hold.[30] Thus, the following singularity theorem assumes implicitly that M is non-compact. It will be seen later that if M is compact, M is geodesically complete and, if this is the case, causality and strong causality are equivalent.[30]

Definition 2.4-6

A set ζ is achronal if there are no two points of ζ with timelike separation, that is, no two points of ζ can be joined by a timelike curve lying within ζ. ζ is acausal if there are no two points of ζ with non-spacelike separation.

Notice every acausal set is necessarily achronal. The 'edge' of an achronal set can be thought of as being a sort of boundary where points have timelike separation. The precise definition is,

Definition 2.4-7

For $\zeta, \mathcal{U} \subset M$, let $I^{\sigma}(\zeta, \mathcal{U})$, ($\sigma$ = + or -), denote the set of all points in $\mathcal{U}$ which can be reached by a future-directed or past-directed timelike curve from ζ in $\mathcal{U}$, respectively. If ζ is achronal, the edge of ζ denoted by edge (ζ), is defined as the set of all points $q \in \bar{\zeta}$ (where $\bar{\zeta}$ denotes the closure of ζ) such that in every

neighbourhood $\mathcal{U}$ of q there exist points $p \in I^-(q, \mathcal{U})$, $r \in I^+(q, \mathcal{U})$ which can be joined by a timelike curve in $\mathcal{U}$ which does not intersect ζ.

Notice an acausal set must be a spacelike hypersurface and it can be shown that a partial Cauchy surface is an acausal set with no edge.[31]

The second singularity theorem can now be stated as,

Theorem 2.4-2[32]

Space-time M is not timelike and null geodesically complete if:-

(i) $R_{ab}K^aK^b \geq 0$ at each point of M for every non-spacelike vector K^a;

(ii) The generic condition holds on M;

(iii) The chronology condition holds on M;

(iv) There exists at least one of the following:-

(a) A compact achronal set without edge

(b) a closed trapped surface

(c) a point $m \in M$ such that the null geodesics from m are focussed by the matter or curvature and start to reconverge.

□

As before, condition (i) holds for null vectors if the weak energy condition holds. It holds for timelike vectors if the following condition holds, using the Einstein equations:-[28]

The Strong Energy Condition

The energy-momentum tensor at each $m \in M$ obeys the inequality $T_{ab}w^aw^b \geq \frac{1}{2} w^aw_aT$ for any timelike vector $w \in T_mM$, where $T = T^a_{\ a}$.

In this case, (i) is called the timelike convergence condition. Condition (i) holds if both the timelike and null convergence conditions hold. Condition (iv)(a) is really equivalent to assuming the existence of a partial Cauchy surface and so is a weaker requirement than condition (ii) of theorem 2.4-1 and this is now only one of three alternatives. Condition (iv)(c) is closely related to the existence of a closed trapped surface. In the case in which the light cone is our own

past light cone, it is possible to determine whether or not this condition is satisfied. Recent observations of the microwave background indicate that it is.

Conditions (i)-(iii) of the theorem imply that strong causality holds, assuming that M is geodesically complete. If condition (iv) (b) or (c) is assumed, that is, a converging effect on the geodesics, together with condition (i), then conjugate points will exist and things go wrong. Conjugate points will also exist under assumptions (i) and (ii). (M assumed to be geodesically complete). As in theorem 2.4-1, the assumption that M is geodesically complete together with the existence of conjugate points leads to a contradiction.

Notice, if in theorem 2.4-1, the chronology condition is also assumed, then it is just a special case of theorem 2.4-2. Notice also,that if condition (iv)(a) is assumed theorem 2.4-2, the theorem is fundamentally different than the first singularity theorem, as in the former no converging effect is assumed in a region of M, whereas in the latter, it is. This point is discussed again in chapter four.

Theorem 2.4-2 has the disadvantage of not showing whether the predicted singularity is in the future or the past. If condition (iv)(c) is strengthened to say that all past-directed timelike as well as null geodesics from $m \in M$ start to reconverge within a compact region to the past of m, it can be shown that there is a singularity in the past. This is the content of the third singularity theorem:-

Theorem 2.4-3[33]

If in M,

(i) the null and timelike convergence conditions hold;

(ii) the strong causality condition holds;

(iii) there exists a past-directed unit timelike vector w at a point m and a constant $b > 0$ such that if V is the unit tangent vector to the past-directed timelike geodesics through m. then on each geodesic, $V^a{}_{;a}$ becomes $< \frac{3c}{b}$

within a distance b/c from m, where $c = -w^a V_a$;
then there is a past incomplete non-spacelike geodesic
through m. □

Notice that the theorem has a dual which predicts the occurrence of a future incomplete non-spacelike geodesic from a point if, in condition (iii), 'past' is replaced by 'future'.

The second and third singulairty theorems both use the assumption that space-time does not contain any closed timelike or non-spacelike curves. If the chronology condition is not assumed, the first theorem still predicts the occurrence of a singularity under the artificial assumption of the existence of a Cauchy surface. The following singularity theorem, which is the timelike analogue of the first, predicts the occurrence of a singularity without the chronology condition and the existence of a Cauchy surface:-

Theorem 2.4-4[34]

Space-time M is not timelike geodesically complete if,

(i) $R_{ab} w^a w^b \geq 0$ for every non-spacelike vector w at each point of M;

(ii) There exists a spacelike three-surface ζ which is compact and without boundary, such that the unit normals to ζ are everywhere converging (the second fundamental form of ζ is negative) and edge (ζ) is empty. □

Condition (ii) may be interpreted as saying that the universe is spatially closed and contracting. It can be shown that the theorem still holds if the unit normals to ζ are everywhere diverging - in this case, the universe is expanding. Notice ζ is necessarily achronal and so edge (ζ) is defined.

The converging effect predicts the existence of conjugate points on every future-directed geodesic orthogonal to the surface (assuming M to be complete). The proof proceeds by taking a covering manifold

$\hat{M}$ in which each connected component of the image of ζ is diffeomorphic to ζ and is a partial Cauchy surface in $\hat{M}$. If M is assumed to be geodesically complete (and hence $\hat{M}$), this leads to a contradiction if conjugate points exist.

2.5) Resumé and Critique

At the present time, the Einstein theory has proved to be one of the best for gravitation theory, one major reason being that it agrees with all current experimental evidence. However, we must bear in mind that although the Einstein model is widely accepted at the present time, it may not be the exact model for our physical universe - it may only be an excellent approximation, within the framework of our present experience. It would be quite naive to think just because the theory presents us with such accurate local results, it is the exact model for large scale structure of which we have no practical experience, or for the very small scale structure.

First of all, it is not clear whether a continuous description of space-time (under the assumption of a manifold structure) is appropriate on a scale small enough so that quantum phenomena become important. If it is accepted that present day quantum and gravitation theory can be accurately extrapolated down to a dimension of 10^{-33} cm, then the quantum fluctuations in the curvature of space-time would be large enough to produce alterations in the topology!

The assumption that there exists an energy-momentum tensor with zero covariant derivative is physically reasonable as this just expresses the fact that all fields have energy and implies approximate conservation of energy, momentum and angular momentum in a small region of space-time. As stated in section 2.4), the weak energy condition on the energy-momentum tensor holds for all experimentally detected fields. A slightly stronger condition is the dominant energy condition which also requires for every timelike vector w, $T^{ab}w_a$ is non-spacelike and implies that creation of matter is impossible

in the sense that space-time must remain empty if it is empty at one time and no matter comes in from infinity. It is this condition which leads to the result that matter cannot travel faster than light.[35] It can be shown that the strong energy condition may break down in extreme cases, but is physically reasonable almost everywhere.

The Einstein field equations are fundamental to the Einstein theory of relativity. For reasons stated earlier, there is no guarantee that the equations hold at or near a predicted singularity - there is no experimental evidence to support that they hold in the strong field region, that is, areas of high gravitational attraction. The singularity theorems predict geodesic incompleteness in a lot of the space-times derived from the Einstein model due to the fact that a lot of the solutions to the field equations became singular in some region. The predicted singularities can be thought of as in the form of a 'black hole' within a collapsed star or as the highly condensed and chaotic, initial state of the universe in a cosmological model - a 'white hole'. In the neighbourhood of such singularities, the radii of curvature would become less than 10^{-33} cm and by previous comments, we would expect the Einstein model to become inadequate at such levels. Thus, the singularity theorems seem to imply that the Einstein theory predicts its own downfall!

Both 'black holes' and 'white holes' are unobservable. By definition, not even light can escape from a black hole and so there can be no communication of any kind from inside. Thus, from the point of view of what can be observed, the 'black hole' has only a metaphysical reality. The only observable facets are the effects the 'black hole' has on its surroundings. Certainly, there are phenomena in the observable universe which agree with the theoretical effects produced by a 'black hole'. However, it is not certain whether the effects are produced by a very high gravitational field in a non-singular region of space-time rather than by an infinitely large gravitational field in a singular region. Therefore, it is impossible to verify the singularity theorems experimentally at the present time and perhaps, it may never be possible and so they remain a theoretical implication of the Einstein theory.

The four singularity theorems are all based on the fundamental assumption that the null and timelike convergence conditions hold. The dirst, third and fourth theorems use these conditions, together with some condition of a converging effect to predict the occurrence of conjugate points, this being inconsistent with the other assumptions of the theorem, such as the existence of a partial Cauchy surface or the strong causality condition, if space-time is geodesically complete. If condition (iv)(b) or (c) of theorem 2.4-2 is assumed, the message is similar to that of the other three theorems - in fact, in this case the generic condition is not needed, as the occurrence of a singularity could be proved along similar lines to the other theorems. If condition (iv)(a) of the theorem is assumed, the generic and the timelike and null convergence conditions are used to predict the occurrence of conjugate points. This is then shown to be inconsistent if space-time is geodesically complete.

If it is assumed that the Einstein equations do not hold in regions of very large gravitational forces, the physically reasonable weak and strong energy conditions may no longer predict the timelike and null convergence conditions (a specific example is given in Section 6.2.4). Also, there is no reason to suppose that there exists a Cauchy surface in the physical universe as this really is just a mathematical device. If large departures are made from spherical symmetry, there is no reason to suppose the existence of a closed trapped surface.

Singularities in space-time are both physically and mathematically objectional features as we have no physical or mathematical control in singular regions. The lack of mathematical control is brought out by the fact there is still no suitable definition of a singularity. For instance, it may be thought that a non-singular space-time is one in which timelike and null geodesically complete. However, Geroch has shown that geodesic incompleteness, which appears to be ideally suited to formulating a definition of a singularity turns out to lead to difficult problems. He gives an example of a space-time which is spacelike, null and timelike geodesically complete and so seems to be non-singular, but which

contains a timelike curve of bounded acceleration and finite total length,[36] which seems to contradict the non-existence of singularities in the space-time!

One might hope that a space-time which was null, timelike or spacelike geodesically complete would also be complete in the other two. Unfortunately, this is not necessarily so. In fact, Geroch shows there are five different categories of space-time where some of the types of geodesics are complete and the others are incomplete![37]

More recently, Schmidt[38] has produced a definition of a singularity which seems to be slightly more satisfactory. He has introduced a generalization of the concept of an affine parameter to all C' curves, geodesic or not. He defines a notion of completeness by requiring that every C' curve can be extended to arbitrary values of its generalized affine parameter. For a geodesic, this generalized affine parameter corresponds to the affine parameter. Thus, completeness in the sense of Schmidt implies geodesic completeness, but the example by Geroch shows that the converse is not true.

In most physically realistic situations the concept of incompleteness by Schmidt corresponds to the existence of regions where the curvature becomes unboundedly large. However, an example of a solution to the Einstein equations where this does not seem to be true is provided by 'Taub-NUT' space.[39] Thus, even Schmidt's definition is not totally satisfactory.

In chapter four, the correspondence between gravitation and singularity theory, as presented in this chapter and dynamical systems theory is discussed. The following chapter records the basic definitions and results of dynamical systems theory needed for this discussion.

Chapter Three : Dynamical Systems Theory

In this chapter, a survey of the theory of dynamical systems is given and the more important results are mentioned. Of course, this is by no means a thorough account and only results which are applicable to the problem in question are given. References are given for the proofs of the various theorems and for results which due to lack of space are omitted.

After giving an abstract definition of a dynamical system, the close link with classical autonomous systems of ordinary differential equations is made explicit. It is then shown how a vector field acting on a manifold gives rise to an autonomous system of ordinary differential equations and hence may be to a dynamical system. The qualitative behaviour of such systems is investigated with reference to such topics as the behaviour of orbits in a neighbourhood of critical elements, limits and minimal sets of the orbits, structural stability and bifurcation theory.

3.1) Dynamical Systems

The abstract definition of a dynamical system is given and a few examples are presented, including the tie-up with systems of ordinary autonomous differential equations.

3.1.1) Definition and Examples

Definition 3.1 -1

A dynamical system (X,ϕ), is a topological space X, together with a continuous mapping ϕ: $X \times \mathbb{R} \to X$ which satisfies the following axioms:-

$$\phi(x,0) = x \qquad \text{(A1)}$$

$$\phi(\phi(x,t),s) = \phi(x,t+s) \qquad \text{(A2)}$$

for all $x \in X$ and $s,t \in \mathbb{R}$.

(A1) is known as the initial value axiom and (A2) is the group axiom.

A semi-dynamical system differs from a dynamical system essentially by the fact that the development of the system is specified into positive values of the parameter only, that is, $\mathbb{R}$ is replaced by

the set $\mathbb{R}_{>0}$. A local dynamical system is defined as a dynamical system, except that $\mathbb{R}$ is replaced by some subset of $\mathbb{R}$.

Example 3.1-1

Let $X = \mathbb{R}$ and define a function ϕ by,

$$\phi(x,t) = x + t \quad \text{for all } x,t \in \mathbb{R}.$$

Clearly this function satisfies both the axioms. Using the usual metric on $\mathbb{R}$ and a metric on $\mathbb{R}^2$, it is seen that ϕ: $\mathbb{R}^2 \to \mathbb{R}$ is continuous on $\mathbb{R}^2$ and so it follows that $(\mathbb{R}^2,\phi)$ is a dynamical system.

Example 3.1-2) :- Involving a Partial Differential Equation

Consider the first order linear partial differential equation

$$u_x + u_t = 0 \;\; ; \;\; u = u(x,t) \qquad (3.1\text{-}1)$$

subject to the initial condition $u(x,0) = f(x)$.

It is well known that the general solution to such an equation is,

$$u(x,t) = f(x-t) \qquad (3.1\text{-}2)$$

Consider a mapping ϕ : $X \times \mathbb{R} \to X$ defined by,

$$\phi(f,t)(x) = u(x,t) \quad \text{for all } f \in X,\ t \in \mathbb{R} \qquad (3.1\text{-}3)$$

where X is the space of bounded continuous functions.

Clearly, the function ϕ satisfies the given axioms (A1), (A2). It remains to show that ϕ is continuous.

Suitable metrics for X and $X \times \mathbb{R}$ are given by,

$$\sigma(f,g) = \underset{p \in \mathbb{R}}{\text{l.u.b}} \; | f(p) - g(p) | \quad \text{for all } f,g \in X \qquad (3.1\text{-}4)$$

and

$$\rho((f,t),(g,s)) = \underset{p \in \mathbb{R}}{\text{l.u.b}} \; |f(p) - g(p)| + |t-s| \qquad (3.1\text{-}5)$$

for all $(f,t),(g,s) \in X \times \mathbb{R}$.

Now $\sigma(\phi(f,t), \phi(g,s)) = \underset{p \in \mathbb{R}}{\text{l.u.b}} \; |\phi(f,t)(p) - \phi(g,s)(p)|$.

Let this value of p be denoted by p_1.

$$\therefore \quad \sigma(\phi(f,t), \phi(g,s)) = |f(p_1 - t) - g(p_1 - s)| \tag{3.1-6}$$

$$\text{Now } \rho((f,t),(g,s)) < \delta \Rightarrow \underset{p \in \mathbb{R}}{\text{l.u.b}} \; |f(p) - g(p)| < \delta_1 \text{ and } |t-s| < \delta_2 \tag{3.1.7}$$

where $0 < \delta_1 < \delta, \quad 0 < \delta_2 < \delta$

The function f is continuous and bounded and so in particular, it is continuous and bounded at $p_1 - s \in \mathbb{R}$.

$$\therefore \quad |t-s| < \delta_2 \Rightarrow |f(p_1 - t) - f(p_1 - s)| < \eta \text{ , say.} \tag{3.1-8}$$

From (3.1-6) it is seen that,

$$\sigma(\phi(f,t), \phi(g,s)) = |f(p_1 - t) - f(p_1 - s) + f(p_1 - s) - g(p_1 - s)|$$

$$\leq |f(p_1 - t) - f(p_1 - s)| + |f(p_1 - s) - g(p_1 - s)| \tag{3.1 -9}$$

Choosing δ suitably to make $\delta_1 + \eta < \varepsilon$, it is seen from (3.1-8) and (3.1-7) that ϕ is continuous at any point (g,s) of $X \times \mathbb{R}$, and so (X,ϕ) constitutes a dynamical system.

Notice that if the initial condition of (3.1-1) is not restricted to bounded functions, then (3.1-1) need not lead to a dynamical system.

It is possible to construct more complicated examples of partial differential equations which lead to dynamical systems.[1]

3.1.2) Systems of Autonomous Ordinary Differential Equations

Consider the system of ordinary autonomous differential equations in $\mathbb{R}^n$,

$$\frac{dx}{dt} = \dot{x} = f(x) \qquad (3.1\text{-}10)$$

where $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}^n$ and $t \in \mathbb{R}$.

Notice (3.1-10) can be taken to represent most autonomous ordinary differential equations, for if an n^{th} order equation is considered, it can be transformed to the form (3.1-10) by taking each derivative, except the n^{th}, to be a new dependent variable.[2] It is assumed that (3.1-10) has the following properties:-

(i) $f: \mathbb{R}^n \to \mathbb{R}^n$ is continuous

(ii) The equation has uniqueness of solutions to the initial problem, that is, there is only one solution of (3.1-10) satisfying $x(0) = x_o$.

(iii) The equation has global existence of solutions. That is, each solution can be extended to a solution defined on the entire real line $\mathbb{R}$.

A function defined on $\mathbb{R}^n \times \mathbb{R}$, called the flow, can be associated with (3.1-10) as follows. For any $x_o \in \mathbb{R}^n$, $t \in \mathbb{R}$, consider the unique solution $\sigma: \mathbb{R} \to \mathbb{R}^n$ of (3.1-10) with $\sigma(0) = x_o$; then define ϕ at (x_o, t) by,

$$\phi(x_o, t) = \sigma(t) \qquad (3.1\text{-}11)$$

It is easily verified that if (3.1-10) satisfies the conditions given, $(\mathbb{R}^n, \phi)$ is a dynamical system, (continuity of ϕ being essentially the classical theorem concerning continuous dependence on initial conditions). It will be seen in the next section that a flow can be associated with any given vector field on a manifold.

This work is concerned mainly with dynamical systems arising from equations of the type (3.1-10) and to simplify terminology, if (3.1-10) leads to a dynamical system, (3.1-10) will be called a dynamical system.

Example 3.1-3.

Consider the one-dimensional equation,

$$\dot{x} = \ln x$$

This is an autonomous equation but has a singularity at x=0, that is, $\ln x$ is discontinuous at $x = 0$. The equation does not give rise to a dynamical system.

Example 3.1-4

In contrast to the above example, consider the one-dimensional autonomous equation,

$$\dot{x} = -1-x^2$$

This time $f(x) = -1-x^2$ is continuous, but the equation still does not give rise to a dynamical system as

$$\phi(x_o,t) = \sigma(t) = \cot(t+ \cot^{-1}x_o)$$

and the solution is not defined on the entire real line.

The two cases above are examples of local dynamical systems. Notice it may be possible for a given system of equations to have global existence of solutions, but which is not dynamical.

Example 3.1-5

Consider the one-dimensional equation,

$$\dot{x} = \frac{1}{x^2}$$

so that $f(x)$ is discontinuous at $x = 0$. The general solution is $x = \sqrt[3]{3(t+t_o)}$ and so the equation has global existence of solutions but is not dynamical - it does not have uniqueness of solutions.

It can be seen that, in general, it is difficult to prove that

a given autonomous system of differential equations is dynamical. In particular, it has to be shown that every solution is defined for all $t \in \mathbb{R}$. This cannot be done directly without solving the equations explicitly which is impossible in general. However, there are some theorems which do assure us of global existence of solutions, if the equations have certain properties:-

Theorem 3.1.-1 [3]

If in (3.1-10), $f(x)$ is continuous for all $x \in \mathbb{R}^n$ and, moreover, if each component of f, $f_i(x_1,\ldots,x_n) = 0$ $(|x_1|+|x_2|+\ldots+|x_n|)$ for $|x_1| + \ldots + |x_n| \to \infty$, then the solutions of (3.1-10) are defined on the whole axis, $-\infty<t<\infty$. □

Other theorems exist which prove a given system is semi-dynamical, either by using Lyapunov functions[4], or by considering perturbations of a known dynamical system[5].

All of these theorems are difficult to apply in practice and theorem 3.1-1 is of strictly limited use.

Having proved that the solutions can be extended to arbitrary values of the defining parameter, it is still necessary to show that the given autonomous system has uniqueness of solutions. This is, in general, a lot easier to show.

Theorem 3.1-2[6]

If, in (3.1-10), each of the partial derivatives $\frac{\partial f_i}{\partial x_j}$ $(i,j = 1,2,\ldots,n)$ exists and is bounded by k in some convex set D of the n-dimensional x space, then (3.1-10) has uniqueness of solutions in D. □

Note that uniqueness of solutions still holds if the conditions of the theorem are replaced by a weaker requirement.[6] The conditions of the theorem also assure us of local existence of solutions.

Example 3.1-6

Consider the one-dimensional equation,

$$\dot{x} = x$$

From theorems 3.1-2, 3.1-1, it has uniqueness and global existence of solutions and so is dynamical. It is one of the simplest examples of a dynamical system.

3.2) Vector fields and Dynamical Systems

Recall that a vector field assigns a tangent vector to each point of a manifold. It is shown in this section how a vector field may give rise to a dynamical system, with the manifold as the underlying topological space.

Notation Let $\mathcal{X}(M)$ denote the set of all vector fields on M.

Recall that when $m \in U$, a domain of a chart x for M, the vectors $(\frac{\partial}{\partial x_i})_m$ form a basis for $T_m M$. Thus, for any vector field X whose domain W meets U,

$$X = \sum_{i=1}^{n} f_i \frac{\partial}{\partial x_i} \tag{3.2-1}$$

where the f_i are differentiable functions defined on $U \cap W$.

Let c be a curve in M with domain D. If $\frac{\partial}{\partial t}$ is the vector field corresponding to the identity chart on $\mathbb{R}$,

$$\dot{c} = c_* \circ \frac{\partial}{\partial t} : \mathbb{R} \to TM \tag{3.2-2}$$

is a curve in TM with domain D. Notice $\dot{c}$ satisfies the equation $\pi \circ \dot{c} = c$. If $X \in \mathcal{X}(M)$,

$$\dot{c} = X \circ c \tag{3.2-3}$$

is a differential equation of the first order on M. Notice if X satisfies (3.2-3), it can be thought of as being the tangent vector to the curve in the classical sense.

Definition 3.2-1

A curve c in M satisfying (3.2-3) and the initial condition $c(0) = m$ for a specific $m \in M$ is called an integral curve starting from m, of the vector field X.

To see that (3.2-3) actually is an autonomous system of differential equations, choose a chart of M whose domain U meets c(D) and let $C = x \circ c$. It follows from (2.2-1) that c is an integral curve of X when,

$$\frac{dC_i}{dt} = f_i(c_1,\ldots,c_n) \quad (i=1,\ldots,n) \tag{3.2-4}$$

using (3.2-1).

The equations (3.2-4) are a set of autonomous equations and from 3.1.2), with slight amendments, from a local dynamical system, but do not, in general, form a global dynamical system.

Notice, in particular, the geodesic equations are derivable from a vector field and constitute a local dynamical system.

It follows from the definition of a vector field and Theorem 3.1-2, that,

Proposition 3.2-1

If the integral curves c_1, c_2 of $X \in \mathcal{X}(M)$ with domains D_1, D_2 respectively, both start from the same point of M, they must coincide on $D_1 \cap D_2$. □

It follows that the union of the domains of all the integral curves of X which start from $m \in M$ is an open interval D(m) of $\mathbb{R}$ on which is defined a 'maximal' integral curve γ_m starting from m.

Definition 3.2-2

A vector field is complete if $D(m) = \mathbb{R}$ for each point $m \in M$.

Example 3.2-1) A System of Differential Equations on the Sphere

The sphere, S^2, is the set of zeros of the function $f: \mathbb{R}^3 \to \mathbb{R}$ defined by,

$$f(z_1,z_2,z_3) = z_1^2 + z_2^2 + z_3^2 - 1. \tag{3.2-5}$$

An atlas for S^2 can be obtained by sterographic projection of S^2 from the points $(0,0,1)$ and $(0,0,-1)$ onto $z_3 = 0$.

Suppose that U,U' are two punctured spheres obtained from S^2 by omitting these two points respectively. The functions $x: S^2 \to \mathbb{R}^2$, $y: S^2 \to \mathbb{R}^2$ defined on U,U' by,

$$x_i z = \frac{z_i}{1-z_3}, \quad y_i z = \frac{z_i}{1+z_3} \tag{3.2-6}$$

can be shown to be charts for S^2.[7] The change of co-ordinates defined on $U \cap U'$ is a diffeomorphism defined on $\mathbb{R}^2 - \{0\}$ by,

$$y_i = \frac{x_i}{(x_1)^2 + (x_2)^2} \tag{3.2-7}$$

Thus, the charts x,y from a C^∞ atlas for S^2, called the sterographic atlas of S^2 into $\mathbb{R}^2$.

In terms of the charts y, a vector field on U' is given by,

$$\frac{1}{2}\left\{9y_1y_2 + y_1 + y_2 + (y_2)^2 \frac{(3(y_2)^2-8y_1)-(y_1)^2(3(y_1)^2+16y_2)}{(y_1)^2+(y_2)^2}\right\}\frac{\partial}{\partial y_1}$$
$$+\frac{1}{2}\left\{3y_1 - y_2 - 3y_1y_2 + \frac{(y_2)^2(9(y_2)^2-16y_1)-(y_1)^2(9(y_1)^2-8y_2)}{(y_1)^2+(y_2)^2}\right\}\frac{\partial}{\partial y_2} \tag{3.2-8}$$

Using the chain rule for partial derivatives, it is easy to show that the above vector field agrees on $U \cap U'$ with the linear vector field,

$$\frac{1}{4}(3 - 2x_1 + 2x_2)\frac{\partial}{\partial x_1} + \frac{1}{4}(6x_1 + 2x_2 - 9)\frac{\partial}{\partial x_2} \qquad (3.2\text{-}9)$$

As (3.2-8), (3.2-9) agree on the intersection of their domains, together they define a vector field on S^2.

As (3.2-9) is more tractable than (3.2-8), its integral curves are found. Such a curve c is defined by $x \circ c = (C_1, C_2)$ where,

$$\begin{aligned} \frac{dC_1}{dt} &= \frac{1}{4}(3 - 2C_1 + 2C_2), \\ \frac{dC_2}{dt} &= \frac{1}{4}(-9 + 2C_2 + 6C_1) \end{aligned} \qquad (3.2\text{-}10)$$

It follows that if the initial condition is $x \circ c(0) = (0,0)$, ie., $c(0) = (0,0,-1)$,

$$x \circ c = \left(\frac{3}{2} - \frac{3}{8}e^{t} - \frac{9}{8}e^{-t}, -\frac{9}{4}\sinh t\right) \qquad (3.2\text{-}11)$$

In general, $x \circ c = \left(\frac{3}{2} + ae^{t} + 3ae^{-t}, 6a\sinh t\right)$ (3.2-12)

where a is uniquely determined by the condition that $c(0)$ is any chosen point of $U \cap U'$. Notice the solution of (3.2-10) is defined for all t and so the vector field is complete.

There is a condition which assures us that the system (3.2-4) is dynamical:-

Proposition 3.2-2

Every vector field on a compact manifold is complete. □

Any complete vector field gives rise to a dynamical system of differential equations, by definition. The function Φ, which when associated with M is a dynamical system in the sense of definition 3.1-1, is known as the flow of the vector field:-

Definition 3.2-3

Given any vector field X on a manifold M there exists a maximal integral curve γ_m, domain D(m), starting from any point $m \in M$. If $\mathcal{S}$ is the set of points (m,s) of M x $\mathbb{R}$ such that $s \in D(m)$, there is a function Φ: M x $\mathbb{R} \to M$ with domain $\mathcal{S}$ defined by

$$\Phi(m,s) = \gamma_m(s) \tag{3.2-13}$$

Φ is called the flow of X.

It can be shown that Φ is differentiable[8] and it is clear from the definition that it satisfies the axioms of definition 3.1-1. Thus, if X is complete, (M,Φ) is a dynamical system.

3.3) Phase Portraits and Critical Elements

In most cases, we cannot hope to determine the maximal integral curve of a given vector field, starting from a given point of the manifold. However, it may be possible to get some qualitative picture of the integral curves of X on the manifold. For example, it may be possible to say whether or not the vector field has a periodic integral curve.

The first step in the qualitative theory is to define the concepts of an orbit of a vector field and the phase portrait. Phase portraits on a given manifold are numerous and complex, so that a complete classification is difficult. However, some progress has been made for the simplest manifolds of dimension two.[9]

A useful concept in the qualitative theory of vector fields is that of a critical point. This is a point of the manifold where the vector field vanishes identically. These points are defined and studied in this section, together with the periodic orbits.

Definition 3.3-1

If $X \in \mathcal{X}(M)$, an orbit of X is the image $\gamma_m(D)$ of a maximal

integral curve γ_m: $D \to M$ of X. The set P(X) of all orbits of X is known as the phase portrait of X.

It is possible for two distinct vector fields to have the same phase portrait.

Example

The equations $\begin{cases} \dot{x} = f_1(x,y) \\ \dot{y} = f_2(x,y) \end{cases}$ and $\begin{cases} \dot{x} = f_1(x,y)g(x,y) \\ \dot{y} = f_2(x,y)g(x,y) \end{cases}$

lead to the same phase portraits on the plane $\mathbb{R}^2$. Thus, it is possible for a non-dynamical system to have the same phase portrait as a dynamical system.

It is also possible for two distinct vector fields to possess phase portraits with the same topological properties. As it is these properties which the qualatative theory is concerned with, the phase portraits can be thought of as being equivalent.

Definition 3.3-2

Let M be a manifold and $X,Y \in \mathcal{X}(M)$. P(X), P(Y) are equivalent phase portraits if and only if there is a homeomorphism $h: M \to M$ such that $\gamma_m \in P(X)$ if and only if $h(\gamma_m) \in P(Y)$.

The above relation is clearly an equivalence relation in the algebraic sense.

Definition 3.3-3

Let $X \in \mathcal{X}(M)$. A point $m \in M$ is a critical point of X if and only if $X(m) = 0$.

As any vector field has integral curves given by equation (3.2-4), any property of a vector field can be thought of as being a property of (3.2-4). In particular, critical points of an autonomous set of differential equations are points at which the right hand sides of the equations vanish.

Definition 3.3-4

A closed orbit γ of $X \in \mathcal{X}(M)$ is an orbit on which X does not vanish and which is compact. A critical element of X is a set $\{m\}$ where $m \in M$ is a critical point or a closed orbit of X. Γ_X denotes the union of critical elements of X.

The following theorem shows that a closed orbit must be a periodic orbit in the classical sense.

Proposition 3.3-1[10]

Let $X \in \mathcal{X}(m)$ with flow $\Phi: \mathcal{S} \subset M \times \mathbb{R} \to M$. If $\gamma \subset M$ is a closed orbit of X, then $\gamma \times \mathbb{R} \subset \mathcal{S}$, γ is a connected one-dimensional submanifold and there exists a positive real number $\tau(\gamma)$, such that for any $m \in \gamma$, $\Phi(m,t) = m$ if and only if $t = k\tau(\gamma)$, for some $k \in \mathbb{Z}$. (The number $\tau(\gamma)$ is the period of γ).

If the period of a critical point of X is defined to be zero, a critical point may be thought of as the degenerate case of a periodic orbit.

The beauty of autonomous systems arising from a vector field as opposed to just an autonomous set of equations is brought out by the following theorems, the results following essentially by local uniqueness of solutions.

Theorem 3.3-2

No orbit $\gamma_m(t)$ of a vector field X enters a critical point for a finite value of t and any point $m \in M$ which is not a critical point has at most one orbit passing through it.

□

Theorem 3.3-3

An orbit $\gamma_m(t)$ of a vector field X that passes through at least one ordinary point cannot cross itself, unless it is a closed orbit. (A point which is not critical being ordinary).

□

Notice the above results hold, whether X is complete or not.

3.4) Systems of non-autonomous Equations

It has been shown that the autonomous system (3.1-10) can be thought of as being derived from a vector field acting on an n-dimensional manifold.

Consider now the non-autonomous generalization of (3.1-10), that is, where f(x) can also depend explicitly on t,

$$\dot{x} = f(x,t) \tag{3.4-1}$$

Clearly, the system (3.4-1), which is n-dimensional, can be re-written as an autonomous system of (n+1) equations: Let $y = (x_1, x_2, \ldots, x_n, t)$ and so (3.4-1) becomes,

$$\frac{dy}{dt_1} = f_1(y) \tag{3.4-2}$$

where now one of the equations of the system is $\frac{dt}{dt_1} = 1$.

Thus, any non-autonomous n-dimensional system can be thought of as having been derived from a vector field acting on an (n+1)-dimensional manifold.

It may happen that (3.4-2) constitutes a dynamical system, so that the solution y can be extended for all values of t_1. Then the solution of (3.4-1) can also be extended for all values of t. If (3.4-1) also has uniqueness of solutions, it is dynamical in some sense, although, being non-autonomous it may not lead to a dynamical system in the sense of definition 3.1-1.

Definition 3.4-1

Any non-autonomous system which has uniqueness of solutions to the initial value problem and which has global existence of solutions is said to be quasi-dynamical.

The results on uniqueness of solutions given in section 3.1)

hold for (3.4-1) independent of t. Thus, if f(x,t) is continuous with respect to x and all the partial derivatives, to prove (3.4-1) is quasi-dynamical, it is sufficient to show that it has global existence of solutions.

Notice that (3.4-2) has no critical points in the sense as defined in the last section, due to the presence of the equation $\frac{dt}{dt_1} = 1$. However, (3.4-1) may have points of equilibrium in the n-dimensional x space. Thus, as extended definition of a critical point is needed to cater for non-autonomous systems.

Definition 3.4-2

A critical point of (3.4-1) is any point $x_o \in \mathbb{R}^n$, which is a point of equilibrium. That is, an integral curve starting at x_o remains there for all values of t and any integral curve not starting at x_o cannot reach x_o in a finite value of t.

In contrast to the autonomous case, a critical point x_o of (3.4-1) may not simply be a point at which the right hand sides of (3.4-1) vanish.

Example 3.4-1

Consider the one-dimensional, non-autonomous equation,

$$\dot{x} = tx^2$$

The solution is $x = \frac{-2}{t^2+K}$ (K constant) and so is not quasi-dynamical. There is a critical point at $x = 0$, but not at $t = 0$.

Example 3.4-2

Consider the one-dimensional, non-autonomous equation,

$$\dot{x} = R(t)x \qquad (3.4\text{-}3)$$

The equation is linear and, as in the autonomous case, can have at most one critical point. Clearly, (3.4-3) has general solution

$$x = \exp\left\{\int R(t)dt + K\right\}.$$

(a) If $R(t) = t^n$, $n \in \mathbb{R}_{\geq 0}$, the equation is quasi-dynamical and $x = 0$ is the critical point.

(b) If $R(t) = \frac{a}{t}$, 'a' constant, the solution of (3.4-3) is $x = Kt^a$. If $a > 0$, the system is quasi-dynamical and there is no critical point. If $a < 0$, the system is not quasi-dynamical but $x = 0$ is a critical point.

(c) If $R(t) = \frac{a}{t^n}$, $n > 0$, $n \neq 1$, for $0 < n < 1$, the system is quasi-dynamical for certain values of n but not others. If $n > 1$, the system is not quasi-dynamical and there is no critical point.

For the purposes of this work, only linear non-autonomous equations need be considered. In this case, the system (3.4-1) can be written as,

$$\dot{x} = A(t)x \qquad (3.4\text{-}4)$$

where $A(t)$ is an $(n \times n)$ matrix function. (Taking the only possible critical point to be at the origin, without loss of generality - any linear system can be reduced to the form (3.4-4) by a transformation of variables).

Clearly, in order for $x = 0$ to be the critical point of (3.4-4), the right hand sides must be zero when $x = 0$ for all finite t.

Lemma 3.4-1

The system (3.4-4) has a critical point at $x = 0$ if the matrix $A(t)$ is continuous for all values of t. □

Notice, however, that (3.4-4) may have a critical point if $A(t)$ is discontinuous when it is not quasi-dynamical.

Although any linear autonomous equation must be dynamical (using

theorem 3.1-1, theorem 3.1-2), (3.4-4) is not necessarily quasi-dynamical, as the corresponding (n+1)-dimensional autonomous system is not necessarily linear. Whether or not (3.4-4) is quasi-dynamical depends on a large extent on the properties of A(t) as shown in the previous example. Notice it may be quasi-dynamical even when A(t) is discontinuous.

3.5) Properties of Orbits in the Neighbourhood of a Simple Critical Point

For the remainder of this chapter, only autonomous systems are considered. An important aspect of their qualatative behaviour is the nature of orbits in the neighbourhood of a critical point.

Consider a critical point m of a vector field X acting on an n-dimensional manifold M. A neighbourhood of U of m can be made small enough so that there exists a chart y: $U \to \mathbb{R}^n$. Then the maximal integral curve of X is given in terms of the autonomous system,

$$\dot{x} = f(x) \qquad (3.5\text{-}1)$$

where $y \circ c = x$, from section 3.2).

It is assumed that the critical point in question is situated at the origin, that is, $f(0) = 0$, without loss of generality. It is always possible to make a linear transformation of (3.5-1) which transforms any critical point to the origin. If f(x) contains linear terms (3.5-1) may be expressed in the form,

$$\dot{x} = Ax + g(x) \qquad (3.5\text{-}2)$$

where A is an n x n constant matrix and g(x) contains only terms of order higher than the first. The linearized system of (3.5-1) is then,

$$\dot{x} = Ax \qquad (3.5\text{-}3)$$

Definition 3.5-1

The eigenvalues of X at the point $m \in M$ are defined to be the

eigenvalues of the matrix A. The characteristic multipliers of X are the exponentials of its eigenvalues at any point $m \in M$.

These concepts may be defined without going into local coordinates by treating the eigenvalues of X as eigenvalues of a certain linear transformation mapping $T_m M \to T_m M$. Also, the above concepts may be generalized to critical elements and so lead to eigenvalues of X at a closed orbit γ.[11]

Definition 3.5-2

If $X \in \mathfrak{X}(M)$ and m is a critical point of X, m is a simple critical point if and only if none of the eigenvalues of X at m are zero.

The characteristic multipliers characterize the flow of the linear approximation near the critical point. Local qualitative behaviour near a simple critical point is especially elementary and it is these types of critical points which are considered in this section.

Only the two-dimensional case will be considered. By hypothesis, let the matrix A have eigenvalues $\lambda, \mu \neq 0$.

It turns out that there are only a few possibilities for the behaviour of integral curves of (3.5-1) in a neighbourhood of (0,0), (assuming (3.5-1) to be two-dimensional).

Definition 3.5-3

Let $\gamma = (\gamma_1, \gamma_2)$ be a solution of (3.5-1). If there exists δ, $0 \leq \delta \leq r_o$ such that for any orbit γ of (3.5-1) which has at least one point in $0 < r < \delta$ (r being the normal plane polar radius), the solution exists over a t half line, and if $(\gamma_1, \gamma_2) \to (0,0)$ as $t \to \infty$, or as $t \to -\infty$, then the origin is called an attractor for (3.5-1).

Definition 3.5-4

The origin is a node of (3.5-1) if it is an attractor for which all

orbits arrive at the origin in a definite direction. A node is proper if every half line through the origin is tangent to some orbit there.

Definition 3.5-5

The origin is called a focus for (3.5-1) if it is an attractor such that $|w(t)| \to \infty$ as $t \to \infty$ or $t \to -\infty$ where $w(t) = \tan^{-1} \frac{\gamma_2}{\gamma_1}$. (A focus is sometimes called a spiral point).

Definition 3.5-6

If there exists a sequence of periodic orbits $\{\gamma_n\}$ of (3.5-1) each of which contains all later orbits and the origin in its interior and such that $\gamma_n \to (0,0)$ as $n \to \infty$, then the origin is a centre of (3.5-1).

Definition 3.5-7

Suppose the two orbits on one eigenvector of A tend to the origin as $t \to \infty$ and the two orbits on the second eigenvector tend to the origin as $t \to -\infty$. Also suppose that, with increasing t, every point tracing an orbit first moves closer to the origin and after reaching a minimal distance, moves away from the origin with its orbit asymptotic to the second eigenvector. Then the origin is a saddle point. The eigenvectors are the separatrices of the saddle point and will be distorted to curves in the non-linear case.

In the case of the linear system (3.5-3), the nature of the critical point is determined completely by the eigenvalues λ, μ:-

Lemma 3.5-1

Consider the linear system (3.5-4), with critical point at the origin.

(i) If the eigenvalues of A are real, district and of the same sign, the origin is a node.

(ii) If the eigenvalues of A are of opposite sign, the origin is a saddle point.

(iii) If the eigenvalues of A are a complex conjugate pair, but not purely imaginary, the origin is a focus.

(iv) If the eigenvalues of A are purely imaginary, the origin is a centre.

(v) If eigenvalues of A are real and equal, the origin is a node. □

Consider now the non-linear system (3.5-2), (considered to be two-dimensional), where it is assumed that $g(x_1x_2) = O(r)$ as $r \to 0^+$ where $r = \sqrt{(x_1^2 + x_2^2)}$. This guarantees that the perturbations rend to the origin 'faster' than the linear terms in (3.5-2). In general, (3.5-2) will have more than one critical point. It is assumed that each one is isolated. That is, there exists an open neighbourhood about each critical point which contains no other critical points.

Theorem 3.5-2 [12]

If the origin is an attractor for (3.5-3), it is one for (3.5-2). □

Theorem 3.5-3 [12]

If the origin is a focus for (3.5-3), it is also one for the non-linear system (3.5-2). □

Although attractors of (3.5-3) are attractors of (3.5-2), it is not, in general, true that a node of (3.5-3) is a node of (3.5-2).

Theorem 3.5-4 [12]

If, in (3.5-2), $g(x_1,x_2) = O(r^{1+\varepsilon})$ as $r \to 0$ for some $\varepsilon > 0$, then if the origin is a proper node of (3.5-3), it is one for (3.5-2) also. □

If the origin is a saddle point of (3.5-3) it is also one for

(3.5-2). More explicitly, suppose that (3.5-2) is assumed to be in the canonical form,

$$\begin{cases} \dot{x}_1 = \lambda x_1 + g_1(x_1, x_2) \\ \dot{x}_2 = \mu x_2 + g_2(x_1, x_2) \end{cases} \tag{3.5-4}$$

Then,

Theorem 3.5-5[12]

(i) There exists at least one orbit tending to the origin at each of the angles 0 and π.

(ii) If $\frac{\partial g_1}{\partial x_2}$ and $\frac{\partial g_2}{\partial x_1}$ exist and are continuous on $0 \leq r \leq r_o$, there exists exactly one orbit tending to the origin at each of the angles 0 and π. Any orbit starting sufficiently near either of these orbits in the neighbourhood of the origin tends away from them as $t \to \infty$.

□

Examples of the use of the above theorems to specific non-linear equations abound.[13]

Unfortunately, if the origin is a centre for (3.5-3), it need not be one for (3.5-2). In fact, it is either a centre, a focus, or enclosed in a limit cycle, a special type of closed orbit (see definition 3.5-10).

There are several procedures which give criteria for distinguishing between a centre and a focus. In the case in which f(x) in (3.5-1) is analytic, Poincaré introduced an alogorithm, requiring only the solution of algebraic equations.[14] Also, in this case, there is one condition, again due to Poincaré, which assures us that the origin is a centre.

Lemma 3.5-6[14]

If the vector field giving rise to the system (3.5-2) is symmetric with respect to either the x_1 axis or x_2 axis, (that is, the system does not change if x_i is replaced by $-x_i$, t by -t, $i = 1$ or 2), then the origin is a centre. □

There are several theorems which give criteria for the non-existence of periodic orbits, the most useful being Dulac's test, which includes Bendixson's test as a special case. (In Bendixson's test, the function ψ defined below is identically equal to one).

Theorem 3.5-7 Dulac's Test.[15]

Assume that (3.5-1) has uniqueness of solutions. Let G be a simply connected subregion of the domain of definition of (3.5-1). Let $\psi(x_1,x_2)$ be a continuously differentiable function defined in G such that the function

$$\Phi = \frac{\partial}{\partial x_1}(\psi f_1(x_1,x_2)) + \frac{\partial}{\partial x_2}(\psi f_2(x_1,x_2)) \text{ does not change sign.}$$

Then there are no simple closed curves in G which are unions of orbits of system (3.5-1). □

Example 3.5-1

Consider Liénards equation, which plays an important role in the theory of nonlinear oscillations,

$$\ddot{x}_1 + F(x_1)\dot{x}_1 + G(x_1) = 0 \qquad (3.5\text{-}5)$$

Putting $\dot{x}_1 = x_2$, the equation is transformed into the form (3.5-1) with $f_1(x_1,x_2) = x_2$ and $f_2(x_1,x_2) = -F(x_1)x_2 - G(x_1)$.

Clearly if $F(x_1)$, $G(x_1)$ are continuous, the partial derivatives of

f_1, f_2 are continuous. Also, $\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} = - F(x_1)$. Thus, if $F(x_1)$ does not change sign within $a \leq x_1 \leq b$ then there are no closed orbits in this strip.

There are various theorems concerning the existence of periodic solutions of second order equations of the Liénard type [16]. However, these are not directly applicable to the type of equations considered in this work.

As stated earlier, if the origin is a critical point of (3.5-2), it may be enclosed in a limit-cycle. This is a particular type of periodic orbit and a rigorous definition is now given in terms of limit sets.

Definition 3.5-8

Let $X \in \mathcal{X}(M)$ have flow $\Phi: \mathcal{S} \subset M \times \mathbb{R} \to M$. A point $m \in M$ is σ complete ($\sigma = +$ or $-$, or $\pm$) if and only if $\mathcal{S} \cap \left\{ \{m\} \times \mathbb{R} \right\}$ contains all $t > 0$, $t < 0$ or $t \in \mathbb{R}$, respectively.

Thus, if all points $m \in M$ are $\pm$ complete, (with respect to $X \in \mathcal{X}(m)$), X is complete, and (M, Φ) is a dynamical system.

Definition 3.5-9

Let $X \in \mathcal{X}(M)$. If m is σ complete, (with respect to X), the limit set $w^\sigma(m)$ is defined as the set of points $m_o \in M$ for which there exists a sequence $\{t_n\}$ with $t_n \to \sigma\infty$ and $\Phi(m, t_n) \to m_o$.

Let Ω^σ_X denote the set of all w^σ limit sets, with respect to X.

It can be shown that if m is not σ complete, $w^\sigma(m)$ is empty. Limit sets will be considered again in section 3.7).

Definition 3.5-10

A closed orbit of X is a limit-cycle if it is either the w^+

or w^- limit set of another orbit of $X \in \mathcal{H}(m)$.

By geometric means, the following theorem on the existence of limit-cycles may be established.

Theorem 3.5-8 [17]

If there exist two constants, r_o, r_1 with $r_o < r_1$ and such that $F(x_1,x_2) = x_1 f_1 + x_2 f_2 \geq 0$ for $x_1^2 + x_2^2 = r_o^2$ and $F \leq 0$ for $x_1^2 + x_2^2 = r_1^2$ and if there are no critical points of (3.5-1) in the region Γ: $r_o^2 \leq x_1^2 + x_2^2 \leq r_1^2$, then Γ contains a limit-cycle of (3.5-1). Similarly if F is negative on the inner circle and positive on the outer circle.

□

Example 3.5-2

Consider the two-dimensional system,

$$\begin{cases} \dot{x}_1 = x_1(x_1^2 - 1 + x_2(x_2 + 2x_1 - 1)) - x_2(x_2 - 1)^2 \\ \dot{x}_2 = x_1(x_1 - 2x_1^2 + 1 + x_2(x_2 + x_1 - 2)) + x_2(x_2^2 - 1) \end{cases}$$

In this case, a suitable function is, $F = (x_1^2 + x_2^2)(x_1^2 + x_2^2 - 1)$ and hence $F < 0$ for $x_1^2 + x_2^2 = 1 - \varepsilon$ and $F > 0$ if $x_1^2 + x_2^2 = 1 + \varepsilon$.

$\therefore$ $x_1^2 + x_2^2 = 1$ is a limit-cycle of the system.

Often, given a system, it is useful to make a transformation of the coordinates to determine any possible limit-cycles.

Example 3.5-3

Consider the two dimensional system in the plane $\mathbb{R}^2$ given in

terms of cartesian coordinates,

$$\begin{cases} \dot{x}_1 = x_1(x_1^2 + x_2^2 - 1)^2 - x_2 \\ \dot{x}_2 = x_2(x_1^2 + x_2^2 - 1)^2 + x_1 \end{cases} \qquad (3.5\text{-}6)$$

In terms of polar coordinates, $x_1 = r\cos\theta$ $x_2 = r\sin\theta$, the system (3.5-7) becomes,

$$\begin{cases} \dot{r} = r(r^2 - 1)^2 \\ \dot{\theta} = 1. \end{cases} \qquad (3.5\text{-}7)$$

The general solution to system (3.5-7) is,

$$\begin{cases} \left\{ 1 + \dfrac{1}{r^2-1} \right\} e^{-\frac{1}{r^2-1}} = Ce^{2t} & r > 1 \\ \left\{ -1 - \dfrac{1}{r^2-1} \right\} e^{-\frac{1}{r^2-1}} = Ce^{2t} & r < 1 \\ r = 1 \end{cases}$$

(C an arbitrary constant).

As $t \to \pm\infty$, $r \to 1$ and $r = 1$ is a limit-cycle of the system (3.5-6). Thus, the critical point at the origin is enclosed in a limit-cycle.

Further methods of differentiating between centres and foci and nodes and foci are given in section 3.10) on bifurcation theory.

The idea of characterizing critical points can be extended to the general case where (3.5-3) is an n-dimensional system. The

nature of the orbits in a neighbourhood of the critical point in question is again determined by the eigenvalues of the matrix A. Various characterizations of the critical point can be made which are generalizations of the two-dimensional definitions. The three dimensional case is treated fully in[18], where the authors define such notions as nodes, saddle-nodes, stellar nodes, foci and centres in three dimensions;[19] goes a stage further and treats the n-dimensional case.

The task of determining the nature of the critical points of (3.5-2) based on the nature of the corresponding critical points of the linear system (3.5-3) is much more formidable in the n-dimensional case than in the two-dimensional case. The more important theorems are instead concerned with the stability of the various critical points of (3.5-2).

The proof of existence of periodic solutions in the n-dimensional case is far more complicated than the two-dimensional case. Some of the methods use the idea of a Lyapunov rotation function and are a generalization of Poincaré's method of contact curves.[20]

3.6) Multiple Critical Points

3.6.1) Multiple Critical Points of the First and Second Kind

In the last section the character of various types of simple critical points was investigated. In this section critical points of two-dimensional non-linear systems of the type (3.5-2) are considered, where one or both of the eigenvalues of the matrix could be zero. Only the elementary case is considered, that is, where the expansions of the right hand sides of the system contain at least one linear term. It is also assumed, once again, that the critical points are isolated. As the section deals exclusively with two-dimensional systems, the notation is changed slightly. The two dimensional form of (3.5-1), by hypothesis is,

$$\begin{cases} \dot{x} = f(x,y) \\ \dot{y} = g(x,y) \end{cases} \qquad (3.6\text{-}1)$$

It is assumed, as usual, without loss of generality, that (0,0) is the critical point in question. By hypothesis, (3.6-1) can be written in the form,

$$\begin{cases} \dot{x} = ax + by + f_1(x,y) \\ \dot{y} = cx + dy + g_1(x,y) \end{cases} \qquad (3.6\text{-}2)$$

where $a,b,c,d \in \mathbb{R}$ are not all zero. Clearly, $ad - bc = 0$, if at least one of the eigenvalues of $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ are zero. Both eigenvalues will be zero if $a + d = 0$.

$$\text{Let } \sigma = \frac{\partial f}{\partial x}(0,0) + \frac{\partial g}{\partial x}(0,0) = a + d \qquad (3.6\text{-}3)$$

(using Taylor's series).

Definition 3.6 -1

If $\sigma \neq 0$, the critical point at the origin is a multiple critical point of the first kind. If $\sigma = 0$, the origin is a multiple critical point of the second kind.

3.6.2) Multiple Critical Points of the First Kind

By hypothesis, $\sigma = a + d \neq 0$, $ad - bc = 0$, $f_1(0,0) = g_1(0,0) = 0$ in (3.6-2). It will also be assumed that $f_1(x,y)$, $g_1(x,y)$ are analytic in a neighbourhood of the origin (the series expansions involving only terms of at least the second order).

Under these assumptions, there exists a nonsingular linear

transformation reducing the system to the following form,

$$\begin{cases} \dot{x} = f_2(x,y) \\ \dot{y} = g_2(x,y) + y \end{cases} \tag{3.6-4}$$

(where the labels of the dependent variables have been changed). Thus, equations of this form are considered without loss of generality.

Consider the equation,

$$y + g_2(x,y) = 0 \tag{3.6-5}$$

By the implicit function theorem, the equation has a solution $y = \phi(x)$ in a small neighbourhood of the origin, where $\phi(x)$ is analytic and $\phi(0) = 0$, $\phi'(0) = 0$.

Using the above solution to (3.6-5), the possible phase portraits in the neighbourhood of the origin can be described, as in the next theorem.

Theorem 3.6-1[21]

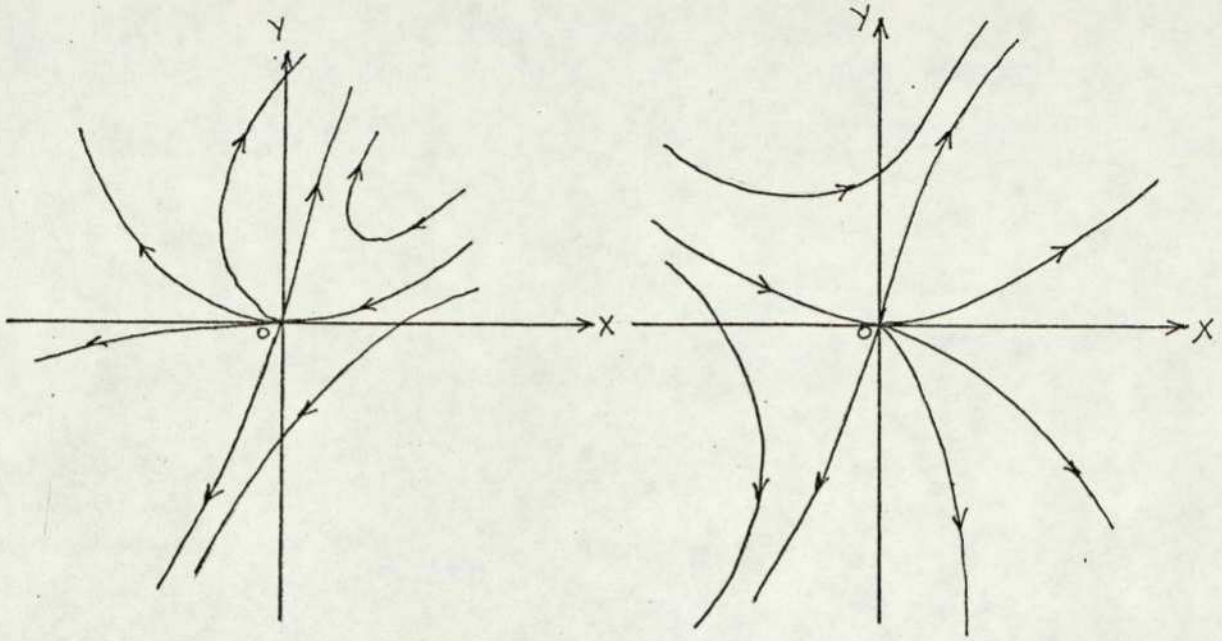
Let (0,0) be an isolated critical point of (3.6-4). Then, let $y = \phi(x)$ be the solution to $y + g_2(x,y) = 0$ in a neighbourhood of the origin and let $\psi(x) = f_2(x,\phi(x))$. $\psi(x)$ can be expressed in the form $\psi(x) = \Delta_m x^m + \ldots$ where $m \geq 2$ and $\Delta_m \neq 0$.

(i) If m is odd and $\Delta_m > 0$, (0,0) is a node

(ii) If m is odd and $\Delta_m < 0$, (0,0) is a saddle point, two of whose separatrices tend to (0,0) in the directions of 0 and π; the other two in the directions $\frac{\pi}{2}$, $\frac{3\pi}{2}$.

(iii) If m is even, (0,0) is a 'saddle-node' (see the figures below).

□



(a) $\Delta_m < 0$ (b) $\Delta_m > 0$.

Figure 3.6-1:- 'Saddle-Node'

3.6.3) Multiple Critical Points of the Second Kind

In this case, the system under consideration is again (3.6-2) where f_2, g_2 satisfy the same conditons as in the previous sub-section and this time, $ad - bc = 0$, $\sigma = a + d = 0$, $|a| + |b| + |c| + |d| \neq 0$. Because of the existence of a non-singular transformation reducing (3.6-2) to the following form, only this special case needs to be considered:-

$$\begin{cases} \dot{x} = y + f_2(x,y) \\ \dot{y} = g_2(x,y) \end{cases} \qquad (3.6\text{-}6)$$

(3.6-6) can be simplified still further by the use of the transformation,

$$\xi = x \ , \ \eta = y + f_2(x,y) \qquad (3.6\text{-}7)$$

where the inverse of the transformation is of the form,

$$x = \xi \ , \ y = h(\xi,\eta) \qquad (3.6\text{-}8)$$

to get,

$$\begin{cases} \dot{\xi} = \eta \\ \dot{\eta} = g_2(\xi, h(\xi,\eta)) + f'_{2x}(\xi,h(\xi,\eta))\eta + f'_{2y}(\xi, h(\xi,\eta))g_2(\xi,h(\xi,\eta)) \end{cases}$$

(3.6 -9)

Clearly, if (0,0) is a critical point of (3.6-6), it is one for (3.6-9). It can also be shown that the paths of the system (3.6-6) passing through a sufficiently small neighbourhood of the origin correspond in one-to-one fashion to the paths of the system (3.6-9) near the transformed origin.

Thus, only the system,

$$\begin{cases} \dot{x} = y \\ \dot{y} = g_3(x,y) \end{cases} \qquad (3.6\text{-}10)$$

need be considered, which is the form of (3.6-9).

Notice this is precisely the form obtained when any analytic second order equation is expressed as a system of first order equations.

By hypothesis the function $g_3(x,y)$ can be expressed as,

$$g_3(x,y) = a_k x^k \left\{1 + A(x)\right\} + b_n x^n y \left\{1 + B(x)\right\} + y^2 C(x,y) \qquad (3.6\text{-}11)$$

where $A(x)$, $B(x)$, $C(x,y)$ are analytic in a neighbourhood of the origin, $A(0) = B(0) = 0$, $k \geq 2$, $a_k \neq 0$. (The coefficient b_n may vanish: if $b_n \neq 0$, $n \geq 1$).

Theorem 3.6-2[21]

Let (0,0) be an isolated critical point of (3.6-10)

(a) If the number k in (3.6-11) is odd, let $k = 2m + 1$ $(m \geq 1)$ and $\lambda = b_n^2 + 4(m + 1)a_{2m+1}$.

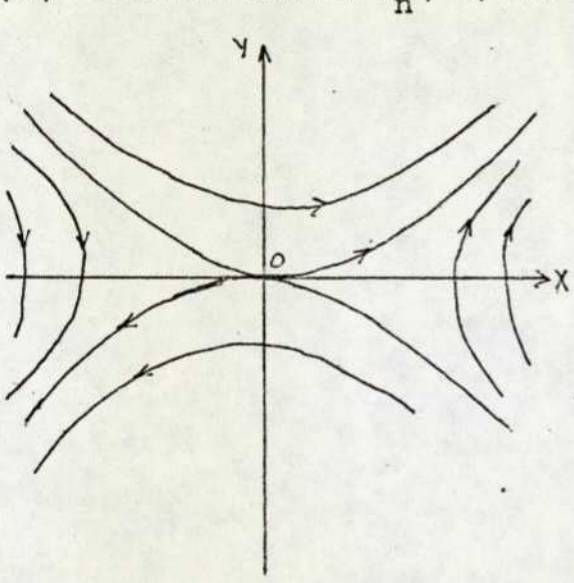
(i) If $a_{2m+1} = a_k > 0$, (0,0) is a saddle point, of a degenerate type (see figure (a) below).

(ii) If $a_k < 0$, (0,0) is (I) a focus or centre if $b_n = 0$, or if $b_n \neq 0$ and $n > m$, or if $b_n \neq 0$, $n = m$ and $\lambda < 0$; (II) a node if $b_n \neq 0$, n is even and $n < m$, or if $b_n \neq 0$, n is even, $n = m$ and $\lambda \geq 0$; (III) a critical point with elliptic region (see figure (b) below), if $b_n \neq 0$, n is odd and $n < m$, or if $b_n \neq 0$, n is odd, $n = m$, $\lambda \geq 0$.

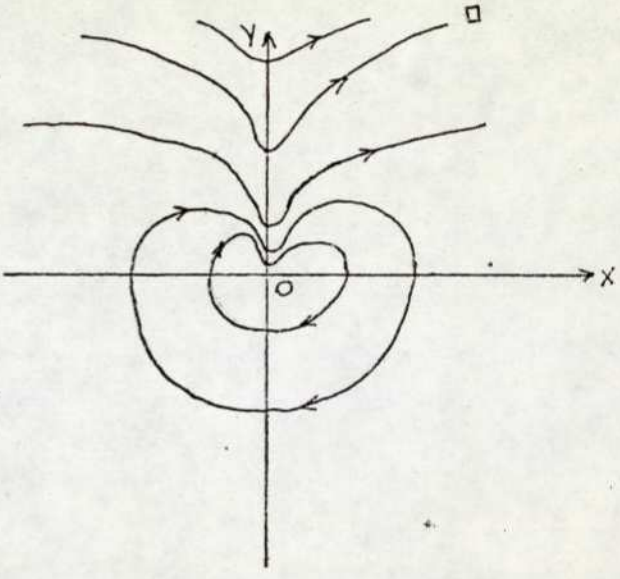
(b) If the number k is (3.6-11) is even, let $k = 2m$ $(m \geq 1)$.

(i) (0,0) is a degenerate critical point if $b_n = 0$ or $b_n \neq 0$ and $n \geq m$. (see figure (c))

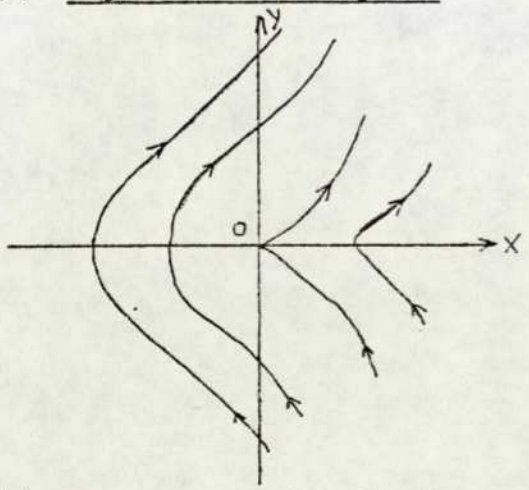
(ii) A saddle-node if $b_n \neq 0$, and $n < m$ (see figure (d)).



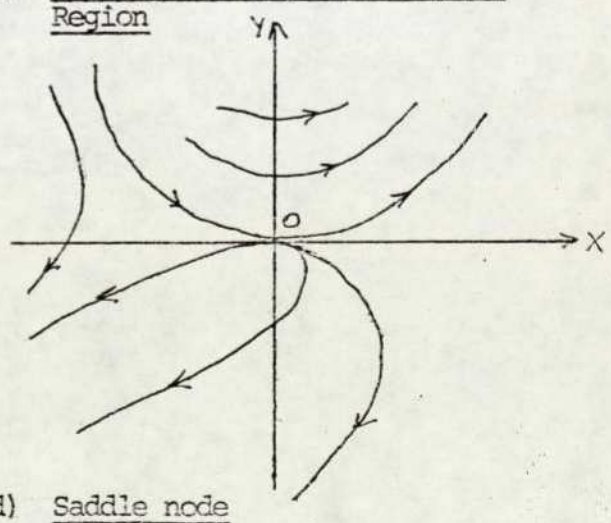
(a) Degenerate Saddle point



(b) Critical Point with Elliptic Region



(c) Degenerate Critical Point ($a_{2m} > 0$)



(d) Saddle node ($b_n > 0$, $a_{2m} < 0$)

Figure 3.6-2

3.6.4) Example

Consider the system,

$$\begin{cases} \dot{x} = y + \alpha x \\ \dot{y} = x^2 + y^2 \end{cases} \qquad (3.6\text{-}12)$$

where α is a constant. The origin is an isolated critical point. It is a multiple critical point of the first kind if $\alpha \neq 0$ and a multiple critical point of the second kind if $\alpha = 0$.

Case One. Suppose that $\alpha \neq 0$, - take for example, $\alpha = 1$. Let $x_1 = y$, $y_1 = x + y$, so that (3.6-12) is transformed to,

$$\begin{cases} \dot{x}_1 = (y^2_1 - 2x_1y_1 + 2x^2_1) \\ \dot{y}_1 = y_1 + (y^2_1 - 2x_1 y_1 + 2x^2_1) \end{cases} \qquad (3.6\text{-}13)$$

This is of the form (3.6-4), with $f_2(x_1,y_1) = g_2(x_1,y_1) = y^2_1 - 2x_1y_1 + 2x^2_1$. $y_1 + g_1(x_1,y_1) = 0 \Rightarrow y_1 + y^2_1 - 2x_1y_1 + 2x^2_1 = 0$ and the solution of this equation for which $\phi(0) = 0$ is easily seen to be,

$$y_1 = \phi(x_1) = \frac{(2x_1 - 1) + \sqrt{1 - 4x_1 - 4x^2_1}}{2}$$

$\therefore$ $\psi(x_1) = f_2(x_1,\phi(x_1)) = \phi^2 - 2x_1\phi + 2x^2_1 = -\phi(x_1)$, and so, by expanding by the binomial theorem in a small neighbourhood of $x_1 = 0$,

$$\psi(x_1) = \tfrac{1}{2} - x_1 - \tfrac{1}{2} + x_1 + x^2_1 + x^2_1 + \ldots = 2x^2_1 + \ldots$$

Thus, by theorem 3.6-1, the origin of (3.6-13) and hence of (3.6-12) is a 'saddle node'.

Case Two

If $\alpha = 0$, system (3.6-12) is already of the form (3.6-10), where $g_3(x,y) = x^2 + y^2$. Hence, in (3.6-11), $a_k = 1$, $k = 2$, $b_n = 0$, $c(x,y) = 1$. $A(x) \equiv 0$, so that, by theorem 3.6-2, the origin is a degenerate critical point.

Notice the topological nature of the phase portrait in a neighbourhood of the origin of (3.6-12) is changed when α passes through 0. The point $\alpha = 0$ is said to be a bifurcation point. (see section 3.10).

3.7). Limit and Minimal Sets:- Poincaré - Bendixon Theory

The notion of a limit set of an orbit of a vector field has already been defined (see section 3.5.). Both attractors and limit-cycles are necessarily members of the limit set, if they exist. Clearly, if the given vector field does not lead to a dynamical system, some of the limit sets will be empty. It could happen that, if the vector field is not complete, all the limit sets are empty.

Closely related to the limit sets are the minimal sets, which are generalizations of the critical elements. It turns out, that, in two dimensions, the possibilities for the limit sets and minimal sets are limited.

The following lemmas are intuitively obvious. ($\sigma = +, -, \pm$ as in 3.5).)

Lemma 3.7-1

If $m \in M$ is σ-complete and the limit set $w^{\sigma}(m)$ contains a unique element, then that point is a critical point.

□

Lemma 3.7-2

If $\Phi(m,t) = \gamma_m(t)$ is a periodic orbit, associated with a vector field

X, for each $m \in M$, $\Omega^+_X = \Omega^-_X = \Phi(m,t)$. □

The definition of a minimal set only applies when the associated vector field leads to a dynamical system:-

Definition 3.7-1

Let $X \in \mathcal{X}(M)$ be complete with flow Φ. A subset S of M is positively invariant if and only if $\Phi/_S \subset S$ for all $t \geq 0$ and negatively invariant if and only if $\Phi/_S \subset S$ for all $t \leq 0$, where $\Phi/_S$, the restriction of Φ to S is given by $\Phi/_S = \{\Phi(s,t), s \in S\}$. S is called invariant if it is both positively and negatively invariant.

Definition 3.7-2

$S \subset M$ is a minimal set if and only if S is closed, non-empty, invariant, and no proper subset of S has these properties.

It can be seen that, as stated earlier, critical elements are examples of minimal sets.

If M is a compact two-dimensional manifold, $X \in \mathcal{X}(M)$ and A is a minimal set of X, then it can be shown that there are only two possibilities. Either A is a critical element of X, or A = M, the torus $T^2 = S^1 \times S^1$. These two alternatives form the basis of the proof of the Poincaré-Bendixon theorem[22]:-

Theorem 3.7-3 (Poincaré - Bendixon - Schwartz)

Let M be a compact, connected, two-dimensional manifold, $X \in \mathcal{X}(M)$ and $m \in M$. Suppose that $w^\sigma(m)$ contains no critical points. Then either,

(a) $w^{\sigma}(m) = M = T^2$; or

(b) $w^{\sigma}(m) = \gamma$ is a closed orbit, that is, γ is a limit-cycle. □

In the second case, the orbit from m can be show to wind towards γ in the usual intuitive sense, that is, it can be thought of as approaching γ in a spiraling motion. This notion of one orbit winding towards another can be defined precisely in terms of a Poincare map acting on a transversal section of a manifold.[23]

The condition that M is compact is no real restriction if it is assumed that the orbit from m lies in a compact subset of M.

Unfortunately, there are no general results concerning the limit and minimal sets of orbits on higher dimensional manifolds at the present time.

3.8) The Stability of Orbits of Vector Fields

There exist many different notions of stability of an orbit of a vector field. Several different types are defined in this section and results concerning the stability of critical elements, in particular critical points are given. One of the most important methods for determining the nature of stability of a critical point is also given - that of Lyapunov.

The concept of a non-wandering point is also introduced and its relationship to stability is studied.

Recall that it is not, in general, true that a manifold M possesses a metric in the topological sense. For example, the manifold model of space-time given in chapter two, which possesses a Lorentz metric (in the geometrical sense) is not metrizable. However, every manifold is a topological space and the various forms of stability can be defined in terms of the continuity of a function mapping M to its power set, also a topological space.[24] If the given manifold does possess a metric, the various forms of stability reduce to the

definitions given in this section. Recall every manifold possessing a positive definite geometrical metric, possesses a topological metric. Thus, if the possibility of space-like vectors on the space-time manifold is excluded, it possesses a metric.

Definition 3.8-1

Let M be a manifold and $X \in \mathcal{X}(M)$, with flow $\Phi: \mathcal{D} \subset M \times \mathbb{R} \to M$. For $m \in M$, let $m_+ = \bigcup \left\{ \Phi(m,t) \mid (m,t) \in \mathcal{D}, t \geq 0 \right\}$

$$m_- = \bigcup \left\{ \Phi(m,t) \mid (m,t) \in \mathcal{D}, t \leq 0 \right\} .$$

Also, let $m_\pm = m_+ \cup m_-$.

Thus, m_+ gives the total range of an orbit in the positive direction and m_- gives the total range of the orbit in the negative direction. As in section 3.5), σ is taken to represent +, - or $\pm$.

Definition 3.8-2

Let M be a manifold with metric ρ, $X \in \mathcal{X}(m)$ and $m \in M$ be σ-complete. Let $\bar{\rho}$ denote the Hausdorff pseudometric on $\mathcal{P}(M)$, the power set of M, that is, the set of all subsets of M.

Then m is α^σ stable if and only if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\rho(m',m) < \delta \Rightarrow$

(i) $\alpha = 1$, $\bar{\rho}(m_\sigma, m'_\sigma) < \varepsilon$

(ii) $\alpha = 2$, $\lim_{t \to \sigma\infty} \rho(m, \Phi(m',t)) = 0.$

(iii) $\alpha = 3$, $\lim_{t \to \sigma\infty} \rho(\Phi(m,t), \Phi(m',t)) = 0.$

(iv) $\alpha = 4$, $\bar{\rho}(m', \Phi(m,t)) < \varepsilon$.

If m is 1^+ stable, it is sometimes called stable and if m is 2^+ stable, it is cometimes called asymptotically stable.

α=1 stability requires that if the orbits start at a distance δ apart, they remain within a given distance ε of each other. α = 2 or 3 stability requires that the orbit starting at m' actually approaches the orbit starting at m as $t \to \infty$ or $t \to -\infty$, if m, m' ε M are sufficiently close. α= 4 stability requires that the orbit starting from m, at a distance δ from m' remains within a distance ε of this point. It can be shown that if m ε M is 3^σ stable, it is 2^σ stable and that if it is 2^σ stable, then it is 1^σ stable.

Definition 3.8-3

If γ is a closed orbit of X ε $\mathcal{X}$(M), γ is α^σ stable if and only if m is α^σ stable for some m ε γ .

(It can be shown that if one such point is α^σ stable, then they all are).

Thus, a limit cycle is 3^+ stable if orbits wind towards it for increasing t and 3^- stable if orbits wind towards it for decreasing t.

Naturally, there is an opposite concept to that of stability - that of instability:-

Definition 3.8-4

If m ε M is not α^σ stable with respect to X ε $\mathcal{X}$(M), it is α^σ unstable with respect to X.

Closely related to the limit sets and critical elements of a vector field are the non-wandering points, which play an important role in the study of stability.

Definition 3.8-5

If X ε $\mathcal{X}$(M) with flow $\phi: \mathcal{S} \subset M \times \mathbb{R} \to M$ and m ε M, then m is a non-wandering point of X if and only if (m,t) ε $\mathcal{S}$ for all $t > 0$

and for all neighbourhoods U of m and all $t_o \geq 0$ there exists a $t > t_o$ such that $U \cap \phi(U \times \{t\})$ is non-empty. Let N_x be the set of all non-wandering points of $X \in \mathcal{X}(M)$.

Thus, the non-wandering points are a generalization of the critical points. It can be shown that, in fact, if M is compact, $\overline{\Gamma}_x \subset (\Omega_x^{\mp}) \subset N_x$. Clearly, every point 4^σ stable is non-wandering.

The non wandering points also play an important role in the study of generic properties of vector fields and structural stability, as seen in section 3.9).

Besides the types of stability already mentioned, there is a different kind, but equally important, that is due to Lagrange, which concerns the boundedness of solutions in some sense.

Definition 3.8-6

Any orbit $\gamma_m t$ is L^σ stable (σ -stable according to Lagrange) if $\overline{m}_\sigma$ is compact.

If M is compact, all motions are stable according to Lagrange. In the case of $\mathbb{R}^n$, stability L^σ implies that the orbit is situated in a bounded portion of $\mathbb{R}^n$. Clearly, for an orbit which is L^σ stable, Ω^σ_x is non empty. It can be shown that, in this case, Ω^σ_x contains a minimal set. It can also be shown that if a system possesses at least one motion stable L^σ, then the set of non-wandering points is non-empty.

Definition 3.8-7

A vector field is completely unstable if all its points are wandering.

If all of the orbits of the vector field are unstable according

to Lagrange, the vector field is said to be unstable.

It can be shown that a completely unstable system is unstable and that in $\mathbb{R}^2$, every unstable system is completely unstable.

The stability of critical points is now considered. Notice if m is critical, then m is 3^σ stable if and only if it is 2^σ stable. 4^σ stability becomes redundant. Thus, we are left with stability and asymptotic stability. Stability of an elementary critical point, which is now defined, can be deduced from the eigenvalues of the matrix of the linearized system in a neighbourhood of the critical point, as the following theorem shows.

Definition 3.8-8

If $X \in \mathcal{H}(M)$ and $m \in M$ is a critical point, m is elementary if and only if none of the characteristic multipliers of X at m have modulus one.

Thus, all elementary critical points are simple, but a point whose matrix has purely imaginary eigenvalues is not elementary. The above definition and following theorem also hold for closed orbits when the multipliers are defined in terms of a Poincaré map.[11]

Theorem 3.8-1

If $m \in M$ is a critical point of $X \in \mathcal{H}(M)$, there exist submanifolds S^+, CS^+, C, CS^-, S^- of M such that,

(i) Each is invariant under X and contains m.

(ii) $T_m(S^+)$ (respectively $T_m(CS^+)$, $T_m(C)$, $T_m(CS^-)$, $T_m(S^-)$) is the sum of the eigenspace in T_mM of characteristic multipliers of modulus < 1. (respectively ≤ 1, $= 1$, ≥ 1, > 1) and $T_m\{m\}$.

(iii) If $m' \in S^\sigma$, then $w^\sigma(m') = m$. ($\sigma = +$ or $-$).

(iv) S^+, S^- are locally unique.

□

S^+, S^- are extendable to globally unique manifolds, by expanding S^+, S^- by means of the flow of X. The manifolds are called respectively, the local stable (S^+), centre-stable (CS^+), centre (C), centre-unstable (CS^-) and unstable (S^-) manifolds of m. Notice, if all the characteristic multipliers of the linearized system about m have modulus <1, m is asymptotically stable. A centre is stable, but never asymptotically stable.

A useful tool in the investigation of the stability of the origin of a non-linear system,

$$\dot{x} = f(x) \qquad (3.8\text{-}1)$$

where $x = (x_1,\ldots,x_n)$, $f = (f_1,\ldots,f_n)$, is the Lyapunov function. It is assumed that f and its partial derivatives with respect to each x_i are continuous in a neighbourhood of the origin, an isolated critical point of (3.8-1).

Let V(x) be a scalar continuous function defined on some region G of $\mathbb{R}^n$ containing the origin.

Definition 3.8-9

V(x) is positive definite on G is and only if $V(0) = 0$ and $V(x) > 0$ for $x \neq 0$ and $x \in G$. V(x) is negative definite on G if and only if $-V(x)$ is positive definite on G.

It is further assumed that V(x) has continuous first order partial derivatives at every point of G.

Definition 3.8-10

The derivative of V(x) with respect to the system (3.8-1) is the scalar product,

$$V^*(x) = \text{grad } V \cdot f = \frac{\partial V}{\partial x_1} f_1 + \ldots + \frac{\partial V}{\partial x_n} f_n.$$

That is, along a solution $x = \psi(t)$, the total derivative of V with respect to t coincide with the derivative of V with respect to the system, evaluated at $x = \psi(t)$.

The original theorems of Lyapunov for autonomous systems may be stated as follows.

Theorem 3.8-2[25]

(i) If there exists a scalar function V(x) that is positive definite and for which $V^*(x) \leq 0$ on some region containing the origin, then the origin of (3.8-1) is stable.

(ii) If V(x) is positive definite and $V^*(x)$ is negative definite on some region containing the origin, the origin is asymptotically stable.

□

Theorem 3.8-3[25]

(i) If there exists a scalar function V(x), V(0) = 0, such that $V^*(x)$ is either positive or negative definite on some region G containing the origin and if there exists in every neighbourhood U of 0, where $U \subset G$, at least one point a such that V(a) has the same sign as $V^*(x)$, then the origin is unstable.

(ii) If $V^*(x) = \lambda V(x) + W$, where $\lambda > 0$ is a constant and W is either identically zero or is a non +ve or non -ve function such that in every neighbourhood U of 0, $U \subset G$, there exists at least one point a such that $V(a).W(a) > 0$, then the origin is unstable.

□

There are also corresponding theorems for non-autonomous systems.[25]

The difficulty in applying the above theorems to any particular

system is in finding a suitable function $V(x)$ satisfying the given conditions. For example, although $V(x) = x_1^2 + x_2^2$ is positive definite in two dimensions, it is not in three dimensions as $V(x) = 0$ at $x_1 = x_2 = 0$ independent of the value of x_3.

However, it may be possible to find a function $V(x)$ such that $V^*(x) \leq 0$. In this case there is a theorem concerning asymptotic stability which does not require $V^*(x)$ negative definite as in theorem 3.8-2. In general, (3.8-1) will possess more than one critical point, some of which may be unstable. Thus, the region of asymptotic stability of a particular critical point may not be the whole of $\mathbb{R}^n$. If the region is the whole of $\mathbb{R}^n$, the critical point is said to be globally asymptotically stable:-

Theorem 3.8-4[25]

Let there exist a scalar function $V(x)$ such that,

(i) $V(x)$ is positive definite on $\mathbb{R}^n$ and $V(x) \to \infty$ as $|x| \to \infty$ (where $|\ |$ denotes the euclidean norm on $\mathbb{R}^n$)

(ii) $V^*(x) \leq 0$ on $\mathbb{R}^n$

(iii) the origin is the only invariant subset of $E = \{x \mid V^*(x) = 0\}$.

Then the origin of (3.8-1) is globally asymptotically stable. □

Example 3.8-1

Consider the third order equation defined on the real line,

$$\dddot{x} + e^{x^2}(\ddot{x} + x) + 2\dot{x} = 0.$$

This is equivalent to the system,

$$\begin{cases} \dot{x} = y \\ \dot{y} = z \\ \dot{z} = -2y - e^{x^2}(x + z) \end{cases} \qquad (3.8\text{-}2)$$

defined on $\mathbb{R}^3$, with a critical point at the origin - in fact, this is the only critical point. A Lyapunov function is given by,

$$2V(x,y,z) = (x^2 + y^2) + (x + z)^2$$

Clearly V is positive definite and,

$$V^* = \frac{\partial V}{\partial x} y + \frac{\partial V}{\partial y} z + \frac{\partial V}{\partial z} \left\{ - 2y - e^{x^2} (x + z) \right\} = - 2e^{x^2} (x + z)^2 \leq 0.$$

Thus, the origin of (3.8-2) is stable by theorem 3.8-2. Notice $V^*(x) = 0$ when $x = -z$, and so $E = \left\{ (x,y,z) \mid x = - z \right\}$. Clearly, $x = -z$ is not a solution of (3.8-2) and the only invariant subset of E is (0,0,0)... By theorem 3.8-4, the origin is globally asymptotically stable.

3.9) Structural Stability

The concept of structurally stable systems of differential equations was first introduced by A. Androvonov and L. Pontrjagen for a certain class of two-dimensional autonomous differential equations on the plane.[26] In simple terms, structurally stable systems are those systems for which a small perturbation does not change the topological nature of the set of orbits. Structural stability is quite different from classical notions of stability of orbits, such as Lyapunov stability; the latter refers to individual orbits, whereas the former refers to the whole system. Structurally stable systems are useful when models of physical systems are being considered. If a structurally stable model is perturbed slightly to fit new experimental evidence, its theoretical qualatative conclusions are not radically altered.

A vector field is structurally stable, if the first order autonomous system it gives rise to is structurally stable.

In this section, the formal definition of a structurally stable vector field is given, together with useful results in this area of study. Before the definition can be given, a topology on the set of all vector fields on a particular manifold has to be defined.

Definition 3.9-1

Let $\mathcal{F}^r(U,F)$ denote the vector space of C^r mappings from U to F, with each kth derivative bounded (k = 0,1,...r where the 0th derivative is the mapping itself), where F is a real, finite dimensional vector space and U is an open subset of a real, finite dimensional vector space. $\mathcal{F}^r(U,F)$ is a Banach space if,

$$||f||_r = \sup \left\{ \sum_{k=0}^{r} || D^k f(u)|| \quad ; \quad u \in U \right\}$$

Let $\mathcal{X}^r(M)$ be the set of C^r vector fields on M. If (U,ϕ) is a chart on M, $\phi(U) = U' \subset \mathbb{R}^n$ and for $X \in \mathcal{X}^r(M)$ there is a coordinate representative X_ϕ: $U' \to \mathbb{R}^n$. If $\bar{U}'$ is contained in the image of a larger chart and is bounded in $\mathbb{R}^n$, then $X_\phi \in \mathcal{F}^r(U', \mathbb{R}^n)$. Thus, there exists an atlas $\mathcal{A} = \left\{(U_i,\phi_i)\right\}$ on M such that, $X_{\phi_i} \in \mathcal{F}^r(U'_i, \mathbb{R}^n)$ for every i and the $\mathcal{A}$ uniform topology on $\mathcal{X}(M)$ is the smallest topology such that the mapping $X \to X_{\phi_i}$ is continuous for all i.

If M is compact, this topology is independent of the atlas $\mathcal{A}$. If M is not compact, the union of topologies for all atlases $\mathcal{A}$ on M having the boundedness property stated above is the Whitney C^r topology on $\mathcal{X}^r(M)$. The space of C^r vector fields with respect to the Whitney topology is denoted by $\mathcal{X}^r(M)$.

One of the more recent aspects of the qualitative theory of vector fields to be investigated is that of generic properties. If $X \in \mathcal{X}^r(M)$, P(X) is a generic property of X is the subspace of vector fields possessing this property is dense in $\mathcal{X}^r(M)$. Thus, if P(X) is a generic property, almost every vector field has this property.[27]

A traditional concept in the theory of differential equations is that of a first integral of the system of equations under consideration. This leads to one example of a generic property.

Definition 3.9-2

A regular first integral of a vector field $X \in \mathcal{X}(M)$ is a function $f: M \to \mathbb{R}$ of class C^n, such that f is not constant on any open subset of M and $df(M).\left[X(m)\right] = 0$ at each point $m \in M$.

If M is compact, the property that X has no regular first integral is generic.[27]

Definition 3.9-3

Let $X,Y \in \mathcal{X}(M)$. If X and Y have equivalent phase portraits, we write $X \sim Y$. X is C^r structurally stable if and only if there is a neighbourhood U_0 of $X \in \mathcal{X}(M)$ in the Whitney C^r topology, such that $Y \in U_0 \Rightarrow X \sim Y$.

X is structurally stable if it is C^r structurally stable for all r.

Let the set of all C^r structurally stable vector fields on M be denoted by $\Sigma^r(M)$.

Example 3.9-1

Consider the n-dimensional disc D, $x_1^2 + \ldots + x_n^2 \leq 1$, $n \geq 2$ in Euclidean space and vector fields $X = \sum_{i=1}^{n} f_i \frac{\partial}{\partial x_i}$, $Y = \sum_{i=1}^{n} g_i \frac{\partial}{\partial x_i}$ of class C' defined on it. These give rise to the autonomous systems,

$$\begin{cases} \dot{x} = f(x) & (3.9\text{-}1)\ (a) \\ \dot{x} = g(x) & (3.9\text{-}1)\ (b) \end{cases}$$

where $x = (x_1,\ldots,x_n)$, $f = (f_1,\ldots,f_n)$, $g = (g_1,\ldots,g_n)$.

A suitable metric for $X,Y \in \mathcal{X}(D)$ is given by,

$$\rho(X,Y) = \max_{D} \left\{ \sum_{i=1}^{n} |f_i - g_i| + \sum_{i,j=1}^{n} \left| \frac{\partial f_i}{\partial x_j} - \frac{\partial g_i}{\partial x_j} \right| \right\}.$$

Then $\dot{x} = -x$, $n \geq 2$ is structurally stable, with respect to the above topology.[28]

The elementary closed orbits can be defined in terms of the eigenvalues of a Poincaré map[11], or can be defined in terms of a local chart, as in the following definition, which is restricted to two dimensions.

Definition 3.9-4

If $x = \phi(s)$, $y = \psi(s)$ is a closed path of the two-dimensional system,

$$\begin{cases} \dot{x} = f(x,y) \\ \dot{y} = g(x,y) \end{cases}$$

having period τ and $I = \int_0^\tau (f'_x(\phi(s), \psi(s)) + g'_y(\phi(s), \psi(s))\, ds \neq 0$,

then the closed path is elementary.

Theorem 3.9-1[29]

Let M be a two-dimensional manifold. $X \in \mathfrak{X}(M)$ is structurally stable in $\bar{G} \subset M$ is and only if

(i) there are only a finite number of critical elements and they are all elementary.

(ii) $N_X = \Gamma_X$

(iii) There are no orbits going from saddle point to saddle point. □

Notice if there are a finite number of closed paths these will usually be limit-cycles. Also, if a closed path is elementary it will still contain a critical point in its interior which will be non-elementary. Thus, in this case, the region of structural

stability cannot include the critical point.

For a compact manifold, it can be shown that $\overline{\Gamma}_X \subset (\overline{\Omega^{\pm}}_x) \subset N_X$. If the number of critical elements are finite, clearly $\overline{\Gamma}_X = \Gamma_X$ and then condition (ii) of the theorem is equivalent to $\Omega_X \subset \Gamma_X$, (as $\Gamma_X \subset \Omega_X$ holds for all manifolds). Thus, by theorem 3.7-3, for a compact, connected, orientable manifold which is not the torus, condition (ii) is virtually redundant.

Example 3.9-2

Consider again the n-dimensional system (3.9-1)(a) with $f(x) = -x$. The system is linear and has only one critical point at the origin. In the two-dimensional case, the origin is a node by lemma 3.5-1 - the system is structurally stable by theorem 3.9-1.

It can be shown that structural stability in two dimensions is a generic property[29]. This is not the case in higher dimensions[30].

The best result for structural stability in higher dimensions at the present time is,

Theorem 3.9-2[31,32]

If $X \in \mathcal{X}(M)$ and X is C' structurally stable then,

(i) critical elements are elementary
(ii) $N_X = \overline{\Gamma}_X$
(iii) X has no regular first integral.

□

3.10 Bifurcation Theory

If a given vector field gives rise to a system of equations

which contains one or more parameters, which can change, then an alteration in the parameter values may lead to changes in the qualitative properties of the system. For example, orbits which were stable could become unstable or the phase portrait could completely alter. The study of such changes is bifurcation theory. If the phase portrait changes at a particular set of parameter values, the set of parameter values is known as the bifurcation point. For instance, $\alpha = 0$ is a bifurcation point of 3.6-12.

Bifurcation theory was originally devised by Poincaré for the purpose of finding periodic solutions of the equations of celestial mechanics. The theory has been extended, developed and applied by various people, but mostly in the case of periodic solutions[33].

The most useful results of the theory in relation to dynamical amendments of the Newtonian and Schwarzchild equations, are those which deal with changes in the nature of certain non-elementary critical elements in two-dimensional phase space. (Clearly, only non-elementary critical elements are considered as elementary critical elements are structurally stable).

First of all, the case of a multiple focus, a certain kind of non-elementary critical point, is considered.

Definition 3.10-1

A critical point whose matrix of the linearized system has purely imaginary eigenvalues, which is not a centre, is a multiple focus.

Before stating the main theorem, some new definitions are needed.

Definition 3.10-2

Let $f_o(x_1, \ldots, x_n)$ be analytic (or C^k) in a region G and r a natural number such that $r \leq k$ if f_o is C^k.

A function $f(x_1,\ldots,x_n)$ of class $k_1 \geq r$ or analytic in G is said to be δ-close to rank r to the function f_o in G if at any point of G,

$$|f - f_o| < \delta \ , \ |f^{(\ell)}_{x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}} - f_o^{(\ell)}_{x_1^{\alpha_1} x_2^{\alpha_2} \ldots \alpha_n}| < \delta$$

where $\ell = 1,2,\ldots r$, all α_i are non-negative and

$\alpha_1 + \alpha_2 + \ldots + \alpha_n = \ell \ ; \ \alpha_i, \ell \ \varepsilon \mathbb{Z}$.

Definition 3.10-3

Consider the two autonomous systems,

$$\begin{cases} \dot{x} = F(x) & (3.10\text{-}1)\ (a) \\ \dot{x} = G(x) & (3.10\text{-}1)\ (b) \end{cases}$$

defined in a bounded region of $\mathbb{R}^n$, where $x = (x_1,x_2,\ldots,x_n)$ and $F = (F_1,F_2,\ldots,F_n)$, $G = (G_1,G_2,\ldots,G_n)$.

The systems are C^k or of analytic class if F, G are C^k functions or anlytic in G.

If both systems are of analytic class or C^k, system (3.10-1)(b) is δ-close to rank r to system (3.10-1)(a) in G is each component of G is δ-close to rank r to each corresponding component of F.

As well as these definitions being useful in bifurcation theory, they are also useful in the study of structural stability. Clearly, the concept of structural stability in $\mathbb{R}^n$ may be re-defined by using the definition of close systems as these represent a neighbourhood of the original system and these neighbourhoods lead to a topology on the space of autonomous systems of analytic class on $\mathbb{R}^n$.

The definitions are also of considerable value when proving certain systems of autonomous equations are dynamical and will be used in Chapter Five.

For the rest of this chapter, only autonomous systems defined on the plane are considered. Before stating any theorems concerning multiple foci, it is necessary to define the concept of the multiplicity of such foci. The method of determining the multiplicity leads to a systematic way of differentiating between a multiple focus and a centre.

It is assumed that the two-dimensional system under consideration has a critical point at the origin, without loss of generality. A necessary condition for the presence of a multiple focus is that the matrix of the linearized system about (0,0) has purely imaginary eigenvalues. Then the system can be reduced to the canonical form,

$$\begin{cases} \dfrac{dx}{dt} = -\beta y + \phi(x,y) = f(x,y) \\ \dfrac{dy}{dt} = \beta x + \psi(x,y) = g(x,y) \end{cases} \qquad (3.10\text{-}2)$$

where β is a constant greater than zero. It is assumed that ϕ, ψ contain no linear terms and that (3.10-2) is analytic or C^k.

Transforming to polar coordinates $x = \rho \cos\theta$, $y = \rho \sin\theta$, (3.10-2) becomes,

$$\frac{d\rho}{dt} + F(\rho,\theta) \ , \ \frac{d\theta}{dt} = \beta + G(\rho,\theta) \Rightarrow$$

$$\frac{d\rho}{d\theta} = R(\rho,\theta) = \frac{F(\rho,\theta)}{\beta + G(\rho,\theta)} \ , \qquad (3.10\text{-}3)$$

where it is assumed that $G(0,\theta) \equiv 0$ for all θ.

Definition 3.10-4

Let $\rho = f(\theta \ ; \ 0, \rho_0)$ be the solution of (3.10-4) satisfying the

initial condition $f(0;0,\rho_o) = \rho_o$, where $|\rho_o| < \delta$, a small positive number. The functions $f(\rho_o)$, $d(\rho_o)$ are defined by

$$f(\rho_o) = f(2\pi;\ 0,\ \rho_o),\ d(\rho_o) = f(2\pi;\ 0,\ \rho_o) - \rho_o.$$

The value of the ith derivative of $d(\rho_o)$ at $\rho_o = 0$ is the ith focal value of the critical point at (0,0).

It can be shown that if there exists $r \in \mathbb{Z}_{>0}$ such that $d'(0) = 0, \ldots, d^{(r)}(0) \neq 0$, r must be odd[34]:-

Definition 3.10-5

If there exists r such that $d'(0) = 0, \ldots d^{(r)}(0) \neq 0$, let $r = 2m + 1$, $m \geq 0$. Then (0,0) is a focus of multiplicity m. $(r \leq k)$

Clearly, in general, equation (3.10-3) is difficult to solve explicitly for $\rho = f(\theta)$. However, if the system under consideration is analytic, focal values may be calculated by first series-expanding $R(\rho,\theta)$ in (3.10-3) in powers of ρ and expanding $f(\theta;\ 0,\rho_o)$ in powers of ρ_o. The coefficients of the later expansion (functions of θ) can be obtained from a set of recursive differential equations. For a multiple focus, it can be shown that the first two focal values must be zero, but, clearly, at least one of the focal values do not vanish. In the case of a centre, all focal values vanish and so a centre is thus observed when an infinite number of conditions are satisfied. An example of the method for calculating focal values is given in Chapter Five - the method is too long for a simple example to be given here.

There are several theorems relating to small perturbations of systems possessing a multiple focus, the most sophisticated being,

Theorem 3.10-1[35]

Let (0,0) be a multiple focus of multiplicity m of a two-dimensional autonomous system of class $N \geq 2m + 1$ or of analytic class. Let $U\varepsilon_o(0)$ be an open ball, radius ε_o, centre (0,0). Let $\delta_o > 0$ where ε_o, δ_o are so small that any system δ_o-close to the original system has a single critical point in $U\varepsilon_o(0)$ which is a focus.

(i) $\exists$ ε_o, δ_o such that any system δ_o-close to rank 2m+1 to the original system has at most m closed paths in $U\varepsilon_o(0)$.

(ii) For any ε and δ, $0 < \varepsilon \leq \delta_o$, $0 < \delta \leq \delta_o$ and for any s, $1 \leq s \leq m$, there exists a system of class N (or an analytic system) which is δ-close to rank $2m + 1$ to the original system and has in $U_\varepsilon(0)$ precisely s paths.

(iii) If a system is δ_o-close to rank 2m+1 to the original system and has m limit-cycles in $U\varepsilon_o(0)$, all these cycles, and likewise the focus of the perturbed system lying in $U\varepsilon_o(0)$, are structurally stable (simple).

□

It is also possible to investigate topological changes in the phase portrait in a neighbourhood of a multiple (that is, non-elementary) limit cycle, resulting from a small change in the right hand side of the two-dimensional autonomous system. It can be shown that any system close to the given system can have at most m closed paths in a neighbourhood of the multiple limit-cycle of multiplicity m[36].

Finally, we consider what happens to closed paths (not necessarily limit-cycles) when the right hand side of a two-dimensional system is changed. The simplest and most useful result in this area relates to Hamiltonian systems.

Definition 3.10-6

A system of the form,

$$\begin{cases} \frac{dx}{dt} = -\frac{\partial H(x,y)}{\partial y} \\ \frac{dy}{dt} = \frac{\partial H(x,y)}{\partial x} \end{cases} \qquad (3.10\text{-}4)$$

where $H(x,y)$ is a single-valued function defined in a closed region G, and is called a Hamiltonian system in G.

If a system has concentric closed paths in the phase portrait as orbits, for instance, if the origin is a centre, it must be related to a Hamiltonian system in a simple manner, as the following theorem states:-

Theorem 3.10-2[37]

Let (3.10-2) be an analytical system and G a closed annular region completely filled with concentric closed paths of (3.10-2). Then there exists an analytic function $\rho(x,y)$ defined in G which satisfies the condition,

$$\frac{\partial \rho}{\partial x} f(x,y) + \frac{\partial \rho}{\partial y} g(x,y) + \rho\left(\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y}\right) = 0$$

and the system $\frac{dx}{dt} = \rho f$, $\frac{dy}{dt} = \rho g$ is Hamiltonian. □

(The theorem applies equally well if the given system is not in canonical form). In the case of bifurcations of Hamiltonian systems, the following theorem is appropriate:-

Theorem 3.10-3[38]

Let C be a closed path of the Hamiltonian system, (3.10-4); $x = \phi(t)$, $y = \psi(t)$ the motion corresponding to the path, τ the period of the functions ϕ, ψ, G_o the region enclosed by C and,

$$\begin{cases} \dfrac{dx}{dt} = -\dfrac{\partial H}{\partial y} + \mu f_1(x,y) + \mu^2 f_2(x,y) + \dots \\ \dfrac{dy}{dt} = \dfrac{\partial H}{\partial x} + \mu g_1(x,y) + \mu^2 g_2(x,y) + \dots \end{cases} \tag{3.10-5}$$

a perturbation of (3.10-4), (μ a small parameter).

Then, if

$$I = \iint_{G_o} (f'_{1x}(x,y) + g'_{1y}(x,y))\, dx\, dy = 0 \text{ and,}$$

$$L = \int_0^\tau (f'_{1x}(\phi(s), \psi(s)) + g'_{1y}(\phi(s), \psi(s)))ds \neq 0,$$

there exist $\varepsilon > 0$, $\delta > 0$ such that

(a) for every μ, $|\mu| < \delta$, (3.10-5) has one and only one closed path in $U\varepsilon_o(C)$ and this contracts to C for $\mu \to 0$.

(b) this path is a structurally stable limit-cycle.

3.11) Concluding Remarks

This chapter has been concerned with dynamical systems and their properties, in particular with the behaviour of orbits in a neighbourhood of a critical element and their stability and structural stability and bifurcation theory of the system itself. A lot of the definitions and theorems given hold also for arbitrary autonomous systems. Dynamical systems have been studied from the point of view of vector fields acting on manifolds and also as sets of autonomous differential equations, although most of the abstract theorems and definitions given in terms of the flow of a vector field hold for dynamical systems in the abstract sense of definition 3.1-1.

As stated at the beginning of the chapter the results presented are far from complete. Indeed, the global properties of orbits on manifolds is an active area of research at the present time. For a more thorough

account, see for instance, the works of A. Andronov, E. Leontovich, I. Gordon and A. Maier, V. Nemytskii and V. Stepanov, and R. Chillingworth[39].

Having looked at dynamical systems in a general setting, the theory is now applied specifically to gravitation theory. In particular, the language of dynamical systems is employed in a translation of the singularity theorems in the next chapter and the various properties presented here hold for the dynamical amendments to the equations of motion in the following two chapters.

Chapter Four:-

The Singularity Theorems in the Language of the Qualitative theory of Differential Equations

In this work, space-time is assumed to have the structure of a Lorentz manifold and so there is a great similarity between the study of the geodesics and their defining second-order differential equations and the general theory of dynamical systems and differential equations as outlined in the previous chapter. In particular, the geodesic equations derived from the given metric are autonomous and the notion of geodesic incompleteness, introduced in order to get some sort of definition of a space-time singularity, corresponds to whether or not the geodesic equations have solutions which can be extended to arbitrary values of the affine parameter. If the geodesic equations are dynamical, the geodesics are complete.

In this chapter, it is shown how the singularity theorems may be translated into the language of autonomous differential equations and vector field theory. It is shown how some of the hypotheses of the theorems may be false in the dynamical case.

As stated in Chapter Two, conjugate points are basic to the proofs of the singularity theorems and in order for a space-time to be considered singular in the sense of definition 2.4-1, it is necessary that there exist conjugate points. However, the presence of conjugate points alone does not guarantee geodesic incompleteness. It is shown in the second subsection that conjugate points divide naturally into two classes:- those of the regular type and those of the singular type. The assumptions of the singularity theorems imply the existence of singular conjugate points, which lead to geodesic incompleteness.

It is assumed throughout this chapter that the model for space-time is as in the Einstein case, except that for reasons discussed in Chapter Two, the field equations do not necessarily hold exactly in the strong field region.

4.1) The Causal Structure of Space-time and Space-time Singularities

By hypothesis, space-time is a four-dimensional manifold,M, which is orientable, and which possesses a Lorentz metric. Any curve on the manifold is an orbit of some particular vector field - its tangent vector field. (assuming that the curve is continuous with respect to the induced topology, where it is defined). In particular, each geodesic curve is derivable from a vector field, although the set of geodesic equations as a whole satisfy the second order autonomous equations,

$$\frac{d^2x_a}{d\lambda^2} + \Gamma^a_{bc} \frac{dx_b}{d\lambda} \frac{dx_c}{d\lambda} = 0 \qquad (4.1\text{-}1)$$

in the co-ordinate domain of a chart $\{x_a\}$, if co-ordinate bases are used. Expressed as an eight-dimensional, first-order system, (4.1-1) becomes,

$$\begin{cases} \dfrac{dx_a}{d\lambda} = y_a \\ \dfrac{dy_a}{d\lambda} = -\Gamma^a_{bc}\, y_b y_c \end{cases} \qquad (4.1\text{-}2)$$

Thus, the geodesics can be thought of as being orbits of one particular vector field acting on the eight-dimensional manifold consisting of points of M and their derivatives. (They can also be thought of as orbits of a particular vector field called the geodesic spray, acting on the tangent bundle of M - the second order equation (4.1-1) is derivable from this field[1]). The completeness of geodesics can thus be defined in terms of the completeness of any one of the vector fields discussed above. Thus, we could have,

Definition 4.1-1

Space-time contains a singularity if there is an incomplete non-spacelike geodesic vector field defined on M. (A geodesic vector field being a vector field all of whose orbits are geodesics).

Notice that it takes more than one vector field on M to define all the

non-spacelike geodesics on M. The above definition is really equivalent to the previous definition of a 'singularity' (definition 2.4-1). Notice that it may be generalized to include any vector field and so include the possibility of 'space-like singularities', or be generalized to include any non-spacelike vector fields. The former case is given in the form of the definition below, while the latter generalization agrees partially with Schmidt's b-completeness.

Definition 4.1-2

Space-time contains a singularity if the geodesic spray is incomplete. (This means in effect that the set of equations (4.1-2) has a solution which cannot be extended to arbitrary values of the affine parameter. It is this definition that will be used in this chapter.)

Any vector field acting on a compact manifold must be complete by proposition 3.2-2, and so the geodesic vector fields must be complete which implies that the geodesic spray is also complete. Thus, we have,

Lemma 4.1-1

If space-time is compact, it is singularity- free.

□

From now on, unless otherwise stated, the space-time manifold is assumed to be non-compact.

The notions of the chronological future (past) and the causal future (past), given in Chapter Two, may be defined in terms of the sets $x^m{}_\sigma$ (σ= + or -) given in Chapter Three (definition 3.8-1):-

Definition 4.1-3

Let $\chi^T(M)$ be the set of all timelike vector fields on M and let $\chi^{NS}(M)$ denote the set of all non-spacelike vector fields. Then,

$$J^\sigma(m) = \bigcup_{X\in\chi^{NS}(M)} x^m{}_\sigma \quad ; \quad I^\sigma(m) = \bigcup_{X\in\chi^{T}(M)} x^m{}_\sigma . \qquad (\sigma = + \text{ or } -)$$

where $m \in M$. If $A \subset M$, $J^{\sigma}(A) = \bigcup_{m \in A} J^{\sigma}(m)$ (similarly for $I^{\sigma}(A)$).

(Notions such as $J^{\sigma}(A,\mathcal{U})$, $\mathcal{U} \subset M$ and $E^{\sigma}(A)$ can be defined in terms of the above sets as in Chapter Two).

Three important concepts as far as the second and third singularity theorems are concerned are the chronology, causality and strong causality conditions. The chronology and causality conditions can be re-phrased as 'no timelike (respectively, non-spacelike) vector field has periodic orbits in M'. It can be seen that closely related to the concept of strong causality is the concept of a non-wandering point (see definition 2.4-5 and definition 3.8-5). A wandering point has a neighbourhood which never returns to intersect itself and so the strong causality condition holds if every non-spacelike vector field has no non-wandering points, except for the critical points:-

Definition 4.1-4

The strong causality condition holds on M if every non-wandering point of a non-spacelike vector field is critical.

Defintion 4.1-5

Let $C: \mathbb{R} \to M$ be a curve on M. $p \in M$ is a future end-point of C if, for every neighbourhood V of p, $\exists\, \lambda \in$ domain of the parameter, such that $C(\lambda_1) \in V$ for every $\lambda_1 \in$ domain of the parameter $\lambda_1 \geq \lambda$.

Notice, if C is an orbit of a complete vector field, the only possibility for a future end point is a critical point.

If c has no future end point, it can either leave M at infinity or a possible singularity, or become imprisoned in a set $A \subset M$:-

Definition 4.1-6

A curve $c: \mathbb{R} \to M$ is imprisoned in a compact subset A of M if it has no future (past) end point and given any λ, $\exists\, \lambda_1 > \lambda$ such that $\Phi(m,\lambda_1) \in A$ where $\Phi(m,0) = m$.

Thus, an imprisoned curve enters and remains in a compact set A or continually re-enters but does not remain in A. For instance, if the limit set of a curve is a limit-cycle, the curve is imprisoned in a neighbourhood of the limit-cycle. Clearly, if a limit-cycle is present in M, the strong causality condition cannot hold. It can be shown generally that imprisonment is incompatible with strong causality:-

Proposition 4.1-2[2]

If the strong causality condition holds on a compact set $A \subset M$, then there is no non-spacelike curve imprisoned in A. □

Notice that the definitions 4.1-4 and proposition 4.1-2 can be generalized to arbitrary curves and, in fact, hold for any general manifold and so the following general corollary can be stated:-

Corollary 4.1-3

If every non-wandering point of the vector field X defined on a compact set $M \subset A$ are critical, there are no orbits of X imprisoned in A. In particular, there are no limit-cycles of X in A.

4.2) Regular and Singular Conjugate Points

It is shown in this subsection that conjugate points can be divided into two sub-classes, one occuring at perfectly well-behaved points of space-time and the other in singular regions.

It has been assumed that the metric for space-time is at least C^2, (recalling that singularities are not assumed to be part of the space-time manifold and so the metric only has to be C^2 in non-singular regions - it may also not be differentiable at apparent singularities due to the fact that it has been expressed in terms of the wrong co-ordinate basis). Then, as the affine connection coefficients Γ^a_{bc} involve the derivatives of g_{ab}, Γ^a_{bc} must be at least once-differentiable. Thus, the right-hand sides of (4.1-2) are continuous and have continuous partial derivatives and so by theorem 3.1-2, solutions of the initial value problem are unique, except in a small

region of the singularity; that is, the orbits of (4.1-2) never intersect except at any singularities and at any critical points. If (4.1-2) is dynamical in all regions of space-time, that is, the geodesic vector fields are complete, the only possible points of intersection are critical points. Thus, there are no regular points of intersection in the eight-dimensional phase space of (4.1-2). However, this does not imply that there are no regular points of intersection of the geodesics on the four-dimensional manifold.

Although there is uniqueness of solutions to the initial value problem, starting at a particular point on the manifold, taking two distinct conditions on the derivatives say, will give two distinct curves passing through the same point. If the curves converge again at some other point, the original point may be conjugate to this, but does not represent a point of termination of the geodesics.

Thus, a conjugate point in a congruence of geodesics can represent a critical point, that is, a critical point of the first integrals of the geodesic equations, a singularity, or a regular point as above. The singularity theorems predict that the conjugate points under consideration are singular.

Notice, as far as the test particle or light ray is concerned, a critical conjugate point can never be reached in a finite value of the affine parameter, by theorem 3.3-2, whereas a singular conjugate point represents a point off the manifold where the geodesics converge such that a test particle or light ray can never escape and a regular conjugate point is a point at which geodesics converge and then diverge, with no effects to the test particle or light ray. In the case where the geodesic equations are dynamical, any conjugate points must be regular or critical.

The above concepts can be defined in terms of a Jacobi field defined along a geodesic as shown shortly. Recall that the Jacobi field along a null or timelike geodesic satisfies (2.2.7), (where in the timelike case, $a,b=1,\ldots,3$ and in the null case, $a,b=1,2$;

λ is the affine parameter). Then, without loss of generality, in both cases, the Jacobi field satisfies the single vector field equation,

$$\ddot{Z} = -\mathcal{R}Z \qquad (4.2\text{-}1)$$

where a dot represents differentiation with respect to the parameter λ and Z represents the column vector of components of Z at any point. $\mathcal{R}$ is a matrix field, whose typical (a,b)th element is R_{a4b4} at any point.

Notice (4.2-1) is a linear equation, but is non-autonomous as the matrix field $\mathcal{R}$ is, via the geodesic equations, implicitly a function of the parameter λ.

Let $\dot{Z} = Y$ in (4.2-1) so that $\dot{Y} = -\mathcal{R}Z$. Also, let $X = \begin{pmatrix} Z \\ Y \end{pmatrix}$ and the block matrix field $\mathcal{C} = \begin{pmatrix} 0 & I \\ -\mathcal{R} & 0 \end{pmatrix}$, all elements being at most 3 x 3, where I is the identity matrix of order two or three. Then (4.2-1) can be written as the single first order vector field equation,

$$\dot{X} = \mathcal{C}X \qquad (4.2\text{-}2)$$

If the geodesic equations (4.1-1) have global existence of solutions, the separation of two neighbouring geodesics must be defined for all values of the parameter, so that, in this case, the system (4.2-2) must be quasi-dynamical, (or at least have global existence of solutions). Similarly, if (4.2-2) is quasi-dynamical, the separation along geodesics is defined for all values of the defining parameter and so the geodesics must have global existence of solutions. Thus, the concept of geodesic completeness could be re-defined in terms of the nature of the Jacobi equation (4.2-1).

It has been stated that the Jacobi field measures the separation of neighbouring geodesics in a congruence and that two points are conjugate if there exists a non-zero Jacobi field which vanishes at the two points. Thus, if the first point is regular, we can start where the separation is zero ($\dot{Z}(\lambda) \neq 0$) and follow the geodesics along to the next point of separation (that is, there are two distinct geodesics passing through each point). However, if the starting point happens to

be a critical point (so that $\dot{Z}(\lambda) = 0$) or a singularity, then we cannot follow the geodesics away. Hence the reason for taking a starting point arbitrarily close to a point of intersection in the definition of conjugate points.

Suppose that $q \in M$, the space-time manifold, is conjugate to $p \in M$. Then $Z(p) = 0$, $Z(q) = 0$. Let $\dot{Z}(p) = B(\neq 0)$ and $\dot{Z}(q) = C$, where B,C are constant column vectors. If q is conjugate to a spacelike surface ζ, and $p \in \zeta$, let the initial conditions for the Jacobi field Z be $Z(p) = B$ so that $\dot{Z}(p) = \chi(p)B$, where $\chi(p)$ is the matrix whose elements are the components of the second fundamental tensor at p. (see section 2.2)). Let $\dot{Z}(q) = C$ as before. Then, in either case, if $\mathcal{R}(\lambda)$ is continuous and $C = 0$, q is also a critical point of (4.2-1), (by lemma 3.4-1). Notice, however, that the majority of conjugate points are not critical points of (4.2-1). (The only possible critical point of (4.2-2) being $X = 0$).

Consider now the concepts of regular, critical and singular conjugate points introduced previously. Clearly, the following definition characterizes each of these:-

Definition 4.2-1

Let q be a conjugate point (with respect to p, or a spacelike surface). Then q is a regular, singular or critical conjugate point (with respect to p, or a spacelike surface), if it is a regular, singular or critical point of (4.2-2) respectively. (In this context, a critical point of (4.2-2) reached infinite parameter time is classed as singular as it represents a finite termination point of the geodesics).

Consider first of all the case of a regular conjugate point, so that $Z(q) = 0$. If $C = \dot{Z}(q) \neq 0$, there is no problem - there is no turning point of separation (negative convergence can be interpreted as divergence and negative divergence as convergence). Suppose then, that $C = 0$. Nsw q is not a critical point, so that $\mathcal{R}(\lambda)$ is necessarily discontinuous at q. Recall, from section 3.4), that this condition is not sufficient, however. If (4.2-2) is quasi-dynamical, q is a regular conjugate point if $C \neq 0$ or if $C = 0$ with

$\mathcal{R}_0(\lambda q)$ singular.

If q is a critical conjugate point, then C = 0 and $\mathcal{R}(\lambda)$ must be continuous at q (by lemma 3.4-1). This is the only possibility. If q is singular, then $\mathcal{R}_0(\lambda)$ must necessarily be singular there, with either C = 0 or C ≠ 0, but, once again, if C = 0, the condition is only necessary.

As stated before, if the geodesic equations are dynamical, the only conjugate points occuring are critical or regular types. However, the geodesics may still be complete if the geodesic equations have global existence of solutions (so that if a singularity is present it can never be reached - the possibility of starting at the singularity being excluded). In this case, there may be singular conjugate points, but these will never be reached (so that the geodesics can only leave the space-time manifold at infinity). Both proposition 2.2-3 and 2.2-4, predicting the occurence of conjugate points, predict regular or singular conjugate points.

Notice that, in the definition of conjugate points, p and q need not be distinct, that is, it is possible for a point to be conjugate to itself, if the geodesics can return to an arbitrarily small neighbourhood of it. Clearly from definition 2.4-5, we have,

Lemma 4.2-1

A point p cannot be conjugate to itself if the strong causality condition holds at p. □

However, usually, a 'singularity' will consist of a set of non-isolated singular points (for instance, the 'singularity' in the Schwarzchild metric, is the region r = 0, for all co-ordinate time) and so it is certainly possible for a singularity to be conjugate to itself, in the sense that infinitesimally neighbouring geodesics arbitrarily close to one point of the region can intersect at another point of the region.

Example 4.2-1

Consider the dynamical system on the plane,

$$\begin{cases} \dot{x} = x \\ \dot{y} = y \end{cases} \qquad (4.2\text{-}3)$$

The critical point at (0,0) is a degenerate node, a point of intersection of all the orbits of (4.2-3), in this case straight lines. However, this critical point is not conjugate to another point (except possibly to infinity), although it can be conjugate to any curve, in which case it is a critical conjugate point.

Example 4.2-2

The geodesic equations derived from the four-dimensional Minkowski metric are,

$$\frac{d^2x_i}{d\lambda^2} = 0 \Rightarrow \frac{dx_i}{d\lambda} = c_i, \quad c_i \in \mathbb{R}, \; i = 1,2,\ldots,4 \qquad (4.2\text{-}4)$$

and so the geodesics are straight lines and complete. There are regular points of intersection on the manifold, but not for constant c_i. However, these points are not conjugate to any other as any pair of geodesics only intersect once. The null and timelike convergence conditions hold trivially as the space is flat, but the generic condition does not hold, hence the reason for the absence of conjugate points.

Example 4.2-3

Consider two-dimensional anti-de sitter space, given by the metric,[3]

$$ds^2 = -(1+x^2)dt^2 + \frac{1}{(1+x^2)}\,dx^2 \qquad (4.2\text{-}5)$$

This is a space of constant negative curvature, that is, the solution of the Einstein field equations in two-dimensions for which $R = -2$ ($T^{ab} = 0$).

The components of the Ricci tensor are,

$$R_{11} = 1 + x^2, \quad R_{22} = -\frac{1}{(1+x^2)}; \quad R_{12} = R_{21} = 0. \qquad (4.2\text{-}6)$$

and, in this case, the generic condition holds. Notice, if ω^a is a null or timelike vector,

$$R_{ab}\omega^a\omega^b = \frac{1}{2}R\, g_{ab}\omega^a\omega^b = -\, g_{ab}\omega^a\omega^b \geq 0,$$

so that the geodesic convergence conditions hold. Thus, by theorem 2.2-4 there will be conjugate points. However, the global metric (4.2-5) is non-singular and so the geodesic equations may be dynamical - this is indeed the case:- The geodesic equations derived from (4.2-5) are,

$$\begin{cases} \dfrac{d^2t}{d\lambda^2} + \dfrac{2x}{1+x^2}\dfrac{dt}{d\lambda}\dfrac{dx}{d\lambda} = 0 \quad \text{(a)} \\ \dfrac{d^2x}{d\lambda^2} - \dfrac{x}{1+x^2}\left(\dfrac{dx}{d\lambda}\right)^2 + x(1+x^2)\left(\dfrac{dt}{d\lambda}\right)^2 = 0 \quad \text{(b)} \end{cases} \tag{4.2-7}$$

(4.2-7)(a) can be integrated directly:-

$$\frac{d}{d\lambda}\left(-(1+x^2)\frac{dt}{d\lambda}\right) = 0 \Rightarrow \frac{dt}{d\lambda} = \frac{K}{1+x^2}, \quad K \text{ constant} \tag{4.2-8(a)}$$

Then (4.2-7)(b) becomes,

$$2\ddot{x} + \frac{2x}{1+x^2}\left\{K^2 - (\dot{x})^2\right\} = 0 \Rightarrow$$

$$\frac{du}{dx} + \frac{2x}{1+x^2}\left\{K^2 - u\right\} = 0, \text{ where } u = \left(\frac{dx}{d\lambda}\right)^2$$

$$\therefore \quad \frac{u}{1+x^2} = c + \frac{K^2}{1+x^2} \Rightarrow \frac{dx}{d\lambda} = \sqrt{c(1+x^2)+K^2} \tag{4.2-8(b)}$$

where c is a constant.

Thus, (4.2-7) has a first integral given by (4.2-8). It can be shown that (4.2-8) has global existence of solutions, simply by solving explicitly. For instance, in the case of timelike geodesics, $\lambda = \tau$, the proper time, $c = -1$ and solving (4.2-8) we get,

$$x = \sqrt{K^2 - 1}\,\sin(\tau+\tau_o); \quad t + t_o = \tan^{-1}(K\tan(\tau+\tau_o)) \Rightarrow \tag{4.2-9}$$

$$t + t_o = \tan^{-1}\left\{\frac{Kx}{\sqrt{K^2 - 1 - x^2}}\right\}$$

τ_o, t_o, constants of integration.

In this case, x and t can be extended for all values of τ. $x = 0$, $K^2 - 1$ (K positive or negative) are points of intersection of the geodesics but are regular points.

Thus, any predicted conjugate points in this case are regular as required.

Example 4.2-4

The geodesic equations, derived from the Schwarzchild metric (2.4-1) are integrable and the first integrals are given by,

$$\begin{cases} \dfrac{dt}{d\lambda} = \dfrac{\pm K}{1 - 2M/r} \\ \dfrac{dr}{d\lambda} = \sqrt{K^2 + (1 - \dfrac{2M}{r})(\varepsilon - \dfrac{c}{r^2})} \\ \dfrac{d\theta}{d\lambda} = \dfrac{1}{r^2}\sqrt{c - L^2\mathrm{cosec}^2\theta} \\ \dfrac{d\phi}{d\lambda} = \dfrac{L}{r^2\sin^2\theta} \end{cases} \qquad (4.2\text{-}10)$$

where K,L,C, ε are constants of integration ($\varepsilon = -1,0,1$, in the timelike, null, spacelike case respectively). The equations are not dynamical as $r = 0$ represents a true singularity of the equations and, in fact, do not have global existence of solutions - see section 6.1). This agrees with the fact that $r = 0$ is a region of dense singular conjugate points.

Having made this distriction between various types of conjugate points, we are now in a position to look at the singularity theorems in a new light.

4.3 The Singularity Theorems

It is clear from the previous section that the existence of

conjugate points alone does not imply that certain geodesics are incomplete. Hence the reason for imposing extra conditions in the singularity theorems, such as the existence of a closed trapped surface. In fact, in all the singularity theorems, the existence of a 'trapping' effect on the geodesics is assumed. (In theorem 2.4-2, the existence of a compact achronal set without edge implies the existence of a future/past trapped set, that is, a set ζ which is achronal and such that $E^+(\zeta)/E^-(\zeta)$ is compact).

As stated in Chapter two, if the Einstein field equations hold, the weak and strong energy conditions on the energy-momentum tensor imply that the null and timelike convergence conditions hold. However, under the assumptions of the present chapter, the field equations may only approximately hold. For instance, dynamical amendments to singular models or amendments possessing global existence of solutions, may not exactly satisfy the field equations. In these cases,

$$R_{ab} - \frac{1}{2} g_{ab}R = 8\pi T_{ab} + K_{ab} \qquad (4.3\text{-}1)$$

where K_{ab} is some correction tensor, at any point. (so that $K_{ab} \doteq 0$ in the weak field region). Note however, that certain dynamical models, such as the one derived from the Minkowski metric satisfy the field equations and there is no need for amendment. In these cases, $K_{ab} = 0$, at all points.

The Null Vector Case

Let ω be an arbitrary null vector at $m \in M$ with components with respect to the co-ordinate basis, ω^a. Then,

$$R_{ab}\omega^a\omega^b = \left\{8\pi T_{ab} + K_{ab} + \frac{1}{2} g_{ab}R\right\} \omega^a\omega^b \Rightarrow$$

$$R_{ab}\omega^a\omega^b \geq K_{ab}\omega^a\omega^b$$

using the weak energy condition. There is no guarantee that $K_{ab}\omega^a\omega^b$ should be positive or zero at or near the predicted singularity. Clearly the null convergence condition only holds if the corrected weak energy condition,

$$(8\pi T_{ab} + K_{ab})\omega^a\omega^b \geq 0 \qquad (4.3\text{-}2)$$

holds for all null vectors ω.

Using proposition 2.2-4, we have,

Lemma 4.3-1

There exist conjugate points in congruence of null geodesics when M is geodesically complete if the corrected weak energy condition (4.3-2) holds and the generic condition holds.

□

The Timelike Vector Case

Let ω be an arbitrary timelike vector at $m \in M$, with components ω^a. Then, $$R_{ab}\omega^a\omega^b = \left\{ \frac{1}{2} g_{ab}R + 8\pi T_{ab} + K_{ab} \right\} \omega^a\omega^b$$

$$\therefore \quad R_{ab}\omega^a\omega^b = \frac{1}{2} R\omega^a\omega_a + 8\pi\, T_{ab}\omega^a\omega^b + K_{ab}\omega^a\omega^b. \qquad (4.3\text{-}3)$$

by (4.3-1). Contracting the field equations (4.3-1), it is clear that,

$$-R = 8\pi T + K \qquad (4.3\text{-}4)$$

and substituting (4.3-4) into (4.3-3) we obtain,

$$R_{ab}\omega^a\omega^b = -\frac{1}{2} (8\pi T + K)\omega^a\omega_a + (8\pi T_{ab} + K_{ab})\omega^a\omega^b,$$

so that the corrected strong energy condition,

$$(8\pi T_{ab} + K_{ab})\omega^a\omega^b \geq \frac{1}{2} (8\pi T + K)\omega^a\omega_a \qquad (4.3\text{-}5)$$

implies that the timelike convergence condition holds. Of course, if $K_{ab} = 0$, (4.3-5) reduces to the strong energy condition. Using proposition 2.2-4, the following lemma may be deduced.

Lemma 4.3-2

There exist conjugate points in congruences of timelike geodesics when M is geodesically complete if the corrected strong energy condition holds on M and the generic condition holds.

□

Of course, the conjugate points predicted by lemma 4.3-1, lemma 4.3-2, must be regular, or if singular, not reached.

In the free space case, that is, $T^{ab} = 0$, if the Einstein equations hold, the null and timelike convergence conditions are trivially satisfied. This is not true in the case $K_{ab} \neq 0$. In chapter six, an example of a dynamical amendment to the Schwarzchild equations will be given which does not satisfy the corrected weak and strong energy conditions.

Notice that lemma 4.3-1 and 4.3-2 basically say that the corrected energy conditions imply that the non-spacelike geodesic convergence condition holds.

Examining the conditions of theorem 2.4-1, we have, from previous comments,

Lemma 4.3-3

If the corrected weak energy condition holds at each point of M, and M contains a closed trapped surface, then there exist points conjugate to the surface. In addition, these points must be regular or singular. □

Proof

By previous analysis, if the corrected weak energy condition holds, then the null convergence condition must hold and so there exist points conjugate to the closed trapped surface by proposition 2.2-3 and these must be reached in a finite value of the affine parameter, as required. ■

Of course, if the geodesics begin to converge at any surface, there will be conjugate points (which will again be regular or singular), but with no 'trapping' effect present, there is nothing to stop the geodesics diverging again and so in most cases, the predicted conjugate points will all be regular. If a closed trapped surface is present, even the 'outgoing' null geodesics are dragged back and the predicted conjugate points are more likely to be singular.

Notice that the two conditions of lemma 4.3-3 alone are not enough

to ensure that the predicted conjugate points are singular.[4] For instance, the 'outgoing' geodesics may get dragged back to a conjugate point different from the first and so on and then all these conjugate points could be regular. This sort of behaviour cannot occur if there is a spacelike hypersurface which every null geodesic intersects exactly once, that is, a Cauchy surface. Hence, we have the following equivalent to theorem 2.4-1:-

Theorem 4.3-4

If the corrected weak energy condition holds at each point of M, and there is a closed trapped surface and a Cauchy surface in M, then there exist singular conjugate points (with respect to the closed trapped surface). □

If the geodesic equations are dynamical, then singular conjugate points cannot occur, so that one of the three conditions of theorem 4.3-4 must break down. It has already been stated that the corrected weak energy condition may not hold in this case. Even if it did, there is no reason to suppose the existence of a Cauchy surface or the existence of a closed trapped surface if the field equations do not hold, so that departures from spherical symmetry are greater than expected. Even if the geodesic spray is complete, but the geodesic equations are not dynamical, no singular conjugate points can occur as in this case, any such points would be reached in a finite value of the affine parameter.

Recall that the timelike analogue to theorem 2.4-1 is theorem 2.4-4. In contrast to the above case, timelike geodesic incompleteness comes about without assuming the existence of a Cauchy surface. This is because a closed trapped surface may not be a partial Cauchy surface, whereas its three-dimensional analogue can always represent one, in effect. Thus, we have,

Theorem 4.3-5

If the corrected strong energy condition holds at each point of M and M contains a compact spacelike three-surface ζ such that unit

normals to ζ are everywhere converging, then there exist singular conjugate points (with respect to the given surface). □

Notice that in the case of lemma 4.3-3, the null geodesics entering $\mathcal{S}$ (the closed trapped surface) will not have a future end point if all the conjugate points are regular (so that there is no Cauchy surface). Thus, the null geodesics are imprisoned within $\mathcal{S}$, and by proposition 4.1-2, this cannot occur if the strong causality condition holds. Thus, we have the following alternative to theorem 4.3-4:-

Theorem 4.3-6

If the corrected weak energy condition and strong causality condition hold on M and there is a closed trapped surface, then there exist singular conjugate points (with respect to the closed trapped surface). □

This can be seen to be equivalent to part of theorem 2.4-2. Considering the first two conditions of the theorem 2.4-2, we have,

Lemma 4.3-7

If both the corrected weak and strong energy conditions hold on M and the generic condition holds on M, then there are conjugate points (which may be regular, singular) in congruences of null and timelike geodesics. □

Proof

By previous comments the corrected weak and strong energy conditions imply that the non-spacelike geodesic convergence condition holds and so by proposition 2.2-4, the result follows. ■

It was stated in Chapter Two, that the chronology and causality conditions are equivalent in physically realistic situations. This is only true if the field equations are assumed to hold exactly. There is no guarantee in the amended cases where the weak and strong energy

conditions do not hold that the chronology and causality conditions are equivalent. The precise statement is,

Proposition 4.3-8[5]

If (a) the null geodesic convergence condition holds,
(b) the generic condition holds on M; and
(c) the chronology condition holds on M,
then the causality condition holds on M. If, in addition, M is null geodesically complete, then the strong causality condition holds on M. □

Thus, the chronology and generic conditions do not, in general imply that the strong causality condition holds. This will only be the case if the corrected weak energy condition holds and the geodesic spray is complete. However, if the strong causality condition is assumed, then we have the following corollary to lemma 4.3-7:-

Corollary 4.3-9

If both the corrected weak and strong energy conditions hold on M and the generic and strong causality conditions also hold then there exist distinct regular or singular conjugate points in congruences of null and timelike geodesics. (That is, no point may be conjugate to itself). □

The case where a closed trapped surface is assumed to exist in theorem 2.4-2, has already been dealt withm in the form of theorem 4.3-6 and so this leaves the remaining hypotheses of theorem 2.4-2 to be dealt with (i.e. conditions (a) and (c)). In fact, all three alternatives can be combined into one hypothesis, due to the following lemma:-

Lemma 4.3-10[6]

If the non-spacelike geodesic convergence condition holds and there exist one of the following in M:-

(a) a compact achronal set without edge,

(b) a closed trapped surface,

(c) a point p such that the null geodesics from p are focused by the matter or curvature and start to reconverge;

then there is an achronal set ζ such that $E^+(\zeta)$ or $E^-(\zeta)$ is compact. □

(If (a) holds, ζ is the achronal set; if (b) holds ζ is the closed trapped surface; and if (c) holds, ζ is the point p). An achronal set ζ with $E^+(\zeta)$ or $E^-(\zeta)$ compact, is usually called future trapped or past trapped respectively.

Similar to the case of theorem 2.4-1, the conditions of corollary 4.3-9 alone are not enough to ensure the existence of singular conjugate points, without a 'trapping' effect, provided by one of the three alternatives of lemma 4.3-10. Notice that the achronal set ζ in lemma 4.3-10, can be thought of as being almost a partial Cauchy surface - it is a non-timelike surface which no timelike curve intersects more than once. Also, if $E^+(\zeta)$ (or $E^-(\zeta)$ is compact), non-spacelike curves through ζ, which are not timelike must have a future (or past) end-point, which must be singular or critical. Thus, these particular curves must contain singular points (noting that points cannot be conjugate to themselves). As null geodesics containing conjugate points can be varied to give timelike curves (as stated in Chapter Two), the timelike geodesics must contain conjugate points, of the singular kind. This is a fairly loose justification for the results of theorem 2.4-2, which can be translated as,

Theorem 4.3-11

If,

(i) the corrected weak and strong energy conditions hold on M,

(ii) the generic condition is satisfied on M,

(iii) the strong causality condition holds on M, and

(iv) there is a future or past trapped set on M, then

M contains singular conjugate points.

□

Once again, if the geodesic equations are dynamical, then one of more of the above conditions cannot hold. By proposition 4.3-8, if the causality condition holds then condition (iii) will hold and if the

space-time is not too symmetric, the generic condition will hold. However, there is no guarantee that the corrected energy conditions hold, and even if they do, there can be no 'trapping' effect on the geodesics, caused by the presence of a trapped set. If the geodesic spray is complete, but the geodesic equations are not dynamical any singularity present cannot be reached and so once again, there can be no singular conjugate points. Such a singularity (that is, one which does not consist of conjugate points) does not hinder the progress of any geodesics and so the situation is really equivalent to having no singularity present at all - none of the singularity theorems predict the existence of such a singularity.

As stated in Chapter Two, theorem 2.4-3 is just a slightly strengthened version of theorem 2.4-2 - it is almost equivalent to strengthening condition (iv) of theorem 4.3-11 by requiring that there exists a past trapped set in M.

The theorems presented in this section predict singular conjugate points in congruence of geodesics on a particular four-dimensional manifold. Notice however, that it may be possible to generalize the results to any given manifold and any set of curves. Looking at the theorems from the point of view of differential equations may give greater insight into the methods of generalization than the topological methods used to prove the singularity theorems of this section.

Chapter Five:

Models of Gravity Approximating to the Newtonian Equations

This chapter sets out to find various amendments to the non-relativistic Newtonian equations of gravity in one and two dimensions. Various properties of the amended equations are investigated, in particular, the change in character of what was the singularity of the Newtonian equations.

The model for the universe is taken to be the flat manifold $\mathbb{R}^3$, where each point has three space co-ordinates only and time is the parameter along the orbit of any observer. In this case, a test particle will deviate from a geodesic (in $\mathbb{R}^3$ a straight line) when acted on by the external gravitational force.

5.1) The One-Dimensional Newtonian Equation

Newton's equations are derived from Newton's second law of gravity and are well known to be, in three dimensions.

$$\frac{d^2\underline{r}}{dt^2} = -\nabla\phi \qquad (5.1\text{-}1)$$

where $\underline{r}$ is the radius vector of a test particle and ϕ represents a gravitational potential and is given by, where $r = |\underline{r}|$,

$$\phi = -\frac{c}{r} \qquad (5.1\text{-}2)$$

in free space, being the basic solution satisfying various boundary conditions, of Laplace's equation.[1] c is a constant, generally the product of two or more constants including the gravitational constant G. For the purpose of amending the equations, it is convenient to treat c as a single, positive constant, without loss of generality.

If the gravitational force is central, that is, it only acts in

the direction of the radius vector, (5.1-1) reduces to, in effect, Newton's equation in one dimension,

$$\ddot{r} = -\frac{c}{r^2} \qquad (5.1\text{-}3)$$

Clearly this autonomous equation is not dynamical as the right hand side is singular at $r = 0$. In fact, in this case we can prove that the equation does not have global existence of solutions as an explicit solution can be obtained as,

$$\frac{1}{a}\sqrt{ar^2 + 2cr} - \frac{2c}{a^{3/2}} \ln\left\{\sqrt{ar} + \sqrt{ar + 2c}\right\} = t + b \qquad (5.1\text{-}4)$$

$$a,b \in \mathbb{R}$$

So, for example, taking $a = 1$, $b = 0$, $r = 0 \Rightarrow -2c \ln(\sqrt{2c}) = t$ and so t is finite at the origin, that is, the singularity is reached in finite time.

5.2). Feasible Amendments of the One-Dimensional Newtonian Equation

Equation (5.1-3) is a member of a class of more general equations given by,

$$\ddot{r} = f(r) \qquad (5.2\text{-}1)$$

A sub-class of equations of the above type with $f(r)$ continuous are certainly dynamical. It is shown in this section that certain members of this sub-class are feasible amendments of (5.1-3). Generalizations of (5.2-1) are then considered, the most useful being of the Liénard type.

5.2.1) Equations of the Form $\ddot{r} = f(r)$

There are basically two ways in which (5.2-1) could be a viable amendment to (5.1-3). First of all, there may be models in which the singularity is still present, but is not reached, so that the equations are not dynamical but the solutions can be extended to arbitrary values of t. Secondly, there may be models in which the

singularity no longer occurs and which are dynamical. As stated in Chapter three, the condition that $f(r)$ is continuous is necessary for (5.2-1) to be dynamical, but not sufficient. For example, $\ddot{r} = r$ is dynamical, whereas $\ddot{r} = r^2$ is not dynamical. (These results can be verified by solving each equation explicitly).

Example 5.2-1

Consider the equation,

$$\ddot{r} = \frac{-c}{r^2 + a^2} \; ; \; a,c \in \mathbb{R} \qquad (5.2\text{-}2)$$

which, as a system can be written

$$\begin{cases} \dot{r} = p \\ \dot{p} = \dfrac{-c}{r^2 + a^2} \end{cases} \qquad (5.2\text{-}3)$$

There are critical points at $(\pm \infty, 0)$. The only singularity occurs in the region of infinite p.

The equation (5.2-2) has a first integral given by,

$$p = \sqrt{\frac{-2c}{a} \tan^{-1}\left(\frac{r}{a}\right) + b}$$

where b is a constant of integration. The right hand sides of (5.2-3) are continuous and have continuous partial derivatives and so any critical points are reached in infinite time by theorem 3.3-2. Also, p is a well-behaved function of r and never becomes infinite. As the critical points and singularities are not eached in finite time and it is precisely these which curtail the progress of a solution, the solutions can be extended for all t, that is, the system (5.2-3) is dynamical. Notice if $|a| \ll 1$, (5.2-2) approximates to (5.1-3).

The possibility of having models in which a singularity is not reached in finite time is now considered. A couple of preliminary

theorems concerning improper integrals are first needed.

Recall that $\int_a^b F(x)dx$ is a Cauchy improper integral of the first kind if a or b are infinite.[2] Suppose that b is infinite. The integral is defined as $\lim_{s\to\infty} \int_a^s F(x)dx$ (similarly for a). If the limit exists, the integral is convergent, otherwise it is divergent. There are various tests for convergence, the most useful being,

Theorem 5.2-1[2]

Let $\lim_{x\to\infty} x^n F(x) = A$. Then,

(i) $\int_a^\infty F(x)dx$ converges if $n > 1$ and A is finite.

(ii) $\int_a^\infty F(x)dx$ diverges if $n \leq 1$ and $A \neq 0$ (possibly infinite). □

If the integral of $\int_a^b F(x)dx$ becomes infinite at $x = c$, $a \leq c \leq b$, the integral is a Cauchy improper integral of the second kind.[3] The sum, $\int_a^{c-\varepsilon} F(x)dx + \int_{c+\varepsilon}^b F(x)dx$ is considered and if the limit as $\varepsilon \to 0$ of the sum is finite, the integral is convergent at $x = c$. Otherwise it is divergent. Again, there are various tests of convergence, the most important being,

Theorem 5.2-2[3]

If in, $\int_a^b F(x)dx$, $F(x)$ becomes infinite at $x = c$, $a \leq c \leq b$,

let $\lim_{x\to c^+} (x - c)^n F(x) = A$.

(i) $\int_a^b F(x)dx$ converges if $n < 1$ and A is finite.

(ii) $\int_a^b F(x)dx$ diverges if $n \geq 1$ and $A \neq 0$ (possibly infinite).

□

(If $F(x)$ becomes infinite at $x = b$, $A = \lim_{x \to b^-} (b-x)^n F(x)$).

Lemma 5.2-3

If, in the equation (5.2-1), $f(r)$ becomes infinite at some finite value $r = a$ and $\int f(r)dr \nrightarrow -\infty$ as $r \to a$, then not all solutions can be extended to arbitrary values of t.

□

Proof

(5.2-1) has a first integral given by,

$$\dot{r} = \sqrt{2\int f(r)dr + K} \Rightarrow \int \frac{dr}{\sqrt{2\int f(r)dr + K}} = t + t_o \qquad (5.2\text{-}4)$$

The singularity at $r = a$ will always be reached in finite time, except possibly when $2\int f(r)dr\Big|_a + K = 0$, if the integrand is real in a neighbourhood of $r = a$, as for most solutions, the integral in (5.2-4) is perfectly well behaved in a region including $r = a$ ($\int_{r_1}^{r_2} f(r)dr$, $r_1 \leqslant a \leqslant r_2$ being either convergent or divergent). If $\int f(r)dr \nrightarrow -\infty$ as $r \to a$, there will always be some choice of K making the root positive in a neighbourhood of $r = a$.

If $2\int f(r)dr\Big|_a + K = 0$, $\int_{r_1}^{r_2} f(r)dr$ must be convergent and

$\int_{r_1}^{r_2} \frac{dr}{\sqrt{2\int f(r)dr+K}}$ is an improper integral of the second kind.

However, the latter integral is convergent when $\lim_{r \to a} \frac{(r-a)^n}{\sqrt{2\int f(r)dr + K}}$

is finite for some $n < 1$, by theorem 5.2-2 and this will always be the case due to the hypothesis on $f(r)$.

Thus, the only way in which the singularity cannot be reached in finite time, for all values of K, is if the integrand is complex in an arbitrarily small neighbourhood of $r = a$ for any K, that is, if

$\int f(r)dr \to -\infty$ as $r \to a$. ∎

From the above lemma, (5.1-3) does not have global existence of solutions as already known, but $\ddot{r} = \frac{2}{r^3}$ does, even though it is not dynamical, (if the possibility of starting from r =0 is excluded).

Theorem 5.2-4

(a) The equation $\ddot{r} = f(r)$ is dynamical if

(i) f(r) and f'(r) are continuous; and

(ii) if f(r) becomes singular at infinity, that is, $\lim_{r \to \pm\infty} f(r)$ does not exist and $\int f(r)dr \nrightarrow -\infty$ as $r \to \pm\infty$, then $\lim_{r\to\pm\infty} f'(r)$ must be finite.

(b) The equation $\ddot{r} = f(r)$ is not dynamical if either

(i) f(r) and f'(r) are discontinuous; or

(ii) if f(r) becomes singular at infinity and $\int f(r)dr \nrightarrow -\infty$ as $r \to \pm\infty$, then $\lim_{r\to\pm\infty} \frac{r^n}{f'(r)}$, $n > 0$ is finite. □

Proof

Condition (i) of (a) is necessary from section 3.1). Thus, if (i) fails to hold, the equation cannot be dynamical. If (a)(i) holds, then by theorem 3.3-2 , all critical points of (5.2-1) are reached in infinite time. This result can also be proved by considering the divergence of the integral in (5.2-4).

As f(r) is assumed to be continuous, the only possible singular regions occur at infinity. Suppose, without loss of generality, that $\lim_{r\to\infty} f(r)$ does not exist, (the analysis being exactly similar if $\lim_{r\to-\infty} f(r)$ does not exist).

From theorem 5.2-1, with n = 1, it is clear that $\int_{r_1}^{\infty} f(r)dr$ is divergent, so that, from (5.2-4), $p \to \infty$ as $r \to \infty$, if the root is real. (If $\int f(r)dr \to -\infty$ as $r \to \infty$, the root is not real and so the singularity cannot be reached. In this case there is no problem).

If the root is real, $\int_{r_1}^{\infty} \frac{dr}{\sqrt{2\int f(r)dr + K}}$ is an improper integral of the first kind, which may or may not be convergent.

$$\text{Let } A = \lim_{r\to\infty} \frac{r^n}{\sqrt{2\int f(r)dr + K}} = \sqrt{\lim_{r\to\infty} \frac{r^{2n}}{2\int f(r)dr + K}} \qquad (5.2\text{-}5)$$

provided that the limit exists.

Let $L = \lim_{r\to\infty} \frac{r^{2n}}{2\int f(r)dr + K}$. By hypothesis on $f(r)$, this is of a form suitable for the application of l'Hopital's rule.

$$\therefore \quad L = \lim_{r\to\infty} \frac{2n\, r^{2n-1}}{f(r)} = \lim_{r\to\infty} \frac{2n(2n-1)\, r^{2n-2}}{f'(r)} \qquad (5.2\text{-}6)$$

Then from (5.2-5), (5.2-6), by theorem 5.2-1, $\int_{r_1}^{\infty} \frac{dr}{\sqrt{2\int f(r)dr + K}}$ converges if $n > 1$ and L is finite and so the singularity is reached in finite time (as in case (b)) and diverges if $n \leq 1$ and $L \neq 0$, so that the singularity is reached in infinite time. (as in case (a)).

As critical points and singularities are the only points which curtail the progress of a solution, the result follows. ∎

Notice that $\ddot{r} = r$ and (5.2-2) are dynamical by theorem 5.2-4, as stated earlier and that $\ddot{r} = r^2$ is not dynamical using the same theorem (choose $n = 1$ in (b) (ii)).

In order that (5.2-1) is a viable amendment to (5.1-3), it must reduce to (5.1-3) in the weak field region, that is, $f(r) \to 0$ as $r \to \pm\infty$, so that there are no singularities at infinity. Thus by theorem 5.2-4, (5.2-1) will be dynamical, provided $f(r)$ and $f'(r)$ are continuous.

Now from Lemma 5.2-3, the only way in which (5.2-1) can be a feasible amendment to (5.1-3) if $f(r)$ is singular at $r = 0$, is if $\int f(r)dr \to -\infty$ as $r \to 0$. In this type of model, any orbit approaching the origin must come to a halt at some point $r = a$, $a > 0$ and then start going backwards. Thus, these models do not seem to be satisfactory

amendments to (5.1-3).

Example 5.2-2

Consider the equation

$$\ddot{r} = \frac{-c}{r^2} + \frac{a}{r^4} \; ; \quad |a| \ll 1 \tag{5.2-7}$$

In the weak field region, it reduces to (5.1-3), but

$\int f(r)dr = \left\{\frac{c}{r} - \frac{a}{3r^3}\right\}$ and so is negative in a small enough neighbourhood of $r = 0$ ($r^2 < a/3c$) and $\int f(r)dr \to -\infty$ as $r \to 0$. Thus, the singularity is never reached. All solutions of (5.2-7) are complete by lemma 5.2-3. Notice $\ddot{r} = 0$ when $r^2 = \frac{a}{c}$ and $\dot{r} = 0$ when $K = \frac{2a}{3r^3} - \frac{2c}{r}$ and this can always be satisfied.

Notice, in these types of models, at some stage the gravitational force becomes repulsive.

From now on, this work concerns itself with dynamical amendments of the one-dimensional Newtonian equation, so that the origin is actually reached, even if this is done in infinite time, (that is, orbits in phase space pass through $r = 0$).

5.2.2) Dynamical Amendments to the One-Dimensional Newtonian Equation of the type $\ddot{r} = f(r)$

In this sub-section, various examples are given of possible dynamical amendments to the one-dimensional Newtonian equation. These can basically be divided into two types. In the first type of model, $f(r)$ is (5.2-1) becomes very large near $r = 0$, but never becomes singular. In the second type of model $f(r)$ may not be large at the origin but becomes large in some neighbourhood of it. In both types of model, $f(r) \to \left(\frac{-c}{r^2}\right)$ as $r \to \infty$, so that they reduce to (5.1-3) in the weak field region.

Example 5.2-3

Consider the equation $\ddot{r} = \frac{-2c}{\pi} \frac{\tan^{-1}(br^n)}{r^2}$; $b,c,n \in \mathbb{R}$ (5.2-8)

As $r \to \infty$, $\tan^{-1}(br^n) \to \frac{\pi}{2}$ and $\ddot{r} \to \frac{-c}{r^2}$

Now $\lim_{r\to 0} \frac{-2c}{\pi} \frac{\tan^{-1}(br^n)}{r^2} = \lim_{r\to 0} - \frac{bcn\, r^{n-2}}{\pi(1+b^2r^{2n})}$

by l'Hopital's rule and so there is no singularity at the origin if $n \geq 2$. In fact, $\ddot{r} = 0$ at $r = 0$ only if $n > 2$ and $\ddot{r} = \frac{-bcn}{\pi}$ at $r = 0$ when $n = 2$.

Thus, by theorem 5.2-4, (5.2-8) is a dynamical amendment of (5.3-1) if $n \geq 2$. If $n = 2$, the first integral of (5.2-8) is given by,

$$(\dot{r})^2 = \frac{4c}{\pi}\left\{ \frac{\tan^{-1}(br^2)}{r} + \sqrt{\frac{b}{2}}\left[\frac{-1}{2}\ln\left(\frac{br^2 + \sqrt{2}br + 1}{br^2 - \sqrt{2}br + 1}\right) + \tan^{-1}\left(\frac{\sqrt{2}br}{br^2 - 1}\right)\right]\right\} + K \qquad (5.2\text{-}9)$$

and, if $n = 3$, the first integral is,

$$(\dot{r})^2 = \frac{4c}{\pi}\left\{\frac{\tan^{-1}(br^3)}{r} - \frac{b}{2}^{1/3}\left[\frac{1}{2}\ln\left(\frac{b^{4/3}r^4 + 2b^{2/3}r^2 + 1}{b^{4/3}r^4 - b^{2/3}r^2 + 1}\right)\frac{\pm 3}{\sqrt{3}}\right.\right.$$
$$\left.\left.\tan^{-1}\left(\frac{2b^{2/3}r^2 - 1}{\sqrt{3}}\right)\right]\right\} + K \qquad (5.2\text{-}10)$$

using integration by parts and standard integrals. With increasing values of n, integration becomes much more difficult - neither it is easy with fractional values of n.

Now for any particular value of r, $\lim_{b\to\infty} \frac{\tan^{-1}(br^n)}{r^2} = \frac{\pi}{2r^2}$

and so, in the limit as $b \to \infty$, the right hand side of (5.2-8) reduces to the Newtonian case. It is desirable to have the amended first integrals reducing to the Newtonian first integral in this limit.

In the case n = 2, it is easily seen that

$$\lim_{b\to\infty} \ln\left(\frac{br^2+\sqrt{2}br+1}{br^2-\sqrt{2}br+1}\right) = 0 \quad \text{and} \quad \lim_{b\to\infty} \tan^{-1}\left(\frac{\sqrt{2}br}{b^2r^2-1}\right) = 0.$$

Applying l'Hopital's rule to (5.2-9) gives $\lim_{b\to\infty} (\dot{r})^2 = \frac{2c}{r} + K$ as required. This will not always be the case.

For $n = 3$, $\lim_{b\to\infty} \ln \frac{b^{4/3}r^4 + 2b^{2/3}r^2 + 1}{b^{4/3}r^4 - b^{2/3}r^2 + 1} = 0$ but

$$\lim_{b\to\infty} \tan^{-1}\left(\frac{2b^{2/3}r^2 - 1}{\sqrt{3}}\right) = \frac{\pi}{2}$$ and so, in this case,

$$\lim_{b\to\infty} (\dot{r})^2 \neq {}^{2c}/_{r} + K.$$

Notice that $-\int f(r)dr = \frac{2c}{\pi}\int \frac{\tan^{-1}br^n}{r^2}\,dr$ represents the amended Newtonian potential, and in both cases (n=2 and n=3) tends to zero or a constant as $r \to \infty$ and so the essential character of the potential as $r \to \infty$ is retained, whereas the singular character at $r = 0$ is erased.

Example 5.2-4

The equation,

$$\ddot{r} = -c\,\frac{\tanh(br^n)}{r^2}; \quad b,n \in \mathbb{R} \qquad (5.2\text{-}11)$$

has similar qualitative behaviour to (5.2-8).

As $r \to \infty$, $\tanh(br^n) \to 1$ $\therefore$ $\ddot{r} \to \frac{-c}{r^2}$ as $r \to \infty$ as required.

Also, $\lim_{r\to 0} -c\,\frac{\tanh(br^n)}{r^2} = -\frac{bcn}{2}\lim_{r\to 0} r^{n-2}$, so that there is no singularity at the origin if $n \geq 2$ and in this case, (5.2-11) is dynamical by theorem 5.2-4. Again, $\ddot{r} = 0$ at $r = 0$ if and only if $n > 2$.

However, with this particular equation, the first integral is far less tractible than in the previous example, for any n.

Example 5.2-5

An example of an amendment of (5.1-3) possessing a different qualitative character in a neighbourhood of the origin is provided by,

$$\ddot{r} = \frac{-cr^{2n-2}}{r^{2n} + a^{2n}} \; ; \quad a \in \mathbb{R}_{>0}, \quad n \in \mathbb{Z}_{\geq 0} \; . \tag{5.2-12}$$

Again, by theorem 5.2-4, (5.2-12) is dynamical and $\ddot{r} = 0$ when $r = 0$ if and only if $n > 1$. As $r \to \infty$, $\ddot{r} \to \frac{-c}{r^2}$ as required. Clearly, as smaller and smaller values of a are taken, the closer (5.2-12) approximates to (5.1-3).

If $n = 1$, the first integral of (5.2-12) is,

$$(\dot{r})^2 = \frac{-2c}{a} \tan^{-1} \frac{r}{a} + K \tag{5.2-13}$$

and if $n = 2$, the first integral is given by,

$$(\dot{r})^2 = c \left\{ \frac{1}{a\sqrt{2}} \tan^{-1} \left(\frac{ar\sqrt{2}}{r^2 - a^2} \right) - \frac{1}{2a\sqrt{2}} \ln \left(\frac{r^2 - a\sqrt{2}r + a^2}{r^2 + a\sqrt{2}r + a^2} \right) \right\} + K \tag{5.2-14}$$

Again, with increasing values of n, integration becomes more difficult, but in this case is possible.[4] Notice if $n = 1$, $\lim_{a \to 0} (\dot{r})^2 \neq \frac{+c}{r}$, but $\lim_{r \to \infty} (\dot{r})^2$ is a constant. In the case $n = 2$,

$\lim_{a \to 0} \tan^{-1} \left(\frac{ar\sqrt{2}}{r^2 - a^2} \right) = \lim_{a \to 0} \ln \left(\frac{r^2 - a\sqrt{2}r + a^2}{r^2 + a\sqrt{2}r + a^2} \right) = 0$ and using l'Hopital's rule on (5.2-14) gives $\lim_{a \to 0} (\dot{r})^2 = \frac{2c}{r} + K$ as required. Also, $\lim_{r \to \infty} (\dot{r})^2 = K$ as in the Newtonian case.

Example 5.2-6

Consider the equation,

$$\ddot{r} = \frac{-cr}{(r^2 + a^2)^{3/2}} \; ; \quad a, c \in \mathbb{R}_{>0}. \tag{5.2-15}$$

It has all the required properties of a feasible dynamical amendment to (5.1-3). In particular, $\lim_{r\to\infty} \ddot{r} = 0$ and as $r \to \infty$, $\ddot{r} \to \frac{-c}{r^2}$, $\lim_{a\to 0} \ddot{r} = \frac{-c}{r^2}$.

In this case, a first integral is particularly simple, being given by

$$(\dot{r})^2 = \frac{2c}{(r^2 + a^2)^{\frac{1}{2}}} + K \qquad (5.2\text{-}16).$$

This clearly reduces to the Newtonian case as a tends to zero.

Example 5.2-7

In all the previous examples, an amended equation was considered and the first integral found. A slightly different approach is to start with a non-singular first integral which approximates to the Newtonian case and derive the amended second order equation.

For instance, consider the first integral of the form,

$$(\dot{r})^2 = \frac{2c\, r^{2n-1}}{(r^{2n} + a^{2n}} + K; \quad a \in \mathbb{R}_{>0}, \quad n \in \mathbb{Z}_{\geq 0} \qquad (5.2\text{-}17)$$

As $r \to \infty$, $(\dot{r})^2 \to \frac{2c}{r} + K$ and as $\lim_{a\to 0} (\dot{r})^2 = \frac{2c}{r} + K$.

Differentiating (5.2-17) gives a second order equation of the form,

$$\ddot{r} = \frac{c(2n-1)a^{2n} r^{2n-2} - cr^{4n-2}}{(r^{2n} + a^{2n})^2} \qquad (5.2\text{-}18)$$

By hypothesis, this is a dynamical amendment to (5.1-3) by theorem 5.2-4. Also, $\lim_{a\to 0} \ddot{r} = \frac{-c}{r^2}$ and $\lim_{r\to\infty} \ddot{r} = 0$ as required.

5.2.3) Equations of the form $\ddot{r} = f(r) + \alpha g(r, \dot{r})$

The amendments so far considered are not sufficiently general to allow certain desirable properties. In particular, they all have first integrals and are structurally unstable. Thus, more general amendments

to the Newtonian equation (5.1-3), of the type,

$$\ddot{r} = f(r) + \alpha g(r,\dot{r}) \qquad (5.2\text{-}19)$$

are considered, where α is a small constant. (5.2-19) includes the special case of (5.2-1), but does not, in general, have a first integral. There are two basic questions to be answered. Firstly, if $f(r) = \frac{-c}{r^2}$ as in (5.1-3), can $g(r,\dot{r})$ be chosen so that (5.2-19) has global existence of solutions? Secondly, if for $\alpha = 0$, (5.2-19) is dynamical, what class of functions do $g(r,\dot{r})$ belong to so that (5.2-19) remains dynamical for $\alpha \neq 0$?

It is assumed that $g(r,\dot{r})$ is not a function of r only, as otherwise the equation is of the previous form. Notice that a necessary condition for (5.2-19) to be dynamical is that both $f(r)$ and $g(r,\dot{r})$ are continuous (g is both r and $\dot{r}$).

Example 5.2-8

Consider the following special case of (5.2-19):-

$$\ddot{r} = f(r) + \alpha(\dot{r})^2 \qquad (5.2\text{-}20)$$

In this case, the equation possesses a first integral which is,

$$(\dot{r})^2 = 2e^{2\alpha r}\left\{\int f(r)\, e^{-2\alpha r}\, dr + K\right\} \Longrightarrow$$

$$\int \frac{dr}{\sqrt{2e^{2\alpha r}\left[\int f(r)e^{-2\alpha r}dr + K\right]}} = t + t_o \qquad (5.2\text{-}21).$$

where K and t_o are constants of integration.

Suppose that $f(r)$ becomes singular at some finite value of $r = a$ say. Then $\int f(r)e^{-2\alpha r}dr$ is an improper integral of the second kind in a neighbourhood of $r = a$. Whether or not this is convergent, the integrand in (5.2-21) is perfectly well-behaved for almost all values of K in a neighbourhood of $r = a$ and so if the root is real in this neighbourhood, the singularity is reached in finite time. Thus, if

(5.2-20) does not have global existence of solutions when $\alpha = 0$, it does not have global existence of solutions for the given choice of $g(r,\dot{r})$ when $\alpha \neq 0$.

Suppose now then that (5.2-20) is dynamical for $r = 0$, so that $f(r)$ is continuous and if $\lim_{r\to\pm\infty} f(r)$ does not exist, $\lim_{r\to\pm\infty} f'(r)$ is finite ($f'(r)$ being continuous), if $\int f(r)dr \to -\infty$ as $r \to \pm\infty$. Whether or not $\lim_{r\to\pm\infty} f(r)$ is finite, $\lim_{r\to\pm\infty}\left[2e^{2\alpha r}\left\{\int f(r)\, e^{-2\alpha r}\, dr + K\right\}\right]$ will be infinite for almost all choices of K (it will be infinite for all choices of K if $\lim_{r\to\pm\infty} f(r)$ is infinite, for if $\left(\int_R^{\infty} f(r)e^{-2\alpha r}dr + K\right) =$

$$\lim_{r\to\infty}\left[2e^{2\alpha r}\left\{\int f(r)e^{-2\alpha r}\, dr + K\right\}\right] = \lim_{r\to\infty}\frac{f(r)e^{-2\alpha r}}{-4e^{-2\alpha r}} \quad)$$

Now as $r \to \infty$, the integral of (5.2-21) is an improper integral of the first kind, if the root is real. By theorem 5.2-2, and l'Hôpital's rule, the integral will be convergent as the denominator will be infinite, independent of the number of times of differentiation from the above comments. Thus, from (5.2-21) $p = \dot{r}$ will always become infinite in finite time (for almost all values of K) and hence, even though (5.2-20) is dynamical for $\alpha = 0$, it is not for $\alpha \neq 0$, (if $\int f(r)dr \to -\infty$ as $r \to \pm\infty$).

Example 5.2-9

Consider now another special case of (5.2-19), namely,

$$\ddot{r} = f(r) + \alpha f(r)\dot{r} \qquad (5.2\text{-}22)$$

The equation has a first integral given by,

$$\frac{e^{\alpha\dot{r}}}{1+\alpha\dot{r}} = e^{\alpha^2 h(r)}, \text{ where } h(r) = \int f(r)dr + K. \qquad (5.2\text{-}23)$$

($\alpha \neq 0$). (K a constant of integration).

Equation (5.2-22) contains possible singularities at infinite r and $\dot{r}$, assuming that for $\alpha = 0$, the equation is dynamical. It will

only contain a singularity at infinite $\dot{r}$ if $\lim_{r\to\pm\infty} f(r)$ is finite.

Suppose in this case that $\lim_{r\to\pm\infty} \int f(r)dr$ is also finite, so that $\lim_{r\to\pm\infty} h(r)$ exists. Now $\lim_{p\to\pm\infty} \frac{e^{\alpha p}}{(1+\alpha p)} = \lim_{p\to\pm\infty} e^{\alpha p}$ and so is infinite.

Thus, if $\lim_{h\to\pm\infty} h(r)$ exists, by (5.2-23), $\dot{r}$ can never become infinite and so the equation is dynamical. However, if $\lim_{r\to\pm\infty} \int f(r)dr$ does not exist, even if $\lim_{r\to\pm\infty} f(r)$ is finite, the equation may not be dynamical.

Notice that the chosen $g(r,\dot{r})$ is bounded in the above example, when (5.2-22) is dynamical and that the chosen $g(r,\dot{r})$ in the first example is not bounded, for values taken in phase space, if $f(r) \nrightarrow -\infty$ as $r \to \pm\infty$.

For the rest of this section, it will be assumed that (5.2-19) is dynamical for $\alpha = 0$. It is seen from the examples above that the properties of (5.2-19) depend on the function $g(r,\dot{r})$ as well as on $f(r)$. Certain properties that $f(r)$ and $g(r,\dot{r})$ must possess in order that (5.2-19) is dynamical are now deduced.

Recall the definitions of closeness of analytical functions and dynamical systems (definitions 3.10-2, 3.10-3). For functions of two values, the usual norm is used, that is, if ϕ,ψ are real functions of two variables and are continuous in a region G of the x,y plane,

$$|\phi - \psi| = \sup_{(x,y)\in G} |\phi(x,y) - \psi(x,y)| \qquad (5.2\text{-}24)$$

Consider the two-dimensional system (5.2-1) in (r,p) space and the perturbed system (5.2-19).

Let $P(r,p) = p$ and $Q_\alpha(r,p) = f(r) + \alpha g(r,p)$. Denote $Q_0(r,p)$ by $Q(r,p)$. Clearly, any function is δ-close to rank s to itself for any s and so (5.2-19) is δ-close to rank s to (5.2-1) in G if Q_α is δ-close to rank s to Q in G, where G is a region of the (r,p) plane.

Theorem 5.2-5

If $g(r,p)$ is a continuous function of both values and is analytic or C^k and bounded for all values of r and p taken in phase space, then $Q_\alpha(r,p)$ is δ-close to rank s to $Q(r,p)$ and so (5.2-19) is δ-close to rank s to system (5.2-1).

□

Proof

$$\text{Now } |Q(r,p) - Q_\alpha(r,p)| = |f(r) + \alpha g(r,p) - f(r)| = |\alpha||g(r,p)|. \quad (5.2\text{-}25)$$

$$\text{and, } |Q^{(n_1+n_2)}_{r^{n_1}p^{n_2}} - Q^{(n_1+n_2)}_{\alpha\, r^{n_1}p^{n_2}}| = |f^{(n_1)}(r) - f^{(n_1)}(r) - \alpha g^{(n_1+n_2)}_{r^{n_1}p^{n_2}}(r,\bar{p})|$$

$$= |\alpha||g^{(n_1+n_2)}_{r^{n_1}p^{n_2}}(r,p)| \quad (5.2\text{-}26)$$

where $n_1 + n_2 = 1,2,\ldots,s$.

$g(r,p)$ is bounded for all values of r and p and so, $|g(r,p)| < M$ say, where $M \in \mathbb{R}_{>0}$. Then from (5.2-25),

$$|Q(r,p) - Q_\alpha(r,p)| \leq |\alpha| M.$$

Thus, if $|\alpha| M < \delta \Rightarrow |\alpha| < \frac{\delta}{M}$, Q_α is δ-close to rank 0 to Q.

If $g(r,p)$ is bounded and analytic, or C^k, $|g^{(n_1+n_2)}_{r^{n_1}p^{n_2}}| < N$, where $N \in \mathbb{R}_{>0}$, for $n_1+n_2 = 1,\ldots s. < k$. If $|\alpha| < \frac{\delta}{\max(M,N)}$, Q_α is δ-close to rank s to Q.

Hence the result. ∎

Having proved that (5.2-19) is δ-close to rank s to (5.2-1), if it is known that (5.2-1) is dynamical, it is an easy step to show (5.2-19) must also be dynamical, using the following theorem ('δ-close' meaning δ-close to rank 1' if no rank is stated).

Theorem 5.2-6[5]

Let $r = \phi(t-t_o, r_o, p_o)$, $p = \psi(t-t_o, r_o, p_o)$ be the solution of (5.2-1) corresponding to initial conditions t_o, r_o, p_o, that is, $\phi(0, r_o, p_o) = r_o$, $\psi(0, r_o, p_o) = p_o$. Let the solution be defined for all t in some interval $\tau < t < T$. Let τ_1, τ_2 be two numbers such that, $\tau < \tau_1 < t_o < \tau_2 < T$. If L is the orbit corresponding to the above solution, let the segment of L corresponding to $\tau_1 \leq t \leq \tau_2$ be completely contained in the closed region G and let $r = \phi^*(t - t_o, r_o^*, p_o^*)$ $p = \psi^*(t - t_o, r_o^*, p_o^*)$ be the solution of (5.2-19) corresponding to the initial conditions t_o, r_o^*, p_o^*. Then for any $\varepsilon > 0, \exists\, \eta > 0, \delta > 0$ such that if,

(i) $|r_o^* - r_o| < \eta$, $|p_o^* - p_o| < \eta$

(ii) system (5.2-19) is δ-close to system (5.2-1) in G, the solution of system (5.2-19) is defined for all t, $\tau_1 \leq t \leq \tau_2$, and in this time interval, $|\phi^* - \phi| < \varepsilon$ and $|\psi^* - \psi| < \varepsilon$.

□

If system (5.2-1) is known to be dynamical, the solution $r = \phi$, $p = \psi$ is defined for all t and so τ and T and hence τ_1, τ_2 can be chosen arbitrarily large. If (5.2-19) is δ-close to (5.2-1) in the whole (r,p) plane, then the region G can be made arbitrarily large, and by theorem 5.2-6, the solution $r = \phi^*$, $p = \psi^*$ can be defined in an arbitrarily large interval of t.

Thus, we have the following Corollary, using theorem 3.3-2.

Corollary 5.2-7

If, for $\alpha = 0$, (5.2-19) is dynamical, and g(r,p) is continuous and analytic or C^k and bounded for all values of r and p taken in phase space, then (5.2-19) is dynamical for $\alpha \neq 0$.

□

Consider again equation (5.2-20) of example 5.2-8. In this case,

$$g(r,\dot{r}) = g(\dot{r}) = \dot{r}^2 = 2e^{2\alpha r}\left\{\int f(r)e^{-2\alpha r}dr + K\right\} \qquad (5.2\text{-}26)$$

from (5.2-21). It was shown in the example that $r \to \infty$ in finite time for almost all values of K if $\dot{r}$ is real, so from (5.2-26), $g(r,\dot{r})$ can certainly become arbitrarily large in finite time. This agrees with the fact that (5.2-20) is not dynamical for $\alpha \neq 0$ in this case.

Consider now equation (5.2-22) of example 5.2-9. Here, $g(r,\dot{r}) = \dot{r}f(r)$. A necessary condition for $g(r,\dot{r})$ to be bounded is that $f(r)$ is bounded for all values of r taken in phase space. From the example, if $\lim_{r\to\pm\infty}\int f(r)dr$ is finite, $\dot{r}$ never becomes unbounded, so that by Corollary 5.2-7, the equation (5.2-22) must be dynamical, agreeing with previous results. This result is generalized in theorem 5.2-8.

One of the most useful sub-class of equations of the type (5.2-19) are equations of the Liénard type

$$\ddot{r} = f(r) + \alpha g(r)\dot{r} \qquad (5.2\text{-}27)$$

By Corollary 5.2-7, if (5.2-27) is dynamical for $\alpha = 0$, it will be dynamical for $\alpha \neq 0$, as long as $pg(r)$ is bounded and analytic (or at least C^1) for values of r and p taken in phase space. Clearly, for this to be so, $g(r)$ must be bounded and analytic (or at least C^1), for values of r taken in phase space. However, this condition, although necessary, is not sufficient.

Theorem 5.2-8

If, in (5.2-27), $f(r)$ and $g(r)$ are analytic and tend to zero as $r \to \pm\infty$, and, in addition, $\lim_{r\to\pm\infty}\int f(r)dr$ and $\lim_{r\to\pm\infty}\int g(r)dr$ are finite, then the equation is dynamical for all values of α.

□

Proof

The equation is dynamical for $\alpha = 0$ by theorem 5.2-4. Also, $f(r)$ and $g(r)$ must necessarily be C^1 in order that (5.2-27) is dynamical from theorem 3.3-2. As a first order system, the equation becomes,

$$\begin{cases} \dot{r} = p \\ \dot{p} = f(r) + \alpha p g(r) \end{cases} \qquad (5.2\text{-}28)$$

From Corollary 5.2-7, it is sufficient to show that p is also bounded for all r taken in phase space. Now from (5.2-28),

$$p\frac{dp}{dr} = f(r) + \alpha p g(r) \qquad (5.2\text{-}29)$$

If $f(r)$, $g(r)$ do not contain singularities for finite values of r, $p(r)$ cannot contain singularities for finite values of r. (To see this let $p = \sum_{n=0}^{\infty} a_n r^{n+\rho}$ where $a_n \in \mathbb{R}$ for each n and ρ may be fractional or negative. Substituting into (5.2-29), it is easily seen that $\rho = 0$, so that if f and g are analytic in r, so is p).

Thus, p(r) can only possibly become unbounded as $r \to \pm \infty$. Let $q = \sqrt{2\int f(r)dr + K + \alpha \int g(r)dr} \implies$

$$qq' = \frac{qf(r)}{(q - \alpha \int g(r)dr)} + \alpha g(r) q \qquad (5.2\text{-}30)$$

Thus, by hypothesis, (5.2-30) is asymptotic to (5.2-29). Notice also that $\lim_{r \to \pm\infty} q$ is finite and so p must also be finite as required. ∎

Of course, (5.2-27) may still be dynamical if not all the conditions of theorem 5.2-8 are satisfied. For instance, $f(r)$ and $g(r)$ need only be C^1. Also, in certain cases, $\lim_{r \to \pm\infty} \int g(r)dr$ need not be zero, so long as $r \to \pm \infty$ in infinite time. For example, $\ddot{r} = \frac{2\alpha r \dot{r}}{r^2 + a^2}$ is dynamical.

5.2.4) Dynamical Amendments to the One-dimensional Newtonian Equation of the type $\ddot{r} = f(r) + \alpha\dot{r}g(r)$

In order for (5.2-27) to be a viable amendment to (5.1-3), $f(r) \to \frac{-c}{r^2}$ as $r \to \pm\infty$ as usual, so that $f(r)$ satisfies the conditions of theorem 5.2-8 if $f(r)$ is also analytic. Notice in (5.1-3) that $\dot{r}$ is bounded as $r \to \pm\infty$ and so in any reasonable amendment $\dot{r}$ should be bounded and $g(r)$ should therefore satisfy the conditions of theorem 5.2-8 if it is analytic. The condition on $\int g(r)dr$ will be satisfied if $\lim_{r\to\infty} r^n g(r)$ is finite for $n > 1$. So, for instance, viable amendments to (5.1-3) are given by (5.2-27) with $g(r)$ satisfying the above criterion and $f(r)$ as in examples 5.2-3 to 5.2-7. Specific examples will be given in the next section when certain properties of the amendments are considered.

5.3) Properties of the Amended One-Dimensional Newtonian Equations

In this section, various properties of the dynamical amendments of (5.1-3) of the Liénard type,

$$\ddot{r} = f(r) + \alpha g(r)\dot{r} \qquad (5.3\text{-}1)$$

are investigated. In order that (5.3-1) is a viable, dynamical amendment to (5.1-3), $|\alpha| << 1$, $f(r) \to \frac{-c}{r^2}$ as $r \to \pm\infty$ and is at least C^1 and $g(r)$ is also at least C^1 with $\int_R^{\pm\infty} g(r)dr$ finite. These conditions will be assumed throughout the section.

5.3.1) Critical Points

As a system, equation (5.3-1) becomes,

$$\begin{cases} \dot{r} = p \\ \dot{p} = f(r) + \alpha g(r)p \end{cases} \qquad (5.3\text{-}2)$$

One of the most important aspects of (5.3-2) is the nature of its critical points, if any, especially if $(r,p) = (0,0)$ is a critical point, for then we are considering, in effect, the nature of the

transfonned singularity. Notice critical points are given by $p = 0$, $f(r) = 0$. It is shown in this sub-section that the critical point can be assigned any topological character by choosing the functions f and g suitably.

Let $(r,p) = (a,0)$ be a critical point of (5.3-2), so that $f(a) = 0$. The linearized system is given by

$$\begin{pmatrix} \dot{r} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ f'(a) & \alpha g(a) \end{pmatrix} \begin{pmatrix} r \\ p \end{pmatrix} \qquad (5.3\text{-}3)$$

expanding the right-hand side of (5.3-2) about (a,0) using Taylor's series.

Eigenvalues of $A = \begin{pmatrix} 0 & 1 \\ f'(a) & \alpha g(a) \end{pmatrix}$ are given by,

$$\lambda^2 - \alpha\lambda g(a) - f'(a) = 0 \implies$$

$$\lambda = \frac{\alpha g(a) \pm \sqrt{\alpha^2 g^2(a) + 4f'(a)}}{2} \qquad (5.3\text{-}4)$$

Clearly, (a,0) is simple if $f'(a) \neq 0$. The following lemma easily follows from lemma 3.5-1, theorems 3.5-2 to 3.5-4.

Lemma 5.3-1

Let (a,0) be a critical point of (5.3-2). (a,0) is simple if $f'(a) \neq 0$ and, in this case,

(i) If $f'(a) < 0$ and $\alpha^2 g^2(a) > -4f'(a)$, (a,0) is a node.

(ii) If $f'(a) > 0$, (a,0) is a saddle point.

(iii) If $g(a) \neq 0$ and $\alpha^2 g^2(a) < -4f'(a)$, (a,0) is a focus.

(iv) If $g(a) = 0$ and $f'(a) < 0$, (a,0) may be a multiple focus (and possibly enclosed in a limit cycle) or a centre.

(v) If $f'(a) < 0$ and $g(a) = \sqrt{\frac{-4f'(a)}{\alpha^2}}$, (a,0) is a node.

□

Notice if $\alpha = 0$ in (5.3-2), this is equivalent to having $g(r) \equiv 0$, and only parts (ii), (iv) of the lemma are appropriate, that is, it is not possible to generate all types of critical points.

Suppose now that (a,0) is a multiple critical point, so that $f'(a) = 0$. Then $\lambda = 0$ and $\lambda = \alpha g(a)$ are the roots of (5.3-4). Thus, (a,0) is a multiple critical point of the first kind only if $g(a) \neq 0$, (and $\alpha \neq 0$).

Theorem 5.3-2

Let (a,0) be a critical point of (5.3-2). (a,0) is multiple of the first kind if $f'(a) = 0$, $\alpha g(a) \neq 0$. Let n_1 be the lowest integer such that $f^{(n)}(a) \neq 0$. If n_1 is even, (a,0) is a saddle node. If n_1 is odd, (a,0) is a node if $f^{(n_1)}(a) < 0$, otherwise it is a saddle point.

□

Proof

It is assumed that, without loss of generality, $a = 0$. (If $a \neq 0$, the proof is the same except that functions are expanded about $r = a$ where appropriate).

In order to apply theorem 3.6-1, (5.3-2) has to be expressed in the canonical form (3.6-4). Let $\alpha g(0) = \beta$ and,

$$\begin{cases} R = p - \beta r \\ P = p \\ \tau = \beta t \end{cases} \qquad (5.3\text{-}5)$$

It is easily seen that the linear transformation (5.3-5) reduces (5.3-2) to the form,

$$\begin{cases} \dot{R} = 1/\beta\, h(R,P) \\ \dot{P} = P + 1/\beta\, h(R,P) \end{cases} \qquad (5.3\text{-}6)$$

where $h(R,P) = f\left(\frac{P-R}{\beta}\right) + \alpha\left(g\left(\frac{P-R}{\beta}\right) - g(0)\right)P \qquad (5.3\text{-}7)$

(so that $h(R,P)$ contains no linear terms and $h(0,0) = 0$ as required), and a dot now denotes differentiation with respect to τ.

Following the notation of theorem 3.6-1, let,

$$P = \phi(R) = a_2R^2 + a_3R^3 + a_4R^4 + \ldots \tag{5.3-8}$$

where $a_i \in \mathbb{R}$ $(i \geq 2)$, be the solution to,

$$f\left(\frac{P-R}{\beta}\right) + \alpha P g\left(\frac{P-R}{\beta}\right) = 0 \tag{5.3-9}$$

as, by hypothesis, ϕ is analytic where $\phi(0) = \phi'(0) = 0$.

As f and g are analytic, the power series (5.3-8) can be substituted into (5.3-9) and coefficients compared, so that values are obtained for the unknown constants a_i. Then,

$$\psi(R) = \frac{1}{\beta}\left\{f\left(\frac{\phi(R)-R}{\beta}\right) + \alpha\left\{g\left(\frac{\phi(R)-R}{\beta}\right) - g(0)\right\}\phi(R)\right\} \tag{5.3-10}$$

using (5.3-6). By theorem 3.6-1, $\psi(R)$ has the form $\Delta_m r^m$ where $m \geq 2$, $\Delta_m \neq 0$. As $\psi(R)$ is analytic, the coefficients of r^m in its power series can be found using the values of its derivatives at 0, that is, $m!\,\Delta_m = \psi^{(m)}(0)$ for each m. Using (5.3-10) it can be seen that,

$$\begin{cases} \psi''(0) = f''(0)/\beta^3 \\ \beta\psi'''(0) = -f'''(0)/\beta^3 - 3f''(0)\phi''(0)/\beta^2 - \frac{3\alpha}{\beta}g'(0)\phi''(0) \\ \beta\psi^{(iv)}(0) = f^{(iv)}(0)/\beta^4 + 6f'''(0)\phi''(0)/\beta^3 \\ \qquad + 3f''(0)\phi''(0)/\beta^2 - 4f''(0)\phi'''(0)/\beta^2 \\ \qquad -4\alpha\phi'''(0)g'(0)/\beta + 6\alpha\phi''(0)\left\{g''(0)/\beta^2 + g'(0)\phi''(0)/\beta\right\} \end{cases} \tag{5.3-11}$$

with higher order derivatives becoming increasingly complicated:- Using Liebnitz theorem,

$$\beta\psi^{(n)}(0) = D^n f\left(\frac{\phi(R)-R}{\beta}\right)\Big|_0 + \alpha_n C_1 \phi^{(n-1)}(R)\, Dg\left(\frac{\phi(R)-R}{\beta}\right)\Big|_0$$

$$+ \alpha_n C_2 \phi^{(n-2)}(R) D^2 g\left(\frac{\phi(R)-R}{\beta}\right)\Big|_0 + \ldots + \alpha_n C_{n-2}\phi''(R) \times$$

$$D^{n-2} g\left(\frac{\phi(R)-R}{\beta}\right)\Big|_0 ; \; n \geq 3. \tag{5.3-12}$$

The above expressions may be simplified by expressing the derivatives of ϕ in terms of the derivatives of f ang g. Substituting (5.3-8) into (5.3-9),

$$\frac{f''(0)}{2}\left\{\frac{-R}{\beta}+\frac{a_2R^2}{\beta}+\frac{a_3R^3}{\beta}+\dots\right\}^2+\frac{f'''(0)}{6}\left\{\frac{-R}{\beta}+\frac{a_2R^2}{\beta}+\frac{a_3R^3}{\beta}+\dots\right\}^3+\dots$$

$$+\alpha(a_2R^2+a_3R^3+a_4R^4+\dots)\left\{\frac{\beta}{\alpha}+g'(0)\left\{\frac{-R}{\beta}+\frac{a_2R^2}{\beta}+\dots.\right\}\right.$$

$$+\frac{g''(0)}{2}\left\{\frac{-R}{\beta}+\frac{a_2R^2}{\beta}\dots\right\}^2+\dots\Bigg\}\equiv 0. \qquad (5.3\text{-}13)$$

Comparing coefficients in the above identity and substituting the results into (5.3-11),

$$\begin{cases}
\psi''(0) = f''(0)/\beta^3 \\
\psi'''(0) = -f'''(0)/\beta^4 + 3f''(0)\left\{f''(0)+\alpha\beta g'(0)\right\}/\beta^6 \\
\psi^{(iv)}(0) = f^{(iv)}(0)/\beta^5 - f''(0)\left\{6f'''(0)+3f''(0)\beta+6\alpha\beta g''(0)\right\}/\beta^7 \\
\qquad + 6\alpha g'(0)(f''(0))^2/\beta^8 \\
\qquad - 4\left\{f''(0)+\alpha\beta g'(0)\right\}\times\left\{f'''(0)/\beta^3-6f''(0)/\beta^3\times\right. \\
\qquad\qquad \left.\left[\frac{\alpha g'(0)}{\beta}+\frac{f''(0)}{\beta^2}\right]\right\}/\beta^3
\end{cases} \qquad (5.3\text{-}14)$$

Higher derivatives are obtained by substituting for the derivatives of ϕ in (5.3-12). Every term except in the expression for $\psi^{(n)}(0)$ is multiplied by an mth derivative of f, evaluated at R = 0, where $m < n$, as can be seen for n = 2,3,4 from (5.3-14). Thus by theorem 3.6-1, if $f''(0) \neq 0$, the critical point is a saddle-node. If $f''(0) = 0$ and $f'''(0) \neq 0$ then the critical point is a node if $f'''(0) < 0$ or a saddle point if $f'''(0) > 0$. If $f'''(0) = 0$, $f^{(iv)}(0) \neq 0$, (0,0) is a saddle-node, and so on. Hence the result. ∎

It is interesting to note that the character of the critical point does not depend on g(r) beyond the fact that $g(a) \neq 0$.

Clearly, if (a,0) is a multiple critical point of the second kind, then $f'(a) = g(a) = 0$, (or else $\alpha = 0$).

Theorem 5.3-3

If (a,0) is a critical point of (5.3-2), it is multiple of the second kind if $f'(a) = 0 = \alpha g(a)$. Let k be the lowest integer such that $f^{(k)}(a) \neq 0$. Let ℓ be the lowest integer such that $g^{(\ell)}(a) \neq 0$.

(i) If k is even, let $k = 2m$. (a,0) is a degenerate point if $\ell \geq m$ and a saddle-node if $\ell < m$.

(ii) If k is odd, let $k = 2m+1$ and $\mu = \left(\frac{g^{(\ell)}(a)}{\ell!}\right)^2 + 4(m+1)\frac{f^{(k)}(a)}{k!}$.

(a) If $f^{(k)}(a) > 0$, (a,0) is a saddle point.

(b) If $f^{(k)}(a) < 0$, (a,0) is a focus or centre if $\ell > m$ or $\ell = m$ and $\mu < 0$; a node if ℓ is even and $\ell < m$ or is even and $\ell = m$ and $\mu \geq 0$; a critical point with elliptic region if ℓ is odd and $\ell < m$ or $\ell = m$, $\mu \geq 0$.

□

Proof

The system (5.3-2) is already in the canonical form (3.6-10), assuming without loss of generality that $a = 0$.

$$\dot{p} = f(r) + \alpha p g(r) \Rightarrow$$

$$\dot{p} = \left\{\frac{f''(0)}{2!} r^2 + \frac{f'''(0)}{3!} r^3 + \frac{f^{(iv)}(0)}{4!} r^4 + \ldots\right\}$$

$$+ \alpha p \left\{g'(0)r + \frac{g''(0)}{2!} r^2 + \frac{g'''(0)}{3!} r^3 + \ldots\right\} \quad (5.3\text{-}15)$$

Comparing (5.3-15) with the canonical form (3.6-11) and using theorem 3.6-2 the result follows.

Example 5.3-1

Consider the system,

$$\begin{cases} \dot{r} = p \\ \dot{p} = \dfrac{-2c}{\pi}\dfrac{\tan^{-1}(b\,r^3)}{r^2} + \dfrac{\alpha p}{1+r^2} \end{cases} \qquad (5.3\text{-}16)$$

where $a,c \in \mathbb{R}_{>0}$ and $\alpha \in \mathbb{R}$.

It is easy to see that $f(r) = \dfrac{-2c}{\pi}\dfrac{\tan^{-1}br^3}{r^2}$ and $g(r) = \dfrac{1}{1+r^2}$ satisfy the conditions of theorem 5.2-8 and so (5.3-16) is a dynamical amendment to (5.1-3).

The only critical point is situated at (0,0), and $f'(0) = \dfrac{-2cb}{\pi}$, $g(0) = 1$. Thus, by theorem 5.3-1, the origin is a simple critical point. Using the same theorem, there are three possibilities:-

(a) (0,0) is a focus if $\alpha^2 < \dfrac{8bc}{\pi}$

(b) (0,0) is a node if $\alpha^2 \geq \dfrac{8bc}{\pi}$

(c) (0,0) is a centre or multiple focus if $\alpha = 0$.

If (5.3-16) is to be a viable amendment of (5.1-3), $|\alpha| << 1$, $p >> 1$ and so case (b) is not at all likely! If, however, $f(r) = \dfrac{-2c}{\pi}\dfrac{\tan^{-1}br^n}{r^2}$, $n \geq 5$ and odd in (5.3-1), (0,0) is a node irrespective of the value of α.

Example 5.3-2

In order that (5.3-1) has a saddle point, say at (0,0), $f'(0) > 0$ by theorem 5.3-1 and so we can take $\alpha = 0$, without loss of generality. If (5.3-1) is to be a reasonable amendment to (5.1-3), $f(r) \to \dfrac{-2c}{r^2}$ as $r \to \infty$ and so $f(r) = 0$ for some r, $0 < r < \infty$. Thus, the equation has at least two critical points.

For example, taking $f(r) = \dfrac{-2c}{\pi}\dfrac{\tan^{-1}(\ln|1-r^3|)}{r^2}$,

$f(r) = 0$ at $r = 0$ and at $r = \sqrt[3]{2}$. Using l'Hopital's rule, $f(0) = 0$, $f'(0) = \dfrac{2c}{\pi} > 0$ and so (0,0) is a saddle point. Also, $f'(\sqrt[3]{2}) = \dfrac{-6c}{\pi} < 0$ and so $(\sqrt[3]{2},0)$ may be a centre or multiple focus.

Example 5.3-3

Consider now the system,

$$\begin{cases} \dot{r} = p \\ \dot{p} = \dfrac{-cr^{2n-2}}{r^n + a^n} + \dfrac{\alpha p r^m}{1 + r^4} \end{cases} \qquad (5.3\text{-}17)$$

where $n \geq 0$, $0 \leq m \leq 2$, $c \in \mathbb{R}_{>0}$, $\alpha \in \mathbb{R}$, $n, m \in \mathbb{Z}$

(5.3-17) is dynamical using theorem 5.2-8 and theorem 5.2-1. Again, the only critical point is situated at the origin.

Case One If $m = 0$, $g(0) = 1$, $f'(0) = 0$, and so by theorem 5.3-2, (0,0) is a multiple critical point of the first kind.

$$\text{Now } f(r) = \frac{-cr^{2n-2}}{r^{2n} + a^{2n}} = \frac{-cr^{2n-2}}{a^{2n}} \left\{ 1 - \frac{r^{2n}}{a^{2n}} + \frac{r^{4n}}{a^{4n}} - \ldots \right\} \quad |r| < |a|$$

and so the first non-zero derivative is $f^{(2n-2)}(0) = \dfrac{-c}{a^n}(2n-2)!$

Thus, (0,0) is a saddle-node for any value of n.

Case Two If $m = 1$ or 2, $g(0) = 0$, $f'(0) = 0$ and the origin is a multiple critical point of the second kind - theorem 5.3-3 is appropriate.

$$\text{Now } \dot{p} = \frac{-cr^{2n-2}}{a^{2n}} \left\{ 1 - \frac{r^{2n}}{a^{2n}} + \frac{r^{4n}}{a^{4n}} - \ldots \right\} + \alpha p r^m \left\{ 1 - r^4 + r^8 - \ldots \right\}$$

Using the notation of the theorem, k is always even and $\ell = 1$ or 2. Thus, if $m = 1$, (so that $\ell = 1$), (0,0) is degenerate if $n = 2$ and a saddle-node if $n > 2$. (0,0) is degenerate for $m = 2$ ($\ell = 2$) and $n = 2$, or 3 and a saddle-node if $n > 3$.

Notice, that if $\alpha = 0$ in (5.3-1), there is no choice of $f(r)$ which will make a given critical point multiple of the first kind. Also, the possibilities for a multiple critical point of the second kind are limited. Using theorem 5.3-3, it can be seen that it can only be a saddle point, focus, centre or a degenerate critical point. It is interesting to note that exactly the same restrictions on the types of critical points obtainable are imposed if in (5.3-1), $\alpha \neq 0$

but $f(r) \equiv g(r)$.

5.3.2) Stability of Orbits

Equation (5.3-1) is of the Liénard type and it is well known that a suitable Lyapunov function is given by,

$$L(r,p) = \frac{1}{2}p^2 - \int_o^r f(r)dr \qquad (5.3\text{-}18)$$

for the equivalent two-dimensional system (5.3-2), where it is assumed, without loss of generality, that (0,0) is the critical point of (5.3-2) in question.

Now $L(r,p) \gtrless 0$ if $\frac{1}{2}p^2 \gtrless \int_o^r f(r)dr$ and so is trivially true if $\int_o^r f(r)$ is negative.

$$\frac{dL}{dt} = L^* = \frac{\partial L}{\partial r}\frac{dr}{dt} + \frac{\partial L}{\partial p}\frac{dp}{dt} = \alpha p^2 g(r) \qquad (5.3\text{-}19)$$

Thus, the stability of (0,0) depends on the nature of $\alpha\, g(r)$ in the neighbourhood of the point, $r = 0$. The following lemma follows immediately from theorem 3.8-2:-

Lemma 5.3-4

If $\frac{1}{2}p^2 - \int_o^r f(r)dr > 0$ and $\alpha g(r) \lesseqgtr 0$ for all r is a neighbourhood of $r = 0$, then the critical point (0,0) of (5.3-2) is stable. □

Notice asymptotic stability cannot be concluded in this case since $\frac{dL}{dt}$ is never negative definite, no matter what the choice of $\alpha g(r)$ is.

Notice also that in the case $\alpha=0$, $\frac{dL}{dt} = 0$ from (5.3-19) and so L is constant - this can also be shown to be true in this case by substituting the known first integral (5.2-4) into (5.3-18). Thus, so long as $L > 0$ in a neighbourhood of (0,0), the origin is stable by lemma 5.3-4.

Notice that from (5.3-18) $L(r,p) \to \infty$ as $|(r,p)| \to \infty$, and so in certain cases, theorem 3.8-4 may be useful:-

Lemma 5.3-5

If $L(r,p)$ is positive definite in the whole of (r,p) space, $\alpha g(r) \lesseqgtr 0$ for all r and $(0,0)$ is the only invariant subset of $p^2 \alpha g(r) = 0$, then the critical point $(0,0)$ of (5.3-2) is globally asymptotically stable.

□

Notice asymptotic stability cannot be concluded in the case $\alpha = 0$ as then $\alpha g(r) = 0$ on the whole of $\mathbb{R}^n$.

Recall that the eigenvalues of the matrix associated with the linearized system of (5.3-2) are given by (5.3-4), so that the following lemma follows from theorem 3.8-1:-

Lemma 5.3-6

If the origin of (5.3-2) is a simple critical point, it is stable if $\alpha g(0) \lesseqgtr 0$ and $f'(0) < 0$. The origin is asymptotically stable if $\alpha g(0) < 0$.

□

Any of the above lemmas may be used on any particular system. It may happen that one gives more information than the others. In certain cases, Lemma 5.3-4 can be used to conclude asymptotic stability by constructing a modified Lyapunov function.[6]

In the case of a multiple critical point, Lemma 5.3-6 is not applicable - clearly a saddle-node, or degenerate critical point are not stable in the usual sense. However, there exist certain stable multiple critical points such as nodes and in order to prove stability in these cases, we must resort to the Lyapunov method.

If $(0,0)$ is the only critical point of (5.3-2), then the region of stability must be the whole of phase space. If there exists more than one critical point (for instance, if $f'(0) > 0$ in the amended Newtonian cases) the region of stability may be calculated using the Lyapunov function (5.3-18).

Example 5.3-4

Consider the system,

$$\begin{cases} \dot{r} = p \\ \dot{p} = \dfrac{-cr}{(r^2 + a^2)^{3/2}} + \dfrac{\alpha p r^2}{1 + r^4} \end{cases} \tag{5.3-20}$$

where $c \in \mathbb{R}_{>0}$ and $|a|, \ |\alpha| << 1$.

Notice $f(r) = \dfrac{-cr}{(r^2 + a^2)^{3/2}}$ and $g(r) = \dfrac{r^2}{1 + r^4}$ are

analytic and bounded and that $\lim_{r\to\infty} \int f(r)dr$ and $\lim_{r\to\infty} \int g(r)dr$ are

both finite, so that (5.3-20) is a viable amendment to (5.1-3) by theorem 5.2-8. (0,0) is the only finite critical point and $f'(0) = -\dfrac{c}{a^3} < 0$, $g(0) = 0$. Thus, by lemma 5.3-6 (0,0) is

stable, (being simple). Notice that from lemma 5.3-1, the origin must be a focus or centre.

Stability may also be concluded by the use of the Lyapunov function,

$$L(r,p) = \frac{1}{2}p^2 + c\int_0^r \frac{r\,dr}{(r^2 + a^2)^{3/2}} = \frac{1}{2}p^2 \quad \frac{-c}{(r^2 + a^2)^{1/2}} + \frac{c}{a} .$$

Clearly $L(0,0) = 0$ and $\dfrac{c}{a} \geq \dfrac{c}{(r^2 + a^2)^{1/2}}$ for all $r \Rightarrow L(r,p)$ is

positive definite. $\dfrac{dL}{dt} = \alpha \dfrac{p^2 r^2}{1 + r^4} \Rightarrow$ (0,0) is stable if $\alpha < 0$,

by Lemma 5.3-4, but need not be asymptotically stable.

Notice, if (5.3-1) is assumed to be dynamical, all solutions must be extendable to all values of t. The equation is stable in the sense of Lagrange, if, in addition, all the solutions are bounded for all values of t and so, in particular, do not go to the edge of the given manifold at infinity as t tends to infinity. If (5.3-1) is defined on $\mathbb{R}^2$, it can be shown that the solutions are bounded for

$t \geq 0$ if the Lyapunov function L is positive definite on $\mathbb{R}^2$ and tends to infinity as $|(r,p)| \to \infty$ and if $\frac{dL}{dt} \leq 0$ on $\mathbb{R}^2$.

5.3.3) Closed Paths of $\ddot{r} = f(r) + \alpha g(r)\dot{r}$

The question of existence of closed paths of equation (5.3-1) is now looked into. Studying Dulac's test (theorem 3.5-7) and the criteria for differentiating between a focus and centre using focal values (section 3.10), it is clear that it is easier to prove there do not exist closed orbits in phase space, rather than prove existence!

Applying Dulac's test to the two-dimensional system, (5.3-2), the following lemma is obtained,

Lemma 5.3-7 (Dulac's test)

Let G be a simply connected sub-region of the domain of the system (5.3-2). Let $B(r,p)$ be a C^∞ function defined in G such that the function,

$$\alpha B g(r) + \frac{p\partial B}{\partial r} + \frac{\partial B}{\partial p}\,(f(r) + \alpha p g(r)) \qquad (5.3\text{-}21)$$

does not change sign. (It may vanish at finitely many isolated points or on smooth curves, but has the same sign at all other points of G). Then there are no simple closed curves in G which are unions of the paths of system (5.3-2). □

In the case $\alpha \neq 0$, Bendixon's test may prove useful if $-B = 1$ in (5.3-21), Lemma 5.3-7 states that there are no simple closed curves in G is $g(r)$ does not change sign in G ($g(r)$ can be zero at a finite number of points). However, if $\alpha = 0$ in (5.3-2), there is no information as then (5.3-21) is identically zero, (and so is zero at an infinite number of points). In this case, some information can be obtained by letting $B = p$ in lemma 5.3-7:-

Corollary 5.3-8

Let G be a simply connected sub-region of the domain of definition of the system (5.3-2). If f(r) does not change sign, then there are no simple closed curves in G which are unions of the paths of the system (5.3-2) with $\alpha = 0$.

□

It can be shown, using the Poincaré index, that any closed curve contains a critical point in its interior[7], so if (5.3-2) has no critical points, there can be no closed orbits.

The most important cases of a system of the type (5.3-2), as far as dynamical amendments of (5.1-3) are concerned are where $f(0) = 0$ and (0,0) is the only critical point. Then in order for there to be a closed orbit, it is necessary for the linearized system about (0,0) to have a matrix with purely imaginary eigenvalues, that is, $g(0) = 0$ and $f'(0) < 0$ (by theorem 5.3-1). If it is possible to differentiate between a focus and a centre, it is possible to prove there exists a pericdic solution, or no such periodic solution exists. One way this may be able to be done is by using Lemma 3.5-6:- In the case of (5.3-2), it can easily be seen that orbits are symmetric with respect to the p axis if $f(r)$, $\alpha g(r)$ are odd functions of r.

Lemma 5.3-9

If (5.3-2) has a critical point at (0,0) and $g(0) = 0$, $f'(0) < 0$, and in addition, $f(r)$ and $\alpha g(r)$ are odd functions of r, then (0,0) must be enclosed in one or more periodic orbits.

□

Notice in the case $\alpha = 0$, all orbits are symmetric with respect to the $\ddot{r}$ axis, no matter what the character of $f(r)$, and so by Lemma 3.5-6, we have,

Lemma 5.3-10

If (5,3-2) with $\alpha = 0$ possesses a critical point at (0,0) and $f'(0) < 0$, then (0,0) must be enclosed in a system of one or more

periodic orbits.

It can be seen that if $\alpha = 0$ in (5.3-2) and $f(r)$ is an odd function, orbits are symmetric with respect to both r and p axes.

Example 5.3-5

Consider the dynamical amendment to (5.1-3),

$$\ddot{r} = f(r) = \frac{-c}{r^2 + a^2} \quad c \in \mathbb{R},\ |a| << 1. \qquad (5.3\text{-}22)$$

(see Example 5.2-1). Clearly $f(r)$ never changes sign and so (5.3-22) has no periodic solutions by Cor. 5.3-8.

(The equation, as a system is (r,p) space, has no critical points and so the result follows without using the Corollary in this case).

Example 5.3-6

Consider the dynamical amendment of (5.1-3), (by theorem 5.2-8),

$$\begin{cases} \dot{r} = p \\ \dot{p} = \dfrac{r}{(r^2+a^2)^{3/2}} (\alpha p - c) \end{cases} \qquad c,a,\alpha \in \mathbb{R},\ |a| << 1. \quad (5.3\text{-}23)$$

Clearly $f(r) = \dfrac{-cr}{(r^2 + a^2)^{3/2}} = -c\, g(r)$ are both odd functions and (5.3-23) possesses a critical point at the origin. Moreover, $g(0) = 0$ and $f'(0) = \dfrac{-c}{a^3} < 0$. Thus, by Lemma 5.3-9, the origin must be enclosed in a periodic orbit. The result also follows by Lemma 5.3-10 if $\alpha = 0$.

Example 5.3-7

Consider the equations,

$$\begin{cases} \dot{r} = p \\ \dot{p} = \frac{-2c}{\pi} \frac{\tan^{-1} br^3}{r^2} + \frac{\alpha p}{1+r^2} \end{cases} \qquad (5.3\text{-}24)$$

These represent a dynamical amendment of (5.1-3) by theorem 5.2-8 The only finite critical point is situated at (0,0).

If $\alpha \neq 0$, $g(r) = \frac{1}{1+r^2} > 0$ for all r and so there are no closed paths. In contrast, if $\alpha = 0$, $f'(0) < 0 \Rightarrow$the origin is a centre or multiple focus, By lemma 5.3-10, (0,0) must be enclosed in one or more closed orbits.

If $\alpha \neq 0$ and f(r) and g(r) are not odd functions, in general it is difficult to prove the existence of closed orbits. However, if a given critical point has a purely imaginary eigenvalue it can be proved to be a focus using a finite number of equations. This then excludes the possibility of the origin being a centre and hence an infinite number of closed orbits, although it does not exclude the possibility of the existence of limit cycles.

Consider the system,

$$\begin{cases} \dot{r} = p \\ \dot{p} = f(r) + \alpha p g(r) \end{cases} \qquad (5.3\text{-}2)$$

with $f(0) = 0$, $f'(0) < 0$, $g(0) = 0$, so that (0,0) is a centre or multiple focus. ($\alpha \neq 0$, otherwise (0,0) is known to be a centre by lemma 5.3-10.)

Consider the linear transformation,

$$\begin{cases} R = \beta r \\ P = p \end{cases} \qquad (5.3\text{-}25)$$

where $\beta^2 = -f'(0) > 0$. Then (5.3-2) is reduced to canonical form,

$$\begin{cases} \dot{R} = \beta P \\ \dot{P} = -\beta R + \left[\alpha P g\left(\frac{R}{\beta}\right) + f\left(\frac{R}{\beta}\right) + \beta R \right] \end{cases} \qquad (5.3\text{-}26)$$

Assuming f,g to be analytic in a neighbourhood of the origin, let

$$\begin{cases} f(r) = f_1 r + \beta^2 f_2 r^2 + \beta^3 f_3 r^3 + \beta^4 f_4 r^4 + \dots \\ g(r) = \beta g_1 r + \beta^2 g_2 r^2 + \beta^3 g_3 r^3 + \beta^4 g_4 r^4 + \dots \end{cases} \tag{5.3-27}$$

where f_i, g_i are constants and $f_1 = -\beta^2$. Transforming to polar co-ordinates $P = \rho\cos\theta$, $R = \rho\sin\theta$, (5.3-26) becomes, using (5.3-27),

$$\frac{d\rho}{d\theta} = \frac{\left\{ f_2\rho^2\sin^2\theta + f_3\rho^3\sin^3\theta + \dots + \alpha\rho\cos\theta\,(g_1\rho\sin\theta + g_2\rho^2\sin^2\theta + \dots) \right\}\cos\theta}{\beta + f_2\rho\sin^3\theta + f_3\rho^2\sin^4\theta + \dots + \alpha\sin\theta\cos\theta\,(g_1\rho\sin\theta + g_2\rho^2\sin^2\theta + \dots)} \tag{5.3-28}$$

Let $\begin{cases} A_n = f_n\sin^n\theta\cos\theta + \alpha\cos^2\theta\sin^{n-1}\theta g_{n-1} \\ B_n = \dfrac{\tan\theta}{\beta} A_{n+1} \end{cases}$ (5.3-29)

Substituting the expressions (5.3-29) into (5.3-28), we get,

$$\frac{d\rho}{d\theta} = e_2\rho^2 + e_3\rho^3 + e_4\rho^4 + \dots + e_n\rho^n + \dots \tag{5.3-30}$$

where

$$\begin{cases} \beta e_n = A_n - A_{n-1}X_1 + A_{n-2}X_2 + \dots + (-1)^n A_2 X_{n-2} \\ X_o = 1 \\ X_n = \underset{\substack{\text{ALL} \\ \text{COMBINATIONS}}}{\Sigma} \left\{ \underset{\Sigma i = n}{\prod_i} (-1)^{i+2n-1} B_i \right\} \end{cases} \tag{5.3.31}$$

As in definition 3.10-4, let $\rho = F(\theta; 0, \rho_o)$ be the solution of (5.3-30) satisfying the initial condition $F(0;0,\rho_o) = \rho_o$. Expanding $F(\theta; 0, \rho_o)$ in terms of powers of the initial value,

$$\rho = F(\theta; 0, \rho_o) = u_1(\theta)\rho_o + u_2(\theta)\rho_o^2 + u_3(\theta)\rho_o^3 + \dots \tag{5.3-32}$$

where, by hypothesis, $u_1(0) = 1$, $u_2(0) = \dots = u_n(0) = 0$.

Substituting (5.3-32) into (5.3-30) yields a differential

recurrence relation for $u_i(\theta)$ by comparing coefficients of ρ_o^n. By definition, focal values are calculated from,

$$\begin{cases} d'(0) = u_1(2\pi) - 1 \\ d^{(k)}(0) = k!\, u_k(2\pi) \end{cases} \qquad (5.3\text{-}33)$$

Thus, by finding each $u_i(\theta)$, it is clearly possible to obtain conditions on $f(0)$, $g(0)$ so that $(0,0)$ is a multiple focus of multiplicity n. In this work, only the third focal value is considered and this leads to a simple condition for $(0,0)$ to be a multiple focus of multiplicity one and hence exclude the possibility of $(0,0)$ being a centre.

From (5.3-32), (5.3-30),

$$\begin{cases} u_1'(\theta) = 0 \\ u_2'(\theta) = e_2 u_1^2 \\ u_3'(\theta) = 2e_2 u_1 u_2 + e_3 u_1^3 \end{cases} \qquad (5.3\text{-}34)$$

$\therefore\ u_1(\theta) = 1 \Rightarrow u_2(\theta) = \frac{1}{\beta} \int_o^\theta A_2\, d\theta.$

$\therefore\ u_2(\theta) = \frac{1}{\beta}\int_o^\theta (f_2 \sin^2\theta \cos\theta + \alpha\cos^2\theta \sin\theta g_1)\, d\theta \Rightarrow$

$$u_2(\theta) = \frac{1}{3\beta}(f_2 \sin^3\theta - \alpha g_1 \cos^3\theta) \qquad (5.3\text{-}35)$$

Thus, from (5.3-34), (5.3-35), (5.3-33),

$$d'''(0) = 6\int_o^{2\pi} \left(\frac{2e_2}{3\beta}(f_2 \sin^3\theta - \alpha g_1 \cos^3\theta) + e_3\right) d\theta \qquad (5.3\text{-}36)$$

Substituting the expressions (5.3-31), (5.3-29) into (5.3-36) and noting some terms are zero when integrated,

$$d'''(0) = 6\int_o^{2\pi} \left\{ \frac{2\alpha g_1 f_2}{3\beta^2}(\sin^4\theta \cos^2\theta - \cos^4\theta \sin^2\theta) + \frac{1}{\beta}(\alpha g_2 \cos^2\theta \sin^2\theta - \frac{2\alpha g_1 f_2}{\beta} \sin^4\theta \cos^2\theta) \right\} d\theta \Rightarrow$$

$$d'''(0) = \frac{3\pi}{2}\left(\frac{\alpha}{\beta} g_2 - \frac{\alpha g_1 f_2}{\beta^2}\right) \qquad (5.3\text{-}37)$$

using Wallis's reduction formula[8].

Substituting for g_1, g_2, β, and f_2 in (5.3-37), the following theorem is obtained:-

Theorem 5.3-11

If (0,0) is a critical point of (5.3-2), it is a multiple focus of multiplicity one if $\alpha \neq 0$, $f'(0) < 0$, $g(0) = 0$ and, $f'(0)g''(0) + g'(0)f''(0) \neq 0$.

□

The conditions of the theorem are no satisfied if $\alpha = 0$ or if $g(r)$, $f(r)$ are both odd, agreeing with lemmas 5.3-9, 5.3-10.

5.3.4) Structural Stability and Bifurcations

Applying theorem 3.9-1 to (5.3-2) defined on $\mathbb{R}^2$, we have,

Theorem 5.3-12

The system (5.3-2) is structurally stable in a closed region G of $\mathbb{R}^2$ if and only if,

(i) there exists a finite number of critical points at $(a_i, 0)$ say and $f'(a_i) \neq 0$ for each a_i and $g(a_i) = 0$, $f'(a_i) < 0$ do not hold simultaneously for any $r = a_i$;

(ii) if there exist closed paths, then these are simple limit cycles, that is, if $r = \phi(s)$, $p = \psi(s)$, then $J = \alpha \int_0^\tau g(\phi)ds \neq 0$ where τ is the period of ϕ and ψ.

(iii) there are no orbits going from saddle point to saddle point.

□

Notice that the conditions of the theorem are not satisfied throughout the plane if (5.3-2) is dynamical and $\alpha = 0$ and this agrees with the fact that (5.3-2) with $\alpha = 0$ has a regular first integral.

Suppose that (5.3-2) has no closed orbits. If it is to be a viable dynamical amendment to Newton's equation, it has been noted earlier that the origin is a saddle point only if there are at least two finite critical points (see example 5.3-2). If the origin is the only finite critical point, condition (iii) of theorem 5.3-12 is trivially satisfied in a closed region $G \subset \mathbb{R}^2$. Thus, we have the following corollary.

Corollary 5.3-13

If in (5.3-2), $f(r) \to \frac{-c}{r^2}$ as $r \to \infty$ ($c \in \mathbb{R}_{>0}$) and the system has only one finite critical point situated at the origin and $g(0) = 0$, $f'(0) < 0$ do not hold simultaneously, then the system is structurally stable in a closed region G of $\mathbb{R}^2$.

□

Example 5.3-8

Consider the dynamical amendment of (5.1-3) given in example 5.3-1, namely,

$$\begin{cases} \dot{r} = p \\ \dot{p} = \frac{-2c}{\pi} \frac{\tan^{-1}(br^3)}{r^2} + \frac{\alpha p}{1+r^2} \end{cases} \qquad (5.3\text{-}16)$$

where $b \gg 1$, $|\alpha| \ll 1$.

The origin is the only finite critical point of the system and $f'(0) < 0$, $g(0) = 1$. By Corollary 5.3-13, the system is structurally stable if $\alpha \neq 0$.

Having looked at the sub-class of equations of the type (5.3-2) which are structurally stable, we now look at the bifurcations of the structurally unstable types. If (5.3-2) has a multiple critical element, there are systems close to (5.3-2) which have a number of simple critical elements equal to the multiplicity.

Suppose that a critical point of (5.3-2) is a centre or multiple focus. A way of differentiating between the two cases has been discussed

in the previous sub-section. If, without loss of generality, (0,0) is a multiple focus of (5.3-2) of multiplicity k, then by theorem 3.10-1 , there is a system close to (5.3-2) which possesses exactly k closed paths.

Example 5.3-9

Consider the following dynamical amendment to (5.1-3),

$$\begin{cases} \dot{r} = p \\ \dot{p} = \dfrac{-2c}{\pi}\dfrac{\tan^{-1}br^3}{r^2} + \alpha p\dfrac{r^2}{1+r^4} \end{cases} \tag{5.3-38}$$

where $b>>1$, $|\alpha|<<1$. The origin is the only finite critical point and $f'(0) = \frac{-2bc}{\pi} < 0$, $f''(0) = 0$, $g(0) = 0$, $g'(0) = 0$, $g''(0) = 2$.

The system is structurally unstable by Corollary 5.3-13 and (0,0) is a multiple focus of multiplicity one if $\alpha \neq 0$, by theorem 5.3-11. Thus, by theorem 3.10-1, , there is a system close to (5.3-38) possessing one simple limit cycle. For example, consider the system,

$$\begin{cases} \dot{r} = p \\ \dot{p} = \dfrac{-2c}{\pi}\dfrac{\tan^{-1}br^3}{r^2} + \alpha p\left(\dfrac{r^2}{1+r^4} + \mu\right) \end{cases} \tag{5.3-39}$$

where μ is a small parameter.

It can be shown that, in this case, for sufficiently small values of μ,

(a) if $\alpha > 0$, for $\mu < 0$, there exists a stable focus and one unstable cycle; for $\mu > 0$, there exists an unstable focus and no cycle.

(b) if $\alpha < 0$, for $\mu > 0$, there exists a stable focus and no cycle; for $\mu < 0$, there exists an unstable focus and a stable cycle.[9]

Consider now the case where (0,0) is a centre of (5.3-2) (for example, if $\alpha = 0$ and $f'(0) < 0$). In this case, there is a neighbourhood of the origin which can be completely filled with concentric paths of (5.3-2). Thus, by theorem 3.10-2 there is an analytic function $\rho(r,p)$ defined in G such that,

$$\begin{cases} \dot{r} = \rho p \\ \dot{p} = \rho f(r) + \alpha\rho\, pg(r) \end{cases} \qquad (5.3\text{-}40)$$

is Hamiltonian. If $\alpha = 0$, $\rho = 1$ trivially satisfies the partial differential equation of theorem 3.10-2 , and the system (5.3-2) is already Hamiltonian in this case, with,

$$H(r,p) = \int f(r)dr - \frac{p^2}{2} .$$

In this case, theorem 3.10-3 , is appropriate and so there exist systems close to (5.3-2) possessing an infinite number of structurally stable limit cycles. One particular perturbation of (5.3-2) with $\alpha = 0$ is (5.3-2) with $\alpha \neq 0$. Using the notation of theorem 3.10-3, it is seen that in this case,

$$f_1(r,p) = \ldots = f_n(r,p) = \ldots = 0,$$

$$g_1(r,p) = pg(r); \quad g_2(r,p) = \ldots = g_n(r,p) = \ldots = 0,$$

so that,

$I = \alpha \iint_{G_o} g(r)drdp$ and $L = \alpha\int_o^\tau g(\phi(s))ds$, where G_o is the region enclosed inside a chosen closed path C and $r = \phi(t)$ is the motion corresponding to the path C.

Notice I can be integrated with respect to p first as (5.3-2) has a first integral (5.2-4) if $\alpha = 0$. Notice also that any closed curve must be symmetric about the r axis by lemma 3.5-6

$$\therefore \quad I = 2\alpha \int_{r_o}^{r_1} (g(r)\sqrt{2\int f(r)dr + A_o})dr$$

where r_o, r_1 are the least and greatest roots of $2\int f(r)dr + A_o = 0$ and L_c is given by $p^2 = 2\int f(r)dr + A_o$.

Thus, we have the following Corollary to theorem 3.10-3

Corollary 5.3-14

Let C be a closed path of (5.3-2) with $\alpha = 0$, $r = \phi(t)$ the

motion corresponding to the path, τ the period of the function ϕ, and G_o the region enclosed by C.

If $I = \int_{r_o}^{r_1} (g(r) \sqrt{2\int f(r)dr + A_o}\, dr = 0$, where r_o, r_1 are the least and greatest roots of $2\int f(r)dr + A_o = 0$ respectively and C is given by $p^2 = 2\int f(r)dr + A_o$, and,

$$L = \int_o^\tau g(\phi(s))ds \neq 0,$$

there exist $\varepsilon > 0$, $\delta > 0$ such that for every α, $|\alpha| < \delta$, (5.3-2) with $\alpha \neq 0$ has a structurally stable limit cycle in U_ε (C) and this contracts to C for $\alpha \to 0$.

□

5.4). The Two-Dimensional Newtonian Equation

In this section, the special two-dimensional case of (5.1-1) is considered. Using cartesian coordinates (x,y), (5.1-1) reduces to,

$$\begin{cases} \dfrac{d^2x}{dt^2} = \dfrac{-cx}{r^3} \\ \dfrac{d^2y}{dt^2} = \dfrac{-cy}{r^3} \end{cases} \qquad (5.4\text{-}1)$$

and so has a four-dimensional phase space. It is again convenient to treat c as a single positive constant, without loss of generality.

It is well known that (5.4-1) possesses periodic solutions:- transforming to polar coordinates, $x = r\cos\theta$, $y = r\sin\theta$, the following equations are obtained,

$$\begin{cases} \ddot{r} - r\dot{\theta}^2 = \dfrac{-c}{r^2} & (5.4\text{-}2)\,(a) \\ 2\dot{r}\dot{\theta} + r\ddot{\theta} = 0 & (5.4\text{-}2)\,(b) \end{cases}$$

The equations have a first integral, obtained by integrating

(5.4-2)(b) and substituting into (5.4-2)(a):-

$$\begin{cases} \dot{r} = \sqrt{2c/r - k^2/r^2 + A} & (5.4\text{-}3)(a) \\ \dot{\theta} = \dfrac{k}{r^2} & (5.4\text{-}3)(b) \end{cases}$$

where k,A are constants of integration.

The solution to (5.4-3) can be expressed in the form,

$$\frac{1}{r} = A^* \cos(\theta+\alpha) + \frac{c}{k^2} \quad ; \quad A^* = \frac{1}{k^2}\sqrt{c^2 + k^2 A} \qquad (5.4\text{-}4)$$

$(k \neq 0)$, where α is a constant of integration. (5.4-4) represents an equation of an ellipse if $\frac{A^*k^2}{c} < 1 \Rightarrow \sqrt{c^2 + k^2 A} < c$ and so A must be negative.

The right-hand sides of (5.4-1) become singular at the origin and the first integrals in polar co-ordinates (5.4-3), also become singular at $r = 0$. It is clear that, by integrating (5.4-3)(a), that this singularity is reached in finite time, if $k = 0$. Thus, (5.4-3) does not have global existence of solutions and so neither can (5.4-1), as both systems of equations must have the same general solution.

5.5). Feasible Amendments of the Two-Dimensional Newtonian Equations

In this section, the two-dimensional Newtonian equations are amended in such a way that the resulting equations have global existence of solutions and reduce to the amended one-dimensional Newtonian equations (5.2-1) in the one-dimensional limit. Amendments of the first order equations (5.4-3) are considered, so that the amendments will, in general, contain two arbitrary constants, as in the Newtonian case. Notice that, assuming amendments of this type, the one-dimensional case will not reduce to equations of the form (5.2-27) as these, in general, do not have first integrals.

By considering first order equations, the phase space will be the space manifold, in this case the plane $\mathbb{R}^2$, as opposed to the position-velocity space, as considered in the one-dimensional case. Thus, orbits in phase space will be orbits of a test particle as seen by an external observer.

5.5.1). Equations of the Form $\dot{r} = F(r)$, $\dot{\theta} = G(r)$

In this sub-section, equations of the form,

$$\begin{cases} \dot{r} = F(r) & (5.5\text{-}1)(a) \\ \dot{\theta} = G(r) & (5.5\text{-}1)(b) \end{cases}$$

are considered, which include (5.4-3) as a special case. At this stage, no special properties of F(r) and G(r) are assumed.

Theorem 5.5-1

The equations (5.5-1) have global existence of solutions if F(r) is imaginary in an arbitrarily small neighbourhood of all finite singular points of F(r) and G(r), or if F(r) and G(r) are continuous, and F'(r) is finite at each finite critical point and at infinity, if F(r) becomes singular there.

□

Proof

From (5.5-1)(a) we obtain,

$$\int \frac{dr}{F(r)} = t + t_o \qquad (5.5\text{-}2)$$

Suppose that F(r) becomes singular at a finite value of r, say $r = r_1$. The integral in (5.5-2) is well-behaved in a neighbourhood of $r = r_1$ and so $r = r_1$ must be reached in finite time if F(r) is real in a neighbourhood of $r = r_1$. If G(r) becomes singular at $r = r_1$, the same result applies, again using (5.5-2).

Suppose now that F(r) and G(r) are continuous and that F(r) becomes singular at infinity. In this case, the integral in (5.5-2) is

an improper integral of the first kind. Then,

$$L_1 = \lim_{r\to\pm\infty} \frac{r^n}{F(r)} = \lim_{r\to\pm\infty} \frac{nr^{n-1}}{F'(r)} \qquad (5.5\text{-}3)$$

by hypothesis. L is finite if $\lim_{r\to\pm\infty} F'(r)$ is finite and n is taken to be 1. In this case, the integral diverges by theorem 5.2-1. If G(r) becomes singular at infinity but F(r) does not, $\int^{\pm\infty} \frac{dr}{F(r)}$ must diverge, again by theorem 5.2-1 and there is no problem.

Suppose now that (5.5-1) possesses a critical point at a finite value of r, say r = a, so that F(a) = G(a) = 0. The integral in (5.5-2) is then an improper integral of the second kind, in a neighbourhood of r = a.

$$\text{Let } L_2 = \lim_{r\to a} \frac{(r-a)^n}{F(r)} = \lim_{r\to a} \frac{n(r-a)^{n-1}}{F'(r)} \qquad (5.5\text{-}4)$$

by hypothesis on F(r). If F'(a) is not infinite, $L_2 \neq 0$ when n = 1, so by theorem 5.2-2, the integral must diverge and the critical point is not reached in finite time.

If (5.5-1) possesses a critical point at infinity, this point can never be reached in finite time as the integral in (5.5-2) diverges by theorem 5.2-1 in this case. Hence the result. ∎

Notice that the derivatives of F(r) and G(r) do not have to be continuous at all values of r for the solutions to be extendable to any parameter value - they only have to be continuous at any critical point. Thus, (5.5-2) may have global existence of solutions without being dynamical in the usual sense.

5.5.2) Feasible Amendments of the Two-Dimensional Newtonian Equations of the Form (5.5-1)

Clearly, any feasible amendment of (5.4-3) must agree as far as possible with (5.4-3) - it must reduce to this case as r tends to infinity. The equation (5.4-3)(a) involves the Newtonian potential

and also depends on $\dot{\theta}$. Thus, in any amendment to (5.4-3) of the form (5.5-1), it is reasonable to have F(r) depending on the amended Newtonian potential and on G(r).

In the one-dimensional case, $\dot{\theta} = 0$, so that $G(r) \equiv 0$. In order that (5.5-1) agrees with (5.2-1) in this case, we have,

$$F(r) = \sqrt{2\int f(r)dr + A} \quad \text{when} \quad \dot{\theta} = 0 \qquad (5.5\text{-}5)$$

(A a constant of integration), where the second order one-dimensional Newtonian equation is given by,

$$\ddot{r} = f(r) \qquad (5.2\text{-}1)$$

that is, where $-\int f(r)dr$ is the amended Newtonian potential. Thus,

$$\begin{cases} \dot{r} = \sqrt{2\int f(r)dr + A + 2H(r)} \\ \dot{\theta} = G(r) \end{cases} \qquad (5.5\text{-}6)$$

where $H(r) \equiv 0$ when $G(r) \equiv 0$, (so that G(r) and H(r) contain an arbitrary constant).

Now from (5.5-6),

$$\begin{cases} \ddot{r} - r\dot{\theta}^2 = \dfrac{(f(r) + H'(r))\dot{r}}{\dot{r}} - rG^2(r) \\ 2\dot{r}\dot{\theta} + r\ddot{\theta} = 2G\sqrt{2\int f(r)dr + A + 2H(r)} + rG'(r)\dot{r} \end{cases}$$

$$\therefore \begin{cases} \ddot{r} - r\dot{\theta}^2 = f(r) + H'(r) - rG^2(r) & (5.5\text{-}7)(a) \\ 2\dot{r}\dot{\theta} + r\ddot{\theta} = (2G(r) + rG'(r))\sqrt{2\int f(r)dr + 2H(r) + A} & (5.5\text{-}7)(b) \end{cases}$$

Transforming back to cartesian co-ordinates,

$$\begin{cases} \ddot{x} = \cos\theta(\ddot{r} - r\dot{\theta}^2) - \sin\theta(2\dot{r}\dot{\theta} + r\ddot{\theta}) \\ \quad = \dfrac{x}{r}\left\{f(r) + H'(r) - rG^2(r)\right\} - \dfrac{y}{r}\left\{(2G+rG')\sqrt{2\int f(r)dr + 2H(r) + A}\right\} & (5.5\text{-}8)(a) \\ \ddot{y} = \sin\theta(\ddot{r} - r\dot{\theta}^2) + \cos\theta(2\dot{r}\dot{\theta} + r\ddot{\theta}) \\ \quad = \dfrac{y}{r}\left\{f(r) + H'(r) - rG^2(r)\right\} + \dfrac{x}{r}\left\{(2G+rG')\sqrt{2\int f(r)dr + 2H(r) + A}\right\} & (5.5\text{-}8)(b) \end{cases}$$

Now in the Newtonian case, by hypothesis, $f(r) = \frac{-c}{r^2}$, and so to agree as close as possible to (5.4-1), we need,

$$H'(r) - rG^2(r) = 0 \implies$$

$$H(r) = \int rG^2(r)dr. \qquad (5.5\text{-}9)$$

$$\text{If, } 2G + rG' = 0 \Rightarrow G(r) = \frac{k}{r^2} \qquad (5.5\text{-}10),$$

the second order equations (5.5-8) are not coupled as in the Newtonian case and the force is central. Otherwise the equations are coupled.

Suppose it is possible to amend the equations (5.4-1), so that the gravitational force is still central, and (5.5-1) has global existence of solutions. From (5.5-10), (5.5-9), (5.5-6) takes the form,

$$\begin{cases} \dot{r} = \sqrt{2\int f(r)dr + A - k^2/r^2} \\ \dot{\theta} = \frac{k}{r^2} \end{cases} \qquad (5.5\text{-}11)$$

By previous analysis, this is a feasible amendment in the one-dimensional case (k=0). However, when $k \neq 0$, the right-hand sides of (5.5-11) become singular at $r = 0$. By theorem 5.5-1, (5.5-11) will only have global existence of solutions if $2\int f(r)dr + A - \frac{k^2}{r^2}$ is negative in an arbitrarily small neighbourhood of $r = 0$. This will be true for any choice of the arbitrary constants A, k if $\int f(r)dr \to -\infty$ as $r \to 0$. (compare with lemma 5.2-3). As shown in section 5.2), this property of f(r) does not give reasonable behaviour and so (5.5-11) may only have global existence of solutions for certain values of A. This means that the second order equations (5.5-8) may not be dynamical. From section 5.4), periodic orbits are given by $A < 0$ and so if A is chosen in the amendment (5.5-11) so that, $\left[2\int f(r)dr + A - \frac{k^2}{r^2}\right]_{o^+} < 0$, these values of A represent periodic or almost periodic orbits which never reach the origin.

Example 5.5-1

Consider the amended one-dimensional equation given in example 5.26, namely

$$\ddot{r} = \frac{-cr}{(r^2+a^2)^{3/2}} \qquad (5.2\text{-}15)$$

where $|a| << 1$, so that $f(r) = \frac{-cr}{(r^2+a^2)^{3/2}}$. The two-dimensional equations (5.5-11) become, in this case,

$$\begin{cases} \dot{r} = \sqrt{\frac{2c}{(r^2+a^2)^{\frac{1}{2}}} - \frac{k^2}{r^2} + A} \\ \dot{\theta} = \frac{k}{r^2} \end{cases} \qquad (5.5\text{-}12)$$

In this case, $\dot{r}$ will be imaginary when,

$2cr^2 - k^2(r^2 + a^2)^{\frac{1}{2}} + Ar^2(r^2+a^2)^{\frac{1}{2}} \doteq 2cr^2 - k^2r + Ar^3 < 0$,
if $|a| << 1$. Thus, $\dot{r}$ will always be imaginary in a neighbourhood of $r = 0$ if $k \neq 0$. If $k = 0$, the model is non-singular anyway.

Clearly, if $k \neq 0$, $\dot{\theta} \neq 0$ for finite values of r, and $\dot{r} \neq 0$ when $r = 0$. Hence, (5.5-12) does not possess any critical points, and so has global existence of solutions.

If (5.5-6) is required to be nonsingular, (5.5-8) are now coupled and so in this case (5.5-6) are not the first integral of (5.5-8) as the second order equations contain the arbitrary constant of G(r) and H(r). However, if G(r) is chosen to be non-singular at infinity and $G(r) \neq 0$ for any values of r in the two-dimensional case, then (5.5-6) are feasible amendments of (5.4-3) if F'(0) is finite and F(r) is non-singular at infinity by theorem 5.5-1.

Example 5.5-2

Consider the amended two-dimensional equation, (5.5-6), with

$G(r) = \dfrac{k}{r^2 + a^2}$, $|a| << 1$. From (5.5-9),

$$H(r) = \int \frac{r\,k^2}{(r^2+a^2)^2}\,dr = \frac{-k^2}{2(r^2+a^2)} \qquad (5.5\text{-}14)$$

and so (5.5-6) becomes, if $f(r) = \dfrac{-cr}{(r^2+a^2)^{3/2}}$,

$$\begin{cases} \dot{r} = \sqrt{\dfrac{2c}{(r^2+a^2)^{\frac{1}{2}}} - \dfrac{k^2}{(r^2+a^2)} + A} \\ \dot{\theta} = \dfrac{k}{r^2+a^2} \end{cases} \qquad (5.5\text{.-}15)$$

If $a = 0$, (5.5-15) reduces to the Newtonian case and if $k = 0$, (5.5-15) reduces to (5.2-15) as (5.5-12). In this case, the amendment is non-singular, even at infinity. Notice $\dot{\theta} \neq 0$ if $k \neq 0$ except at infinity.

Now $\dot{r} = 0$ at $r = 0$ if $A = \dfrac{k^2}{a^2} - \dfrac{2c}{a}$.

In this case, $F(r) = \sqrt{\dfrac{2c}{(r^2+a^2)^{\frac{1}{2}}} - \dfrac{k^2}{(r^2+a^2)} + \dfrac{k^2}{a^2} - \dfrac{2c}{a}}$ $\Rightarrow$

$$F'(0) = \lim_{r\to 0} \left\{ \frac{-cr}{(r^2+a^2)^{3/2}} + \frac{k^2 r}{(r^2+a^2)^2} \right\} \left\{ \frac{2c}{(r^2+a^2)^{\frac{1}{2}}} - \frac{k^2}{(r^2+a^2)} + \frac{k^2}{a^2} - \frac{2c}{a} \right\}^{-\frac{1}{2}}$$

Using l'Hôpital's rule,

$$F'(0) = \left[\lim_{r\to 0} \left\{ \frac{2\left\{ \dfrac{-cr}{(r^2+a^2)^{3/2}} + \dfrac{k^2 r}{(r^2+a^2)^2} \right\} \left\{ \dfrac{-cr}{(r^2+a^2)^{3/2}} + \dfrac{k^2 r}{(r^2+a^2)^2} \right\}'}{\left\{ \dfrac{-2cr}{(r^2+a^2)^{3/2}} + \dfrac{2k^2 r}{(r^2+a^2)^2} \right\}} \right\} \right]^{\frac{1}{2}}$$

$$\Rightarrow F'(0) = \left\{ \lim_{r\to 0} \frac{d}{dr} \left\{ \frac{-cr}{(r^2+a^2)^{3/2}} + \frac{k^2 r}{(r^2+a^2)^2} \right\} \right\}^{\frac{1}{2}}, \text{ and so is finite.}$$

Thus, (5.5-15) has global existence of solutions by theorem 5.5-1. Notice $F'(0)$ is finite as $f(0) = 0$ in this case.

$$\text{Now } 2G + rG' = \frac{2k}{r^2+a^2} + r\left(\frac{-2rk}{(r^2+a^2)^2}\right) = \frac{2ka^2}{(r^2+a^2)^2} \qquad (5.5\text{-}16)$$

and so (5.5-8) are coupled in this case, although the coefficients of $\frac{y}{r}$ in $\ddot{x}$ and $\frac{x}{r}$ in $\ddot{y}$ can be made arbitrarily small.

Obviously, it is possible to consider the more general amendments of (5.4-3) of the form,

$$\begin{cases} \dot{r} = F(r,\theta) \\ \dot{\theta} = G(r,\theta) \end{cases} \qquad (5.5\text{-}17)$$

The system (5.5-17) will have global existence of solutions if it is δ close to rank p to (5.5-1) at all points of (r,θ) space, so that $F(r,\theta)$, $G(r,\theta)$ remain bounded for all θ (for fixed r). This type of amendment will not be considered further in this work, the most important reason being that any amended set of Schwarzchild equations deriable from a spherically symmetric metric must take the form (5.5-1) and only amendments which do not affect spherical symmetry are considered in Chapter six.

5.6) Properties of the Amended Two-Dimensional Newtonian Equations

In this section, various properties of equations of the type,

$$\begin{cases} \dot{r} = \sqrt{2\int f(r)dr + 2\int rG^2(r)dr + A} = F(r) \\ \dot{\theta} = G(r) \end{cases} \qquad (5.6\text{-}1)$$

are investigated, for reasons stated in the last section. For (5.6-1) to be a feasible amendment of (5.4-3), it is necessary for $\int f(r)\,dr \to \frac{c}{r}$ as $r \to \infty$ and $G(r) \to \frac{k}{r^2}$ as $r \to \infty$. It will also be assumed that $f(r)$ is continuous so that (5.6-1) leads to a one-dimensional, second order equation which is dynamical, for $G(r) \equiv 0$. Due to the hypothesis on $F(r)$ and $G(r)$, $F(r)$ and $G(r)$ are non-singular at infinity. In addition, it will be assumed that (5.6-1) has global existence of solutions, so that by theorem 5.5-1, $F'(r)$ is finite at each finite critical point and $F(r)$ and $G(r)$ are continuous or $F(r)$ is imaginary in an arbitrarily small neighbourhood of all finite singular points of F and G.

Reverting to cartesian co-ordinates, (5.6-1) becomes

$$\begin{cases} \dot{x} &= \frac{x}{r} F(r) - y\, G(r) = P(x,y) \\ \dot{y} &= \frac{y}{r} F(r) + x\, G(r) = Q(x,y) \end{cases} \qquad (5.6\text{-}2)$$

(5.6-2) has the plane $\mathbb{R}^2$ as its phase space. Qualitative behaviour of orbits of (5.6-2) will be exactly the same as qualitative behaviour of solutions of (5.6-1) in two-dimensional space. We first consider possible critical points of (5.6-2).

5.6.1) Critical Points

Now critical points of (5.6-2) are given by,

$$\frac{x}{r} F(r) - yG(r) = \frac{y}{r} F(r) + xG(r) = 0. \qquad (5.6\text{-}3)$$

Thus, critical points are given by $F(r) = G(r) = 0$ and there is also a critical point at $(0,0)$ if $F(0) = 0$, (noting that $F'(0)$ is assumed to be finite and $\lim_{r \to 0} \frac{F(r)}{r} = F'(0)$).

Suppose there exists a finite value of r, $r = a$ say, at which F and G both vanish. This region is a solution of (5.6-1) and represents a circle of non-isolated critical points in $\mathbb{R}^2$. Any orbit approaching $r = a$ will take an infinite time to get there, and so in this sense, it has the same effect as a limit-cycle. Thus, it has the same effect for any observer as the Schwarzchild radius does for a distant observer in the relativistic case. It represents an inpenetratable barrier surrounding the origin, and so in this case, if (5.6-1) are singular at $r = 0$, the singularity cannot be reached. Notice, however, this will only be the case for the particular choice of arbitrary constants in $F(r)$, $G(r)$, making $r = a$, a critical circle.

Consider now the case where the only critical point is situated at the origin, so that $F(0) = 0$.

$$\frac{\partial P}{\partial x}(0,0) = \lim_{h \to 0} \frac{1}{h} \left\{ P(h,0) - P(0,0) \right\} \qquad \text{by definition.}$$

$$\therefore \quad \frac{\partial P}{\partial x}(0,0) = \lim_{h \to 0} \frac{F(h)}{h} = F'(0). \quad \text{Similarly,} \quad \frac{\partial Q}{\partial y}(0,0) = F'(0) \qquad (5.6\text{-}4)$$

$$\frac{\partial P}{\partial y}(0,0) = \lim_{h \to 0} \frac{1}{h} \left\{ P(0,h) - P(0,0) \right\}$$

$$\therefore \quad \frac{\partial P}{\partial y}(0,0) = \lim_{h \to 0} \frac{1}{h} \left\{ -hG(h) \right\} = -G(0). \quad \text{Similarly,} \quad \frac{\partial Q}{\partial x}(0,0) = G(0) \qquad (5.6\text{-}5)$$

Thus, by Taylor's series, the linearized system in a neighbourhood of (0,0) must be,

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} F'(0) & -G(0) \\ G(0) & F'(0) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \qquad (5.6\text{-}6)$$

Eigenvalues of the matrix in (5.6-6) are given by,

$$(\lambda - F'(0))^2 = -G^2(0) \Rightarrow \lambda = F'(0) \pm i\, G(0). \qquad (5.6\text{-}7)$$

The following theorem follows from lemma 3.5-1, theorem 3.5-3, and section 3.6:-

Theorem 5.6-1

Let (0,0) be a critical point of (5.6-2).

(i) If $F'(0) \neq 0$, $G(0) \neq 0$, then (0,0) is a focus.

(ii) If $F'(0) = 0$, $G(0) \neq 0$, then (0,0) is a centre or multiple focus.

(iii) If $F'(0) \neq 0$, $G(0) = 0$, then (0,0) is a node or focus.

(iv) If $F'(0) = 0$, $G(0) = 0$, (0,0) is a multiple critical point.

(v) If $F'(0)$ is imaginary, the linearized system in a neighbourhood of the origin is complex and so there is a region of phase space containing the origin where the orbits of (5.6-2) are not defined. In this case, the origin is completely isolated.

□

The above theorem, and indeed, much of the analysis of this section will prove useful for the discussion of amended Schwarzchild models in Chapter six.

Using (5.6-1), the expression for $F'(0)$ is easily obtained in terms

of $f(0)$, $G(0)$:-

$$F'(0) = \sqrt{\lim_{h\to 0} \frac{1}{h^2}\left\{2\int f(h)dh + 2\int hG^2(h)dh + A\right\}}\,, \quad \text{provided the limit exists.}$$

$$= \sqrt{\lim_{h\to 0} \left(\frac{2f(h) + 2hG^2(h)}{2h}\right)} \qquad (5.6\text{-}8)$$

In order that the limit exists, $f(0) = 0$, and then,

$$F'(0) = \sqrt{\lim_{h\to 0} \left(f'(h) + G^2(h)\right)} = (f'(0) + G^2(0))^{\frac{1}{2}}. \qquad (5.6\text{-}9)$$

Notice that, by hypothesis, $f'(0)$ exists. Also, from theorem 5.5-1, and (5.6-8), any viable amendment of (5.4-3), possessing a critical point at the origin, must have $f(0) = 0$. Using (5.6-9) and theorem 5.6-1, it is seen that the origin is a simple critical point if $f'(0) + G^2(0) > 0$, multiple if $f'(0) + G^2(0) = 0$ and completely isolated if $f'(0) + G^2(0) < 0$.

Recall that for negative values of A in the Newtonian case, there exist periodic, closed orbits of the two-dimensional Newtonian equations. Thus, any viable amendment to the Newtonian equations should have some periodic or almost periodic solutions. The conclusions of theorem 5.6-1 are encouraging in this respect. In case (iv) of the theorem, both eigenvalues are equal and the linearized matrix is zero, and so the analysis of section 3.6) is inappropriate.

Consider now case (v) of theorem 5.6-1. As the origin is completely isolated, it can only have a limited effect on the topological nature of other orbits, and so the critical point is no longer important.

Example 5.6-1

Consider again the models (5.5-12) and (5.5-15), which can be combined into the following system of equations,

$$\begin{cases} \dot{r} = \sqrt{\dfrac{2c}{(r^2+a^2)^{\frac{1}{2}}} - \dfrac{k^2}{(r^2+\alpha^2)} + A} \\ \dot{\theta} = \dfrac{k}{r^2+\alpha^2} \end{cases} \qquad (5.6\text{-}10)$$

where $|a| << 1$, $\alpha = 0$ or a. If $\alpha = 0$, the model is singular and there are no critical points at finite values of r. Suppose that $\alpha = a$. There are no critical circles - the only critical point is at $(0,0)$ and this is when, $A = \dfrac{k^2}{a^2} - \dfrac{2c}{a}$. In this case,

$f'(0) = \left\{ \dfrac{-c}{a^3} + \dfrac{k^2}{a^4} \right\}^{\frac{1}{2}}$, by (5.6-9).

By theorem 5.6-1, the origin is a focus if $k^2 - ac > 0$, a centre or multiple focus if $k^2 = ac$ and completely isolated if $k^2 < ac$.

Example 5.6-2

Consider the models,

$$\begin{cases} \dot{r} = \sqrt{\dfrac{2\,cr^n}{(r^{n+1}+a^{n+1})} - \dfrac{k^2}{r^2+\alpha^2} + A} \\ \dot{\theta} = \dfrac{k}{r^2+\alpha^2} \end{cases} \qquad (5.6\text{-}11)$$

where $n = 1$ or 3 and $|\alpha| << 1$. If $k = 0$, (5.6-11) reduces to the one-dimensional models of example 5.2-7. Once again $\dot{\theta} \neq 0$ for any finite values of r. There are no critical points if $\alpha=0$ and in this case, (5.6-11) is a feasible amendment to (5.4-3) if $\dot{r}$ is real is an arbitrarily small neighbourhood of $r = 0$ by theorem 5.5-1. There is a critical point at $(0,0)$ $(r = 0)$ if

$A = \dfrac{k^2}{\alpha^2}$.

If $n = 1$, $F'(0)$ does not exist using (5.6-8).

If $n = 3$, $F'(0) = \left\{ \frac{k^2}{\alpha^4} + 0 \right\}^{\frac{1}{2}} = \pm \frac{k}{\alpha^2}$ by (5.6-9).

Thus, for $n = 1$, (5.6-11) is not a suitable amendment for $A = \frac{k^2}{\alpha^2}$ but for $n = 3$, (5.6-11) is a suitable amendment for all values of A by theorem 5.5-1 and in this case, if $A = \frac{k^2}{\alpha^2}$, the origin is a focus by theorem 5.6-1.

In contrast to the one-dimensional Newtonian case, it is difficult to construct a Lyapunov function for (5.6-2), so that a region of stability is not easily concluded if (0,0) is a multiple critical point. A completely isolated critical point need not be stable in the usual sense - it may be if it is surrounded by a limit cycle, but it is certainly never asymptotically stable. If (0,0) is a focus, the following lemma follows from theorem 3.8-1:-

Lemma 5.6-2

If the origin of (5.6-2) is a simple critical point, it is asymptotically stable if $F'(0) < 0$ and unstable if $F'(0) > 0$.

□

In (5.6-1), F(r) need not be analytic, even if G(r) and f(r) are, due to the presence of a root, so that (5.6-2) is not necessarily analytic (and so not necessarily dynamical). Suppose that f(r), G(r) are analytic functions of r. Then F(r) can be written,

$$F(r) = \left\{ a_0 + a_1 r + a_2 r^2 + a_3 r^3 + \ldots\ldots \right\}^{\frac{1}{2}} \qquad (5.6\text{-}12)$$

in a neighbourhood of the origin. Suppose that $F(0) = F'(0) = 0$, so that $(x,y) = (0,0)$ is a multiple focus or centre of (5.6-2), by theorem 5.6-1.

$F(0) = 0 \Rightarrow a_0 = 0 \therefore a_1 = 0$ for $F'(0)$ to exist.

$$\therefore F(r) = r \left\{ a_2 + a_3 r + a_4 r^2 + \ldots\ldots \right\}^{\frac{1}{2}} \Rightarrow$$

$a_2 = 0$ if $F'(0) = 0$.

Thus, $F(r)$ can be written in the form,

$$F(r) = r\left\{a_3 r + a_4 r^2 + a_5 r^3 + \ldots\right\}^{\frac{1}{2}} \qquad (5.6\text{-}13).$$

Let a_n be the first non-zero coefficient of r^{n-2} in (5.6-13). Then,

$$f(r) = a_n^{\frac{1}{2}} r^{n/2}\left\{1 + \tfrac{1}{2}\left(\frac{a_{n+1}r}{a_n} + \frac{a_{n+2}r^2}{a_n} + \ldots\right) - \frac{1}{8}\left(\frac{a_{n+1}r}{a_n} + \frac{a_{n+2}r^2}{a_n} + \ldots\right)^2 + \ldots\right.$$

$$(n \geq 3) \Rightarrow F(r) = a_n^{\frac{1}{2}} r^{n/2}\,(1 + b_{n+1}r + b_{n+2}r^2 + b_{n+3}r^3 + \ldots); \quad n \geq 3 \qquad (5.6\text{-}14)$$

for some coefficients b_{n+i}. Thus, $F(r)$ will only be analytic if n is even, (so that $n \geq 4$). By hypothesis, $G(0) \neq 0$ and so in a neighbourhood of the origin, $G(r)$ can be expanded in the form,

$$G(r) = c_o + c_1 r + c_2 r^2 + c_3 r^3 + \ldots \;;\; c_o \neq 0. \qquad (5.6\text{-}15)$$

From (5.6-14), (5.6-15), (5.6-1), we get,

$$\frac{dr}{d\theta} = \frac{F(r)}{G(r)} = \frac{a_n^{\frac{1}{2}} r^{n/2}\,(1 + b_{n+1}r + b_{n+2}r^2 + \ldots)}{c_o + c_1 r + c_2 r^2 + c_3 r^3 + \ldots} \Rightarrow$$

$$\frac{dr}{d\theta} = \frac{a_n^{\frac{1}{2}} r^{n/2}}{c_o}(1 + b_{n+1}r + b_{n+2}r^2 + \ldots)\left\{1 - \left(\frac{c_1 r}{c_o} + \frac{c_2 r^2}{c_o} + \ldots\right)\right.$$

$$\left. + \left(\frac{c_1 r}{c_o} + \frac{c_2 r^2}{c_o} + \ldots\right)^2 \ldots.\right\}$$

$$\therefore \frac{dr}{d\theta} = r^{n/2}\left\{d_n + d_{n+1}r + d_{n+2}r^2 + \ldots\right\} \qquad (5.6\text{-}16)$$

for some coefficients d_{n+i} $(d_n \neq 0)$.

In the analytic case, it can easily be seen that from (5.6-16), the second focal value will not vanish, as the coefficients of r

are constants in this case and do not depend on θ. Thus, from section 3.10, the origin must be a multiple focus (not necessarily of definite multiplicity). The only exception will be if $F(r) \equiv 0$ as then all focal values vanish and $\frac{dr}{d\theta} = 0$ and so $(0,0)$ is clearly a centre.

It can be seen that investigation of the critical points of (5.6-2) does not give nearly so much information as in the one-dimensional case due to the fact that there may only be critical points for particular values of k and A, whereas solutions of (5.6-2) need to be investigated for all values of A,k and also, in some cases, the critical points are completely isolated and give no information.

5.6-2) Existence of Closed Paths

The two-dimensional Newtonian equation has closed curves (elliptic orbits) in its phase space and so one of the most important aspects of the properties of the (5.6-2) concerns the existence of closed paths or almost periodic solutions. Clearly, if for a given choice of $F(r)$, $G(r)$, (5.6-2) has no periodic or almost periodic solutions, it is not a feasible amendment to (5.4-3). If the right-hand sides of (5.6-2) are non-singular and differentiable at most values of r and the equations have no finite critical points, then there can be no closed orbits. This is not, in general, true if the right-hand sides of (5.6-2) are singular - a system of closed curves can contain a non-critical point if it is singular, as in the Newtonian case.

In the non-singular case, Dulac's test gives conditions for the non-existence of closed curves. A particularly useful form of the test is the following lemma:-

Lemma 5.6-4

Let G be a simply connected sub-region of $\mathbb{R}^2 - \{0\}$. There are not closed curves of (5.6-2) in G if the function,

$$\Phi_1(r) = 2\int f(r)dr + 2\int rG^2(r)dr + A + r(f(r) + r\,G^2(r))$$

does not change sign in G.

□

Proof

Applying theorem 3.5-7 to (5.6-2),

$$\Phi(r) = \frac{\partial B}{\partial x}\left(\frac{x}{r}F - yG\right) + \frac{\partial B}{\partial y}\left(\frac{y}{r}F + xG\right) + B\left(\frac{1}{r}F + F'\right)$$

Let $B = rF(r)$ (and so is not continuously differentiable at the origin). Then,

$$\Phi = \left(\frac{x}{r}F + rF'\frac{x}{r}\right)\left(\frac{x}{r}F - yG\right) + \left(\frac{y}{r}F + rF'\frac{y}{r}\right)\left(\frac{y}{r}F + xG\right) + F^2 + rFF'$$

$$\therefore \Phi = 2(F^2 + rFF') \quad \Rightarrow$$

$$\Phi = 2\left\{2\int f(r)dr + 2\int rG^2(r)dr + A + r(f(r) + rG^2(r))\right\}$$

Notice that the test is also applicable in the singular case, if the only singularity occurring in $F(r)$ and $G(r)$ is the origin.

In the non-singular case, lemma 5.6-4 could possibly be used to differentiate between a focus and a centre, and could also be used to detect limit-cycles in the case where (0,0) is completely isolated, but it may not detect a closed curve $r = a$ where $F(a) = 0$.

Example 5.6-3

Consider again the model (5.6-11) with $n = 1$, $\alpha = 0$. (If $n = 3$ and $\alpha \neq 0$, the model is non-singular and the only possible finite critical point is situated at (0,0) and is a focus by theorem 5.6-1 and so there are no closed curves except possibly at $r = a$ where $\dot{r}(a) = 0$.) In this case, the model is singular and may only be viable for certain values of k,A - theorem 5.6-1 gives no information about closed curves.

From lemma 5.6-4,

$$\Phi_1(r) = \frac{2cr}{r^2+a^2} - \frac{k^2}{r^2} + A + r\left\{\frac{a^2c - cr^2}{(r^2+a^2)^2} + \frac{k^2}{r^3}\right\}$$

$$= \frac{A(r^2+a^2)^2 + cr^3 + 3cra^2}{(r^2+a^2)^2} . \qquad (5.6\text{-}17)$$

If $A = 0$, (5.6-17) becomes,

$$\Phi_1(r) = \frac{cr(r^2+3a^2)}{(r^2+a^2)^2} \qquad (5.6\text{-}18)$$

For positive r, (5.6-18) does not change sign and so there can be no closed orbits in $\mathbb{R}^2 - \{0\}$, (If $A = 0$, k can be chosen so that the singularity is not reached), except possibly at $r = a$ where $\dot{r}(a) = 0$.

In contrast to the one-dimensional Newtonian amendments, there are no specialized theorems concerning the existence of periodic orbits of equation (5.6-2). Lemma 3.5-6 is not appropriate either as, $P(x,-y) \neq \pm P(x,y)$, $P(-x,y) \neq \pm P(x,y)$ and similarly for $Q(x,y)$, so that the orbits are not symmetrical about either axis, (assuming that $F(r) \not\equiv 0$). This agrees with previous analysis which concluded that if (0,0) is a critical point of (5.6-2), it cannot be a centre.

Theorem 3.10-8, concerning perturbations of the two-dimensional Newtonian equations (5.4-3) is difficult to apply in practice, although theoretically possible as $G \subset \mathbb{R}^2 - \{0\}$, where G is any closed annualar region surrounding the origin is completely filled with elliptic orbits of (5.4-3) for certain values of A. This is because the first order partial differential equation for ρ given in theorem 3.10-2 is impossible to solve in general. A particular solution is $\rho = \frac{1}{rF(r)}$ (if $F(r)$ is analytic in G). However, with this choice of $\rho(x,y)$, is the notation of theorem 3.10-3, $f_1(x,y), g_1(x,y)$ are given by,

$$f_1(x,y) = -y\eta(r); \quad g_1(x,y) = x\eta(r), \text{ for some function } \eta(r) \Rightarrow$$

$$f'_{1x}(x,y) = -y\eta'(r)\frac{x}{r}; \quad g'_{1y}(x,y) = x\eta'(r)\frac{y}{r} \Rightarrow$$

$$f'_{1x}(x,y) + g'_{1y}(x,y) \equiv 0.$$

Thus, the theorem cannot be applied, with this particular choice of ρ.

Consider now equation (5.6-1), and suppose that $F(a) = 0$, $a \neq 0$.

$$F'(a) = \lim_{r \to a} \frac{f(r) + r\,G^2(r)}{F(r)} \qquad (5.6\text{-}19)$$

Thus, in order for $F(r)$ to be differentiable at $r = a$, $f(a) + aG^2(a) = 0$

Clearly, if $F(a) = 0$, $r = a$ is an isolated closed orbit of (5.6-1) if a is real. As mentioned earlier, if $G(a) = 0$ (so that $f(a) = 0$ in order that $F(r)$ is differentiable at $r = a$), $r = a$ is a critical circle and must have the effect of a limit cycle. (If, in this case, $F'(a)$ does not exist, the system (5.6-1) does not have global existence of solutions and is not a viable amendment to (5.4-3).) Suppose that $G(a) \neq 0$. If $F'(a)$ does not exist, (5.6-1) does not have uniqueness of solutions on $r = a$ and so any other orbit may cross $r = a$ or reach $r = a$ in finite time. Thus, a necessary condition for $r = a$ to be a limit cycle is that, from (5.6-19), $f(a) + aG^2(a) = 0$. (This can always be accomplished by choosing the arbitrary constant in $G(r)$ suitably).

Let $\phi(x,y) = x\left\{\frac{xF(r)}{r} - y\,G(r)\right\} + y\left\{\frac{y}{r}F(r) + xG(r)\right\}$

$$\therefore \quad \phi(x,y) = rF(r). \qquad (5.6\text{-}20)$$

Then, using theorem 3.5-8, from (5.6-20) we have,

Lemma 5.6-5

In (5.6-1), let $F(a) = 0$. Then $r = a$ is a limit-cycle if $F'(a)$ exists (so that $f(a) + a\,G^2(a) = 0$), and $F(r)$ is real in a neighbourhood of $r = a$. □

In this case, even if $r = 0$ is a singularity, (5.6-1) still represents a feasible amendment as orbits cannot pass $r = a$.

Example 5.6-4

Consider the model (5.6-11) with $n = 1$, $\alpha = 0$, and $\dot{\theta} = \frac{k}{r^2} + \gamma$,

so that,

$$\begin{cases} \dot{r} = \sqrt{\frac{2cr}{r^2+a^2} - \frac{k^2}{r^2} + \gamma^2 r^2 + A} \\ \dot{\theta} = \frac{k}{r^2} + \gamma \end{cases} \qquad (5.6\text{-}21)$$

Now $F(a) = \left(\frac{c}{a} - \frac{k^2}{a^2} + \gamma^2 a^2 + A \right) \Rightarrow$

$$F(a) = 0 \text{ when } A = \frac{-c}{a} + \frac{k^2}{a^2} - \gamma^2 a^2 \qquad (5.6\text{-}22)$$

$$G(a) = \frac{k}{a^2} + \gamma \Rightarrow G(a) = 0 \text{ when } \gamma = \frac{-k}{a^2}. \qquad (5.6\text{-}23)$$

Notice that in the case of (5.6-21), $f(a) = 0$ and so from (5.6-19), $r = a$ is a critical circle if $A = \frac{-c}{a}$ (combining (5.6-22), (5.6-23)).

Suppose now that $\gamma = 0$, so that $G(r) \neq 0$ for any finite r. Then from (5.6-19), $r = a$ cannot be a limit-cycle. However, if $r = b \neq a$,

$$f(b) + bG^2(b) = \frac{c(a^2-b^2)}{(a^2+b^2)^2} + \frac{k^3}{b^3} = 0 \Rightarrow$$

$$k^2 = \frac{cb^3(b^2-a^2)}{(a^2+b^2)^2} \qquad (5.6\text{-}24)$$

so that if A is chosen is in (5.6-22), with k^2 as given in (5.6-24), $r = b$ is a limit-cycle of (5.6-21), by lemma 5.6-5.

If (5.6-1) is a viable amendment to (5.4-3), $F(r)$ may be written in the form,

$$F^2(r) = \frac{2c}{r} - \frac{k^2}{r^2} + A + \varepsilon\eta(r) \qquad (5.6\text{-}25)$$

where ε is a small constant and $\eta(r) \to 0$ as $r \to \infty$. For $\varepsilon = 0$ and

certain values of A, orbits of (5.6-1) are closed. If $\varepsilon \neq 0$, the rate of change of r will be affected more and more as the origin is approached and so orbits of (5.6-1) will be perturbations of closed orbits for these values of A and provided $|\varepsilon| << 1$, it is reasonable to suppose that the periodic orbits will be perturbed to almost periodic orbits. This conjecture is supported by the numerical results of various orbits of models given in section 5.8).

In polar form, equation (5.6-2) becomes (5.6-1) and so the following integral is obtained:-

$$\int \frac{G(r)}{F(r)}\, dr = \theta + \theta_o \qquad (5.6\text{-}26)$$

where θ_o is a constant of integration. The question of the type of orbits obtained in the cases where (0,0) is a multiple critical point of (5.6-2) or completely isolated and the question of closed orbits may be answered if the integral (5.6-26) can be solved. Substituting the expression for F(r) into (5.6-26) using (5.6-1), the following integral is obtained,

$$\int \frac{G(r)dr}{\sqrt{2\int f(r)dr + 2\int rG^2(r)dr + A}} = \theta + \theta_o \qquad (5.6\text{-}27)$$

and the conditions on f(r) and G(r) make the evaluation of the integral by analytical means, in general, impossible. However, there are certain cases where (5.6-27) can be evaluated in terms of elliptic functions.

Example 5.6-5

Consider the two-dimensional model,

$$\begin{cases} \dot{r} = \sqrt{\dfrac{2cr^3 + 2\gamma}{(r^2 + a^2)^2} - \dfrac{k^2}{r^2 + a^2} + A} \\ \dot{\theta} = \dfrac{k}{r^2 + a^2} \end{cases} \qquad (5.6\text{-}28)$$

where $|a|, |\gamma| << 1$. From theorem 5.5-1, it is a viable amendment to the two-dimensional Newtonian equation. (The only finite critical point is at the origin, when $A = \frac{k^2}{a^2} - \frac{2\gamma}{a^4}$). In this case, (5.6-26) becomes

$$I = \int \frac{k\,dr}{(r^2+a^2)\sqrt{\frac{2cr^3+2\gamma}{(r^2+a^2)^2} - \frac{k^2}{r^2+a^2} + A}} = \theta + \theta_o \Rightarrow$$

$$I = \int \frac{k\,dr}{\sqrt{Ar^4 + 2cr^3 + (2a^2A-k^2)r^2+(2\gamma-k^2a^2+Aa^4)}} \qquad (5.6\text{-}29)$$

which is expressible in terms of elliptic functions[10].

Now (5.6-9) gives,

$$F'(0) = f'(0) + G^2(0) = \frac{-4\gamma}{a^6} + \frac{k^2}{a^4} = \frac{1}{a^6}\left\{k^2a^2 - 4\gamma\right\}. \qquad (5.6\text{-}30)$$

Using (5.6-30), from theorem 5.6-1, there are three cases to consider when $A = \frac{k^2}{a^2} - \frac{2\gamma}{a^4}$, so that (0,0) is a critical point of (5.6-2):-

(i) if $\gamma = \frac{k^2a^2}{4}$, (0,0) is a centre or focus

(ii) if $\gamma < \frac{k^2a^2}{4}$, (0,0) is a focus

(iii) if $\gamma > \frac{k^2a^2}{4}$, (0,0) is completely isolated.

Case (i)

Let $\gamma = \frac{k^2a^2}{4}$ and $A = \frac{k^2}{a^2} - \frac{2\gamma}{a^4} = \frac{k^2}{2a^2}$, in (5.6-29).

$$\therefore I = \int \frac{k\,dr}{\sqrt{Ar^4+2cr^3}} = \int \frac{k\,dr}{r^{3/2}\sqrt{Ar+2c}}$$

$$\therefore \quad I = -\frac{k}{c}\sqrt{\frac{Ar+2c}{r}} = \theta + \theta_o \quad \Rightarrow$$

$$r = \frac{2c}{(c^2/k^2)(\theta+\theta_o)^2 - A} \qquad (5.6\text{-}31)$$

and it can be seen that the orbit in the (x,y) plane is a focus and so in this case, we can differentiate between a centre and a focus directly.

Case (ii)

Take, for example, $\gamma = 0$ and $A = \frac{k^2}{a^2}$.

$$\text{Then } I = \int \frac{k\ dr}{\sqrt{Ar^4+2cr^3+Aa^2r^2}} = \int \frac{k\ dr}{r\sqrt{Ar^2+2cr+Aa^2}} \qquad (5.6\text{-}32)$$

from (5.6-29). The integral (5.6-32) is standard and the following results are obtained, (noticing that $A > 0$):-

$$I \begin{cases} = \frac{k}{a\sqrt{A}} \ln\left(\frac{r}{r+a}\right) = \ln\left(\frac{r}{r+a}\right); & c = Aa \\ = -\ln\left\{\frac{2k\sqrt{Ar^2+2cr+Aa^2}+2cr+2Aa^2}{r}\right\} & ; \quad c \neq Aa. \end{cases} \qquad (5.6\text{-}33)$$

and so, from (5.6-33), it can be seen that in either case, the orbits are foci, agreeing with theorem 5.6-1.

Case (iii)

$$\text{Now } A = \frac{k^2}{a^2} - \frac{2\gamma}{a^4} \quad \Rightarrow$$

$$I = \int \frac{k\ dr}{\sqrt{Ar^4+2cr^3+(2a^2A-k^2)r^2}} = \int \frac{k\ dr}{r\sqrt{Ar^2+2cr+(2a^2A-k^2)}} \qquad (5.6\text{-}34)$$

which is of the form of (5.6-32), but now, $A < 0 \Rightarrow 2a^2-k^2 < 0$ (as $\gamma > k^2a^2/4$).

$$\therefore I = \frac{1}{\sqrt{k^2-2a^2A}} \sin^{-1}\left\{\frac{cr + (2a^2A - k^2)}{|r|\sqrt{c^2-A(2a^2-k^2)}}\right\} \qquad c \neq A(2a^2A-k^2)$$

$$\therefore r = \frac{2a^2A - k^2}{(\sqrt{c^2-A(2a^2A-k^2)}\sin(\theta\sqrt{k^2-2a^2A}) - c)} \qquad (\theta_o=0) \quad (5.6\text{-}35)$$

The orbits will be periodic in this case if,

$$\frac{c}{\sqrt{c^2-A(2a^2A-k^2)}} > 1$$

(as then r cannot become infinite). This will always be the case since $A(2a^2A-k^2) > 0$. (It is assumed that $c^2 > A(2a^2A-k^2)$ so that I in (5.6-34) is real).

If (0,0) is not a critical point of (5.6-28) expressed in terms of cartesian co-ordinates (5.6-29) is expressible in terms of inverse elliptic functions so that r is expressible in terms of elliptic functions of θ, which are periodic. If r does not become infinite for some value of θ, then the orbits must be closed, for given specific values of A,c,k,γ. However, (5.6-29) is difficult to evaluate for arbitrary values of A,c,k,γ.

5.6.3) Structural Stability

Although (5.6-1) is not structurally stable in the four-dimensional $(r,\dot{r},\theta,\dot{\theta})$ phase space it may be structurally stable in the plane. If arbitrary constants in F(r), G(r) are chosen so that (5.6-2) possesses no critical points and F(r), G(r) are chosen non-singular, then there can be no closed orbits and by theorem 3.9-1, the system must be structurally stable. If (5.6-2) possesses a critical circle, each point on the circle is a non-elementary critical point and so the system must be structurally unstable. If (0,0) is the only critical point of (5.6-2), by theorem 5.6-1, there can be no orbits going from saddle point to saddle point. If (0,0) is completely isolated, orbits in the phase plane are not directly affected by its presence and so from this point of view, it must be regarded as non-elementary. (Notice if F(r) or G(r) are chosen singular in (5.6-2), we cannot conclude structural stability as the equations are not derivable from a vector field in a neighbourhood of the singularity).

In conclusion, we have, using theorem 5.6-1, and theorem 3.9-1,

Lemma 5.6-5

The system (5.6-2) is structurally stable in a closed region G of $\mathbb{R}^2$ if, F(r) and G(r) are non-singular, and differentiable at all points in G, and

(i) (5.6-2) contains no critical points, or
(ii) $F'(0) \neq 0$ and real if (0,0) is the only critical point. □

If (0,0) is the only possible critical point of (5.6-2), only specific values of the arbitrary constants in (5.6-2) will bring about its presence, so that for most values of the arbitrary constants, there are no critical points. Thus, no matter what the character of the critical point at (0,0), (5.6-2) will be structurally stable for almost all values of the arbitrary constants.

Notice that the amendments (5.5-16), and (5.6-11) with $a \neq 0$, $n = 3$, are structurally stable by lemma 5.6-5.

5.7 Various Numerical Solutions of Specific Models

So far, viable amendments to the one and two-dimensional Newtonian equations have been considered in a fairly general context. In this section, various specific models are considered and numerical and graphical solutions are obtained. It is thus seen how close the considered model agrees with the one or two dimensional Newtonian model within a given region. These solutions also support the general theory of the previous section in the given specific cases.

A few selected models are considered, which have certain desirable properties. Of course, from the given theory, it is possible to generate any given number of models any of which are viable.

Example 5.7-1

Consider the equation.

$$\ddot{r} = -\frac{2c}{\pi}\,\frac{\tan^{-1}(br^n)}{r^2} + \alpha g(r)\dot{r}, n\varepsilon^{+}, \qquad (5.7\text{-}1)$$

From example 5.2-3, theorem 5.2-8, (5.7-1) is a dynamical amendment of the one-dimensional equation (5.3-1) if $n \geq 2$, and $g(r)$ is analytic with $\lim_{r\to\pm\infty} \int g(r)dr$ finite. In order to agree as close as possible with (5.3-1), while still retaining the non-singular character, α and b are chosen so that $b \gg 1$, $|\alpha| \ll 1$, $\alpha \in \mathbb{R}$, $b \in \mathbb{R}^+$. The equation has a critical point at (0,0) in $(r,\dot{r})$ space if $n > 2$, and $f'(0) = -\frac{2bc}{\pi}$, $n = 3$, or 0, otherwise. Thus, (0,0) if simple if $n = 3$ and a focus if $g(0) \neq 0$ or centre/multiple focus if $g(0) = 0$ (including the case $\alpha = 0$), by lemma 5.3-1. If $n > 3$, $\alpha \neq 0$ (0,0) is any type of multiple critical point of the first kind, depending on the choice of $g(r)$ and if $\alpha = 0$, (0,0) is any type of multiple critical point of the second kind depending on the choice of $g(r)$. (See theorems 5.3-2, 5.3-3). If $n = 3$, (0,0) is stable if $\alpha = 0$ or asymptotically stable if $\alpha g(0) < 0$ by lemma 5.3-6. The equation is not structurally stable in $(r,\dot{r})$ space if $\alpha = 0$, but may be made structurally stable if $\alpha \neq 0$ by choosing $g(r)$ suitably (i.e. $g(0) \neq 0$) (see theorem 5.3-12).

Various solutions are plotted, in $(r,\dot{r})$ space for various values of ($\alpha = 0$) and compared with solutions of (5.3-1). If $\alpha = 0$ and $n = 2$, the curves are not symmetric about the $\dot{r}$ axis. This seems to contradict the fact that there is a central force at $r = 0$. However, there is no problem as in this case, two equations are defined by a symmetric vector field about $r = 0$ - one for $r \geq 0$ and one for $r < 0$. The equations change sign at $r = 0$. Thus, curves (1) and (2) in figure 5.2, showing the cases $b = 1$, 10 respectively, do not represent the true path of a test particle, but rather the solution curves of (5.7-1). The true paths of the test particle in this case, will be the portion of these curves for $r \geq 0$, together with their reflections in $r = 0$. (In the figure, the test particle starts from rest at $r = 5$ and c is taken to be 1 without loss of generality. For $r > 2$, the amended paths more or less agree with the Newtonian orbit, shown by curve (5). By increasing the value of b, the time taken to reach $r = 0$ can be made arbitrarily close to the time taken in the Newtonian case.

If $n = 3$, the solutions of (5.7-1) for $\alpha = 0$ are periodic.

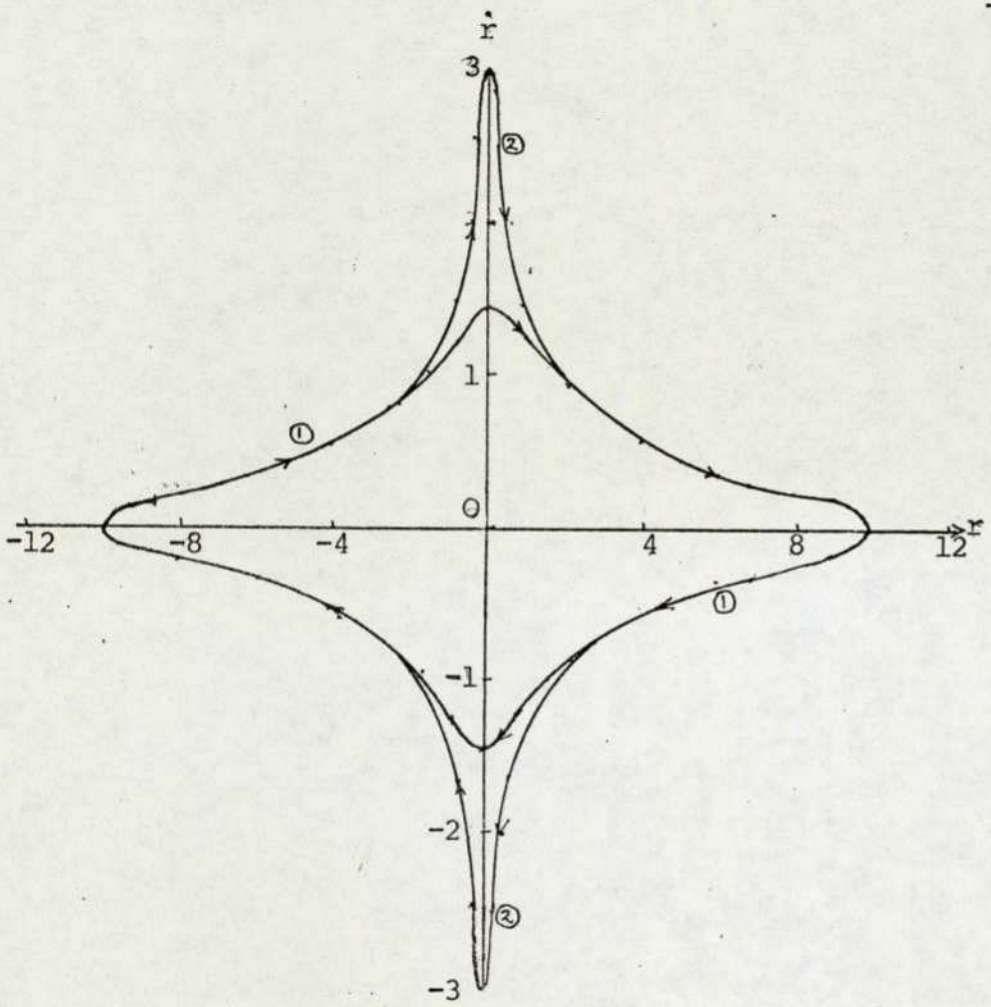


Figure 5.1

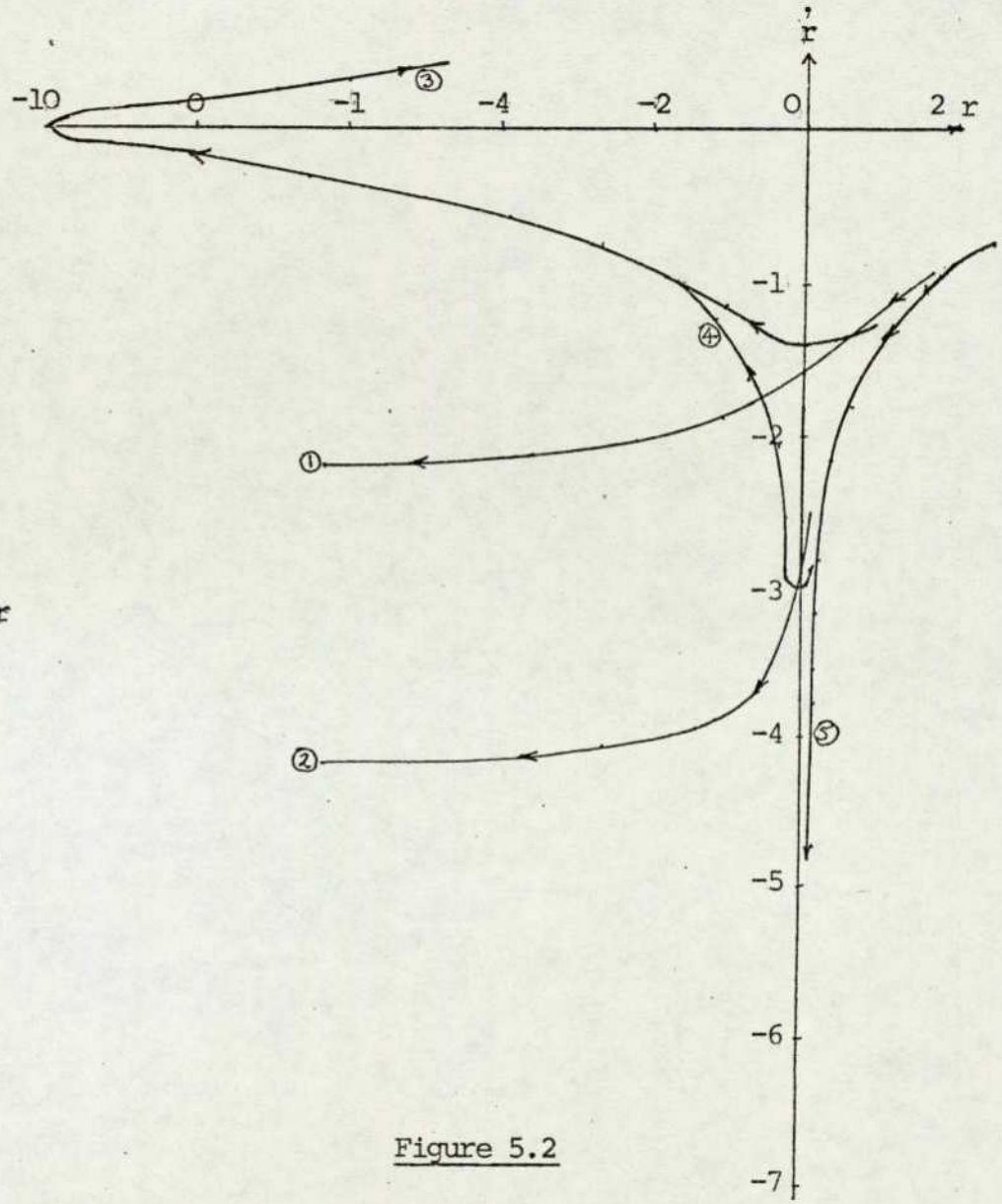


Figure 5.2

Curves (3), (4) of figure 5.2, and (1), (2) of figure 5.1, show the cases $b = 1, 100$ respectively. In these cases, a test particle will oscillate backwards and forwards through the origin, attaining its maximum velocity there.

If α is chosen non-zero in (5.7-1), the orbits in the $(r,\dot{r})$ plane will no longer be symmetric with respect to $t = 0$. As stated earlier, if $\ddot{r}(0) = 0$, $(0,0)$ can be given any chosen character by choosing $g(r)$ suitably. For instance, if $n = 3$ and $g(r) = \dfrac{r^2}{1+r^4}$, $(0,0)$ is a focus. In this case, a test particle will oscillate irregularly through the origin, the oscillations gradually increasing or decreasing, depending on the choice of α. A specific orbit around a focus is plotted in a later example.

The two-dimensional model of the form (5.6-1) corresponding to (5.7-1) is,

$$\begin{cases} \dot{r} = \left\{ 2\int f(r)dr + 2\int rG^2(r)dr + A \right\}^{\frac{1}{2}} \\ \dot{\theta} = G(r) \end{cases} \qquad (5.7\text{-}2)$$

where $f(r) = -\dfrac{2c}{\pi}\dfrac{\tan^{-1}br^n}{r^2}$. In the numerical solutions, $G(r)$ is chosen to be, $G(r) = \dfrac{1}{r^2 + a^2}$, $|a| << 1$. 'n' in $f(r)$ is chosen to be 2 or 3, so that $\int f(r)dr$ is easily obtained as (5.2-9) or (5.2-10) respectively. Thus, the following equations are considered,

$$\begin{cases} \dot{r} = \left\{ F(r) - \dfrac{K^2}{r^2 + a^2} + A \right\}^{\frac{1}{2}} \\ \dot{\theta} = \dfrac{K}{r^2 + a^2} \end{cases} \qquad (5.7\text{-}3)$$

where A, K are constants and $F(r)$ is given by (5.2-9) or (5.2-10). By theorem 5.5-1, the equations will have global existence of solutions for almost all values of A, K and will approximate to the Newtonian equations (5.4-3) as $r \to \infty$, $b \to \infty$, $a \to 0$. If $a = 0$, the force producing the motion will be central and (5.7-3) will still have global

existence of solutions if the possibility of starting at r = 0 is excluded. The equations have a critical point in the ($x = r\cos\theta$, $y = r\sin\theta$) plane at (0,0) only if $2\int f(r)dr - \frac{K^2}{a^2} + A\Big|_{r=0} = 0$, ($a \neq 0$). By theorem 5.5-1, (5.7-3) will only have global existence of solutions for these values of A,K, if n = 3. Then by theorem 5.6-1 and (5.6-9), (0,0) is either completely isolated, multiple, or a focus. If a is made small enough, it will be a focus.

By lemma 5.6-5 for most choices of A,k,a and b, the system (5.7-3) will be structurally stable in (x,y) space.

Solutions to (5.7-3) with n = 2 or 3, are plotted for particular values of A,k, as there are only two basic cases. Those which produce elliptic orbits in the Newtonian case, and those which produce non-closed orbits in this case. These values of A,k have to be amended slightly when plotting solutions of (5.7-3) so that amended solutions have the same initial conditions as the corresponding Newtonian solutions. It can be seen from the diagrams that closed orbits are perturbed slightly to almost periodic orbits.

The Newtonian orbit is taken to be a fairly eccentric one - this is curve (1) in fig. 5.3, and starts from rest at x = 5, as do all the amended orbits. The values of A and k chosen are A =-0.98, k = 1.0. (Recall that if A>0, the Newtonian orbits are parabolic or hyperbolic. The orbits of (5.7-3) agree reasonably closely to thse orbits.)

Curves (2) and (3) in fig. 5.3 show the orbits of (5.7-3) for b = 1000, n = 2 and a = 0.1 and approaching zero respectively, while curve (4) shows the orbit of (5.7-3) for b = 1000, n = 3 and a approaching zero. In all cases, the ellipse is lengthened at one end, (curves (2) -(4) following approximately the same path) and a perihelion shift is induced - the greater the value of a, the greater the extent of the perihelion shift. The shift can be made clockwise or counter-clockwise, depending on the value of a.

Example 5.7-2

Consider the set of equations given by,

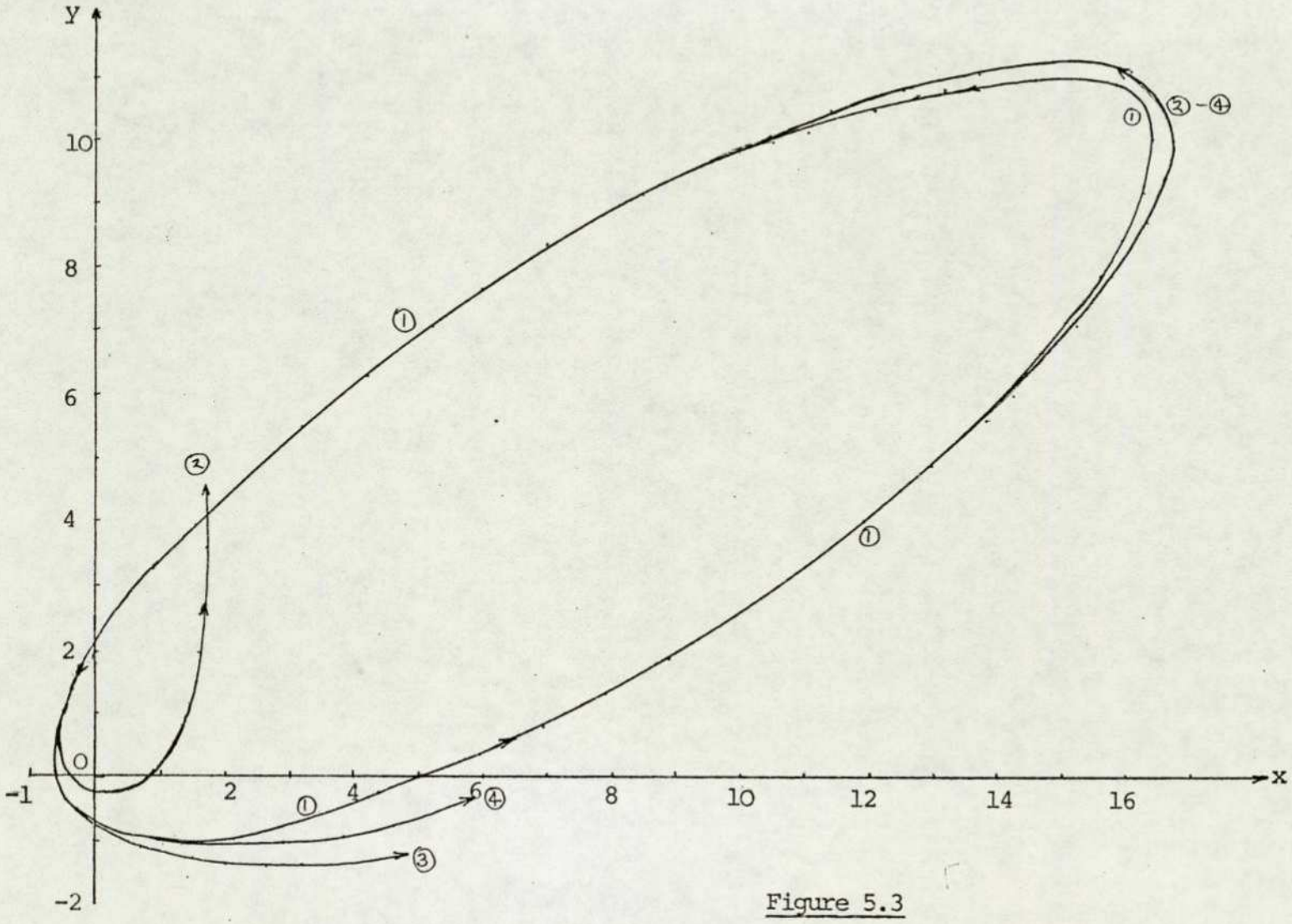


Figure 5.3

$$\ddot{r} = \frac{-cr^n}{r^{n+2} + a^{n+2}} + \alpha g(r)\dot{r}\,, \qquad (5.7\text{-}4)$$

where $|\alpha| << 1$, $|a| << 1$, $n \in \mathbb{Z}^+$.

From example 5.2-5, the equations are dynamical if $a \neq 0$ and n is even ($\alpha = 0$). However, if $a = 0$ the equation may be perturbed so that it is still singular but has global existence of solutions, except at $r = 0$ - as in example 5.2-2. To keep the integration of $r(r)$ in the two-dimensional case, as simple as possible, n is chosen to be 0 or 2. (5.7-4) with $\alpha \neq 0$ is then dynamical if $a \neq 0$ and $\lim_{r \to \pm\infty} \int g(r)dr$ is finite, by theorem 5.2-8. The equation has a critical point at (0,0) if $n = 2$. In this case, it is multiple and can be taken to have a chosen character by choosing $g(r)$ suitably - see example 5.3-3. Thus, it is not stable in the usual sense and unlike in the previous example, (5.7-4) cannot be made structurally stable by suitable choice of $g(r)$.

Once again, various solutions can be plotted in $(r,\dot{r})$ space for various values of a and choices of $g(r)$ ($\alpha \neq 0$) and compared with solutions of (5.3-1). If $a \neq 0$ and $\alpha = 0$, and the test solutions start from rest at $r = 5$, they have a very similar behaviour to the amended orbits in the previous case. For instance if $n = 0$, the acceleration at $r = 0$ is finite and the orbits of (5.7-4) will not be symmetric about $r = 0$, but the orbit of a test particle in this case will be an orbit of (5.7-4) for $r \geq 0$, together with its reflection $r = 0$ and will be symmetric about $r = 0$. For $n \geq 4$, (0,0) is a critical point - a centre, and the motion of a test particle will be oscillatory. This oscillatory nature can be made irregular by taking $\alpha \neq 0$ and $g(r)$ suitably, so that (0,0) is a focus.

The two-dimensional model of the form (5.6-1) corresponding to (5.7-4) to (5.7-2) with

$$f(r) = \frac{-cr^n}{r^{n+2} + a^{n+2}}\,; \quad n = 0,2.$$

Once again in the numerical solutions, $G(r)$ is chosen to be $G(r) = \frac{1}{r^2 + b^2}$, $|b| << 1$. $\int f(r)dr$ is easily obtained as (5.2-13) or

(5.2.-14) for n = 0 or 2 respectively. The chosen system of equations will approximate to the Newtonian equations (5.4-3) as $r \to \infty$, $b \to 0$, $a \to 0$. If b = 0, the force producing the motion will be central. The equations have a critical point in the (x,y) plane at (0,0) only if $2 f(r)dr - \frac{K^2}{b^2} + A\Big|_{r=0} = 0$ (b=0), and, like in the previous case there are no critical circles. By theorem 5.5-1, the chosen equations will only have global existence of solutions for these values of A,K if n = 2. Once again, if b is chosen small enough, (0,0) will be a focus in this case. For most choices of A,K, a and b, the chosen equations will be structurally stable in (x,y) space.

Solutions with n = 0 or 2 are plotted for particular values of A,k and once again, perturbed closed orbits become almost periodic orbits. (Once again A =-0.98, k = 1.0 in the Newtonian case, so that a fairly eccentric orbit is obtained). In the case n = 0, the amended values of A and k differ markedly from the Newtonian values - the program is difficult to run due to the presence of large constants, although the orbits have the same general character of those of the previous example. The cases n = 2 with b = 1000 and a = 0.1 and approaching 0 are shown as curves (2) and (3) respectively, in figure 5.4. These have very similar behaviour to orbits in previous cases. The curve (3) is extended to exhibit the perihelion shift in figure 5.5.

Example 5.7-3

Consider now the set of equations given by,

$$\ddot{r} = \frac{-cr}{(r^2+a^2)^{3/2}} + \alpha g(r)\dot{r} \qquad (5.7\text{-}5)$$

where $|\alpha| \ll 1$, $|a| \ll 1$.

The equations are dynamical so long as $\lim_{r\to\pm\infty} \int g(r)dr$ is finite ($a \neq 0$) or $\alpha = 0$ ($a \neq 0$). (see, for instance, example 5.2-6). In fact, the equations are considerable more tractible than in the previous cases - for instance, the first integral is given by (5.2-16) if $\alpha = 0$. The

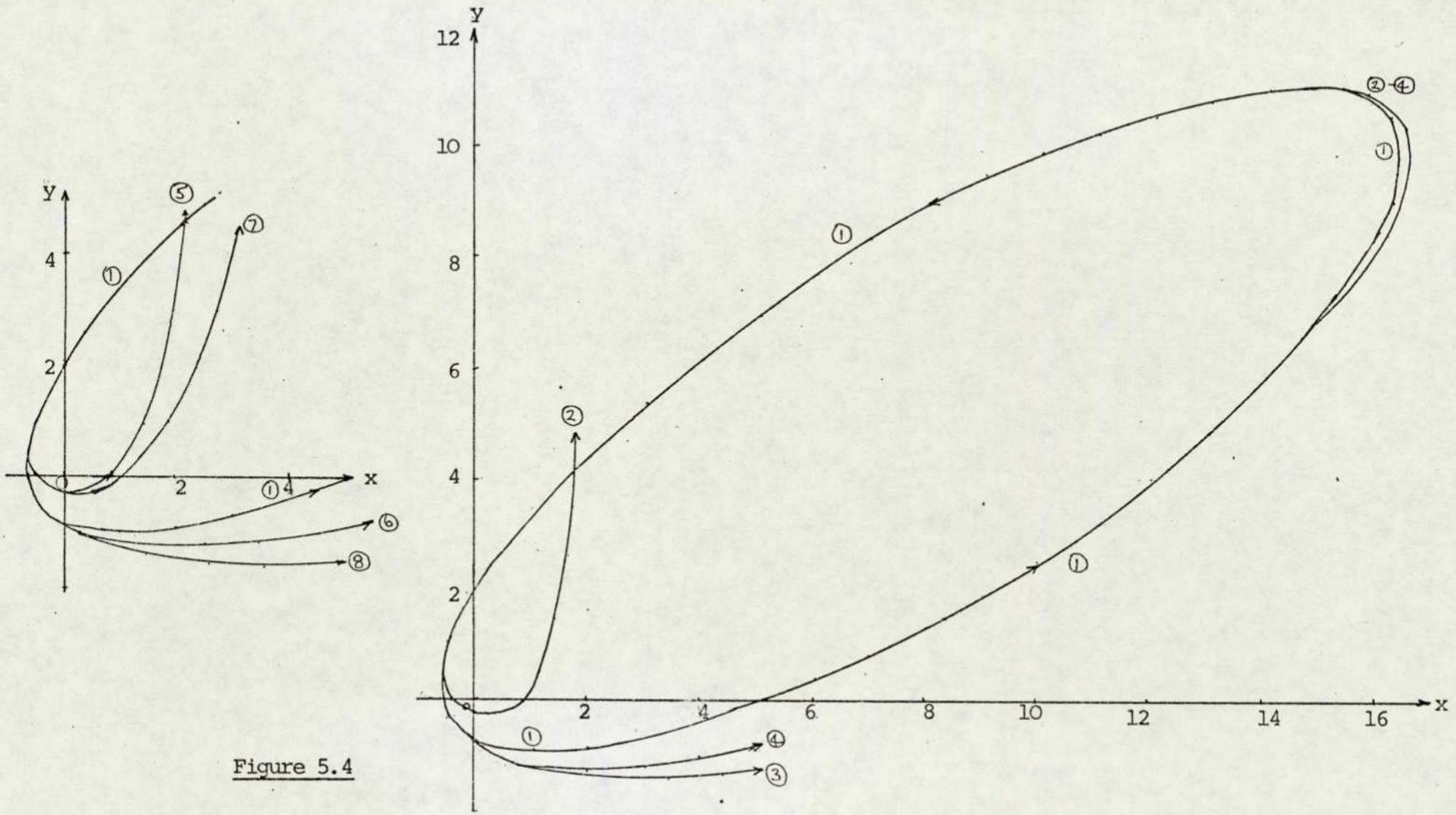


Figure 5.4

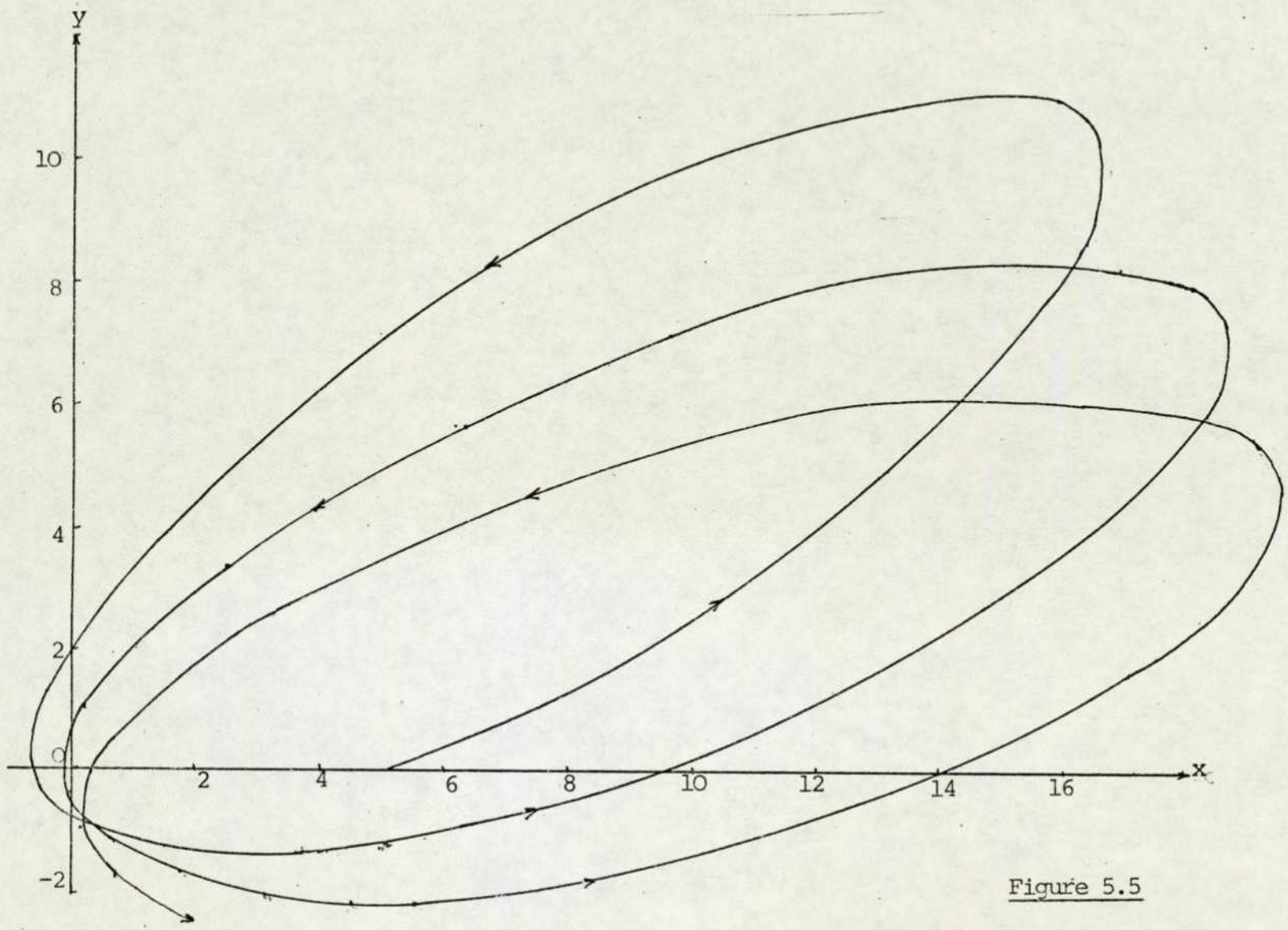


Figure 5.5

equations have a critical point in $(r,\dot{r})$ space at (0,0), which is simple and a focus, multiple focus, or centre in any feasible case, by lemma 5.3-1 (see also example 5.3-4). (0,0) must be a centre if $\alpha = 0$ by lemma 5.3-10. If $g(r)$ is chosen so that $g(0) = 0$ and $f'(0)g''(0) + g'(0)f''(0) \neq 0$, where $f(r) = \dfrac{-cr}{(r^2+a^2)^{3/2}}$, then (0,0) can be made a multiple focus (of multiplicity one) by theorem 5.3-11. The equations can be made structurally stable in $(r,\dot{r})$ space by choosing $g(r)$ so that $g(0) \neq 0$ by corollary 5.3-13.

Various solutions are plotted for choices of a and $g(r)$ ($\alpha \neq 0$) and compared with solutions of (5.3-1). Once again, $c = 1$, without loss of generality and a test solution starts from rest at $r = 5$. Notice (5.7-5) always has a critical point at (0,0) in $(r,\dot{r})$ space and is a centre if the force is central by the above comments. In this case, there is oscillatory motion, as in most previous cases. The curves (1),(2),(3) of figure 5.6 show the cases $a^2 = 1, .01, .001$ respectively. Curve (1) of figure 5.7 shows the complete orbit when $a = 1$. The original Newtonian orbit initially follows curves (2) and (3) in figure 5.6 for $r > 0$. If $\alpha \neq 0$ and $g(r) = \dfrac{r^2}{r^4+a^4}$, the oscillatory motion is irregular - the critical point at (0,0) becomes a focus, and orbits are no longer symmetric with respect to $r = 0$. Curve (4) in figure 5.6 and (3) in figure 5.7 show the case $a^2 = 0.001$ $\alpha = 0.001$.

The two-dimensional model of the form (5.6-1) corresponding to (5.7-5) ($\alpha = 0$) is (5.7-2) with $f(r) = \dfrac{-cr}{(r^2+a^2)^{3/2}}$. Once again, for simplicity, $G(r) = \dfrac{1}{r^2+b^2}$, $|b| \ll 1$, and so the model becomes,

$$\begin{cases} \dot{r} = \sqrt{\dfrac{2c}{(r^2+a^2)^{\frac{1}{2}}} - \dfrac{k^2}{(r^2+b^2)} + A} \\ \dot{\theta} = \dfrac{k}{r^2+b^2} \end{cases} \qquad (5.7\text{-}6)$$

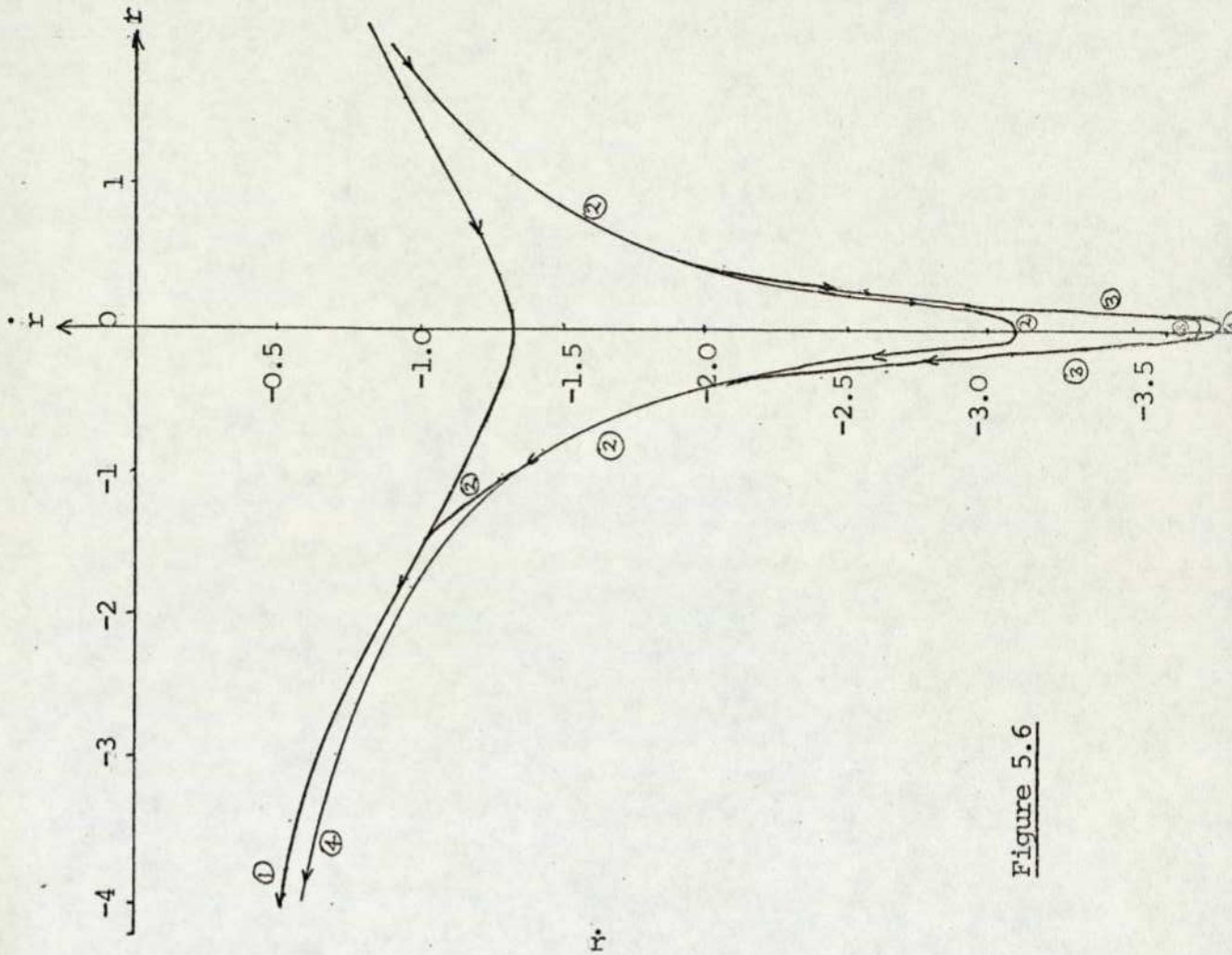


Figure 5.6

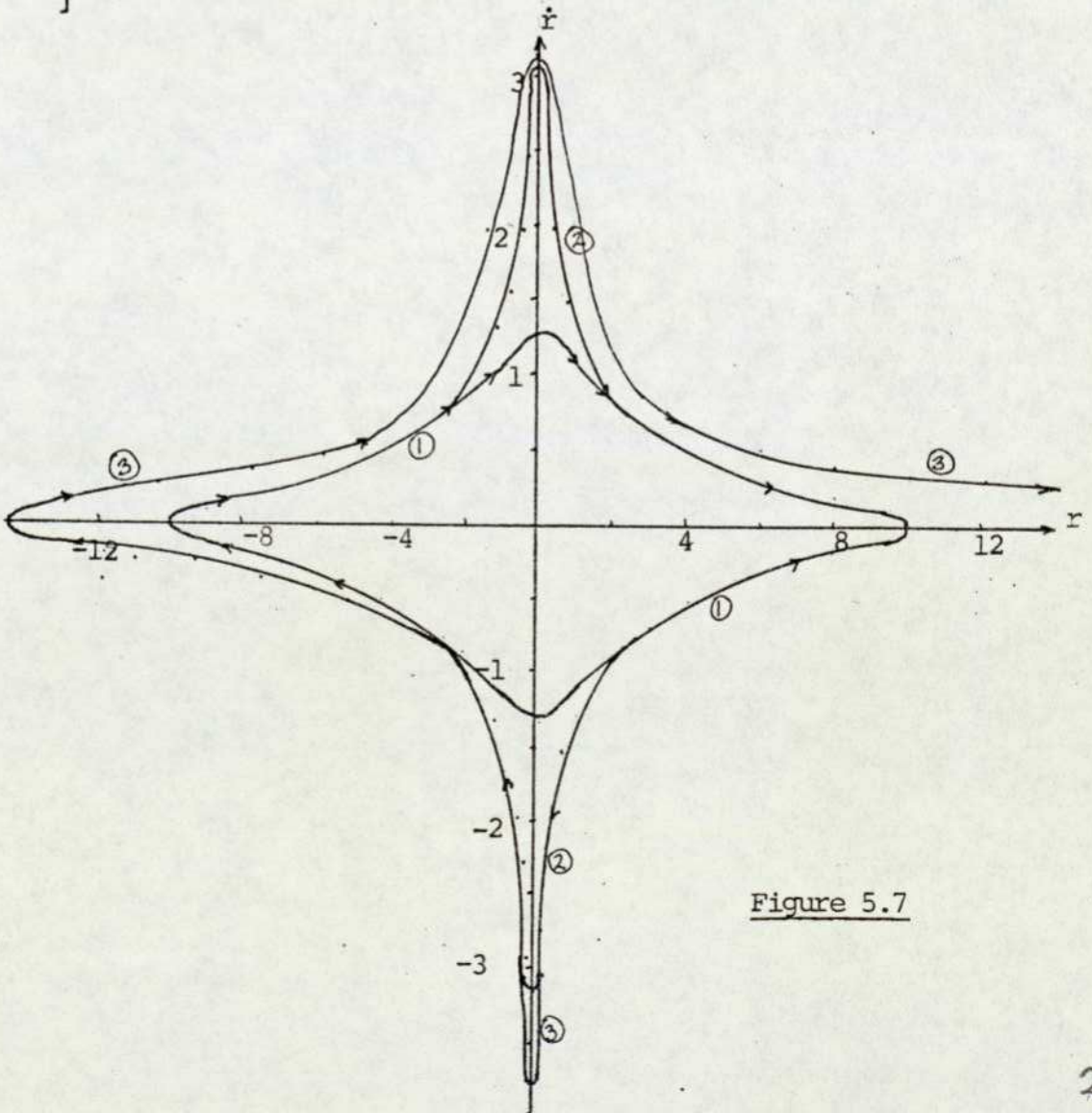


Figure 5.7

The equations have global existence of solutions and reduce to (5.4-3) as $r\to\infty$, $a\to 0$, $b\to 0$. (see examples 5.5-1 and 5.5-2). The character of the only possible critical point in (x,y) space, namely (0,0), is investigated in example 5.6-1. Once again, (5.7-6) is structurally stable for almost all values of A,k.

Solutions of (5.7-6) can be plotted for particular values of A,k, a and b and are compared with the orbits in the Newtonian case. (Once again A = -0.98, k = 1.0 in the Newtonian case and the test solutions start at r = 5, from rest). The solution of (5.7-6) for $a^2 = 0.001$, $b^2 = .001$ is shown as curve (4) in figure 5.4. Notice it is closer to the original Newtonian orbit than the orbit of the previous amendment for the same parameter value. In this sense, it seems a better amendment than the previous one. (Once again, the Newtonian orbit is shown as curve (1) in fig. 5.4).

Example 5.7-4

Finally, consider the set of equations given by,

$$\ddot{r} = \frac{c(2n-1)a^{2n}r^{2n-2} - cr^{4n-2}}{+} + \alpha g(r)\dot{r} \qquad (5.7\text{-}7)$$

These are feasible dynamical amendments to (5.1-3), provided that $a \in \mathbb{R}_{>0}$ and $a<<1$; $\alpha \in \mathbb{R}$, $\alpha<<1$, $n \in \mathbb{Z}^+$ and even; and $\lim_{r\to\pm\infty} \int g(r)dr$ is finite (see example 5.2-7) and theorem 5.2-8). As in the previous example, the equations are quite tractible, having a first integral given by (5.2-17). As far as numerical solution is concerned, n is taken to be 1 or 2 - these are the only two fundamentally different cases. If n = 1, unlike all the previous cases, the critical point in $(r,\dot{r})$ space is not situated at the origin, but arbitrarily close at $r = +\sqrt{2}a$ $(-\sqrt{2}a)$, $\dot{r} = 0$. If n=2, critical points are given by (0,0) and $(\sqrt[4]{3}a,0)$. If n=1, $(\sqrt{2}a,0)$ is a focus, multiple focus, or centre depending on the nature of g(r) (see lemma 5.3-1), and if n = 2, $(\sqrt[4]{3}a,0)$ have this same character and (0,0) is multiple, being a saddle-node, or degenerate (see theorems 5.3-2, 5.3-3). The simple critical points must be centres in both cases if $\alpha=0$, by lemma 5.3-10.

If $n = 1$, the equations may be made structurally stable by choosing $g(r)$ suitably (theorem 5.3-12).

In comparing solutions of (5.7-7) with the original Newtonian orbits, the usual starting values are assumed. Curves (1) and (2) of figure 5.8 show the cases $n=1$, $\alpha=0$ with $a^2 = .1$, $.01$ respectively, while curves (3) and (4) show the cases $n=2$, $\alpha=0$, with $a^2 = .1, .01$ respectively. Curve (5) is the original Newtonian case, and curves (1) and (2) of figure 5.8 are the complete orbits of curves (2) and (4). As the force is central, other possible orbits are reflections of the curves in the figures in the line $r = 0$. These amendments are different from previous cases as here, there is a critical point at a finite positive value of r and the acceleration changes sign there. Thus, the gravitation becomes repulsive and a test particle is thrown back out again.

The two-dimensional model of the form (5.6-1) corresponding to (5.7-7) ($\alpha=0$) is, if $G(r) = \dfrac{1}{r^2+b^2}$, $|b| << 1$, given by,

$$\dot{r} = \left\{ \frac{2cr^{n-1}}{(r^n + a^n)} - \frac{k^2}{r^2+b^2} + A \right\}^{\frac{1}{2}} \qquad (5.7\text{-}8)$$

$$\dot{\theta} = \frac{k}{r^2 + b^2}$$

where $n = 2$ or 4. (see examples 5.6-2). There are not critical circles in (x,y) space, the only possible critical point being $(0,0)$ when $A = \dfrac{k^2}{b^2}$, $\alpha \neq 0$. If $n = 2$, (5.7-8) is not a suitable amendment for these values, but for $n = 4$, $(0,0)$ is a focus. The system is structurally stable for almost all values of A and k.

Comparing with the original Newtonian orbit for $A = -0.98$, $k = 1.0$, the amended orbits have an exactly similar behaviour to previous amendments. Curves (5) and (6) of figure 5.4 show orbits of (5.7-8) with $n = 2$, $a^n = 0.001$, and $b = 0.1$, 0.001 respectively, while curves (7) and (8) show orbits of (5.7-8) with $n = 4$, $a^n = 0.001$, and $b = 0.1$, 0.001 respectively, in a neighbourhood of the origin.

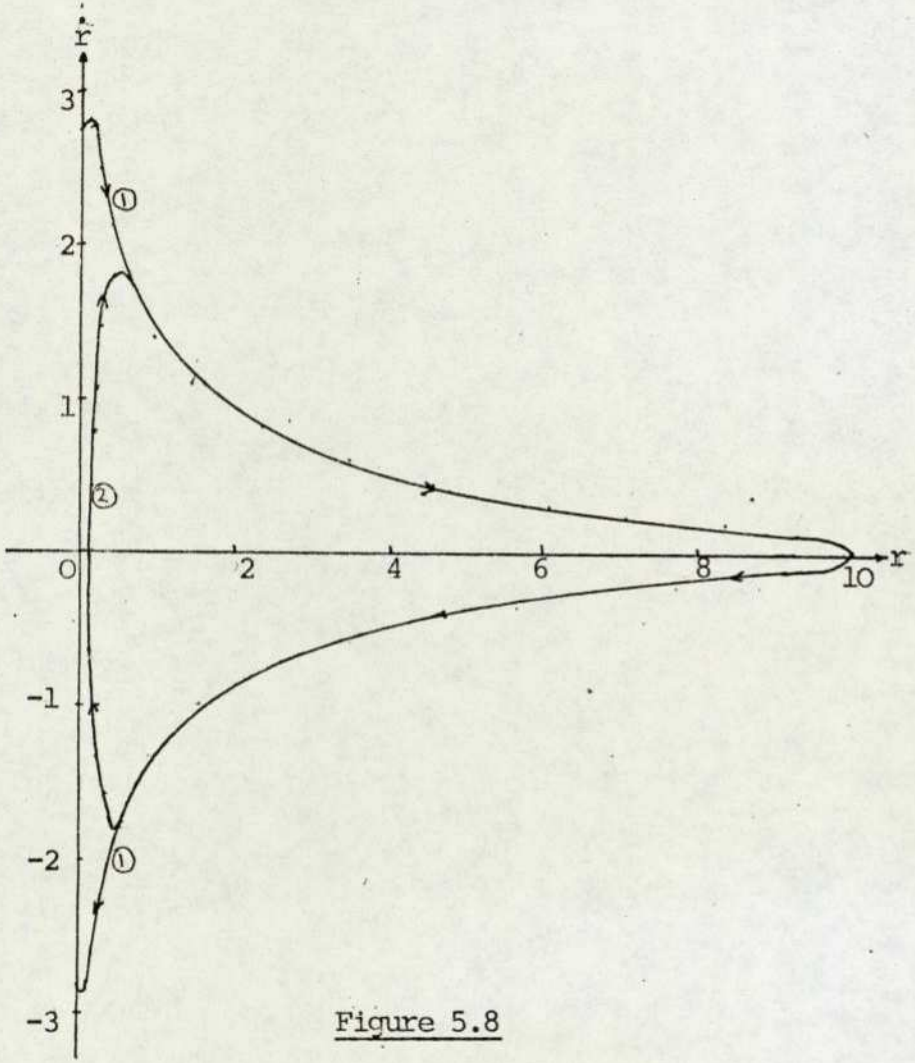


Figure 5.8

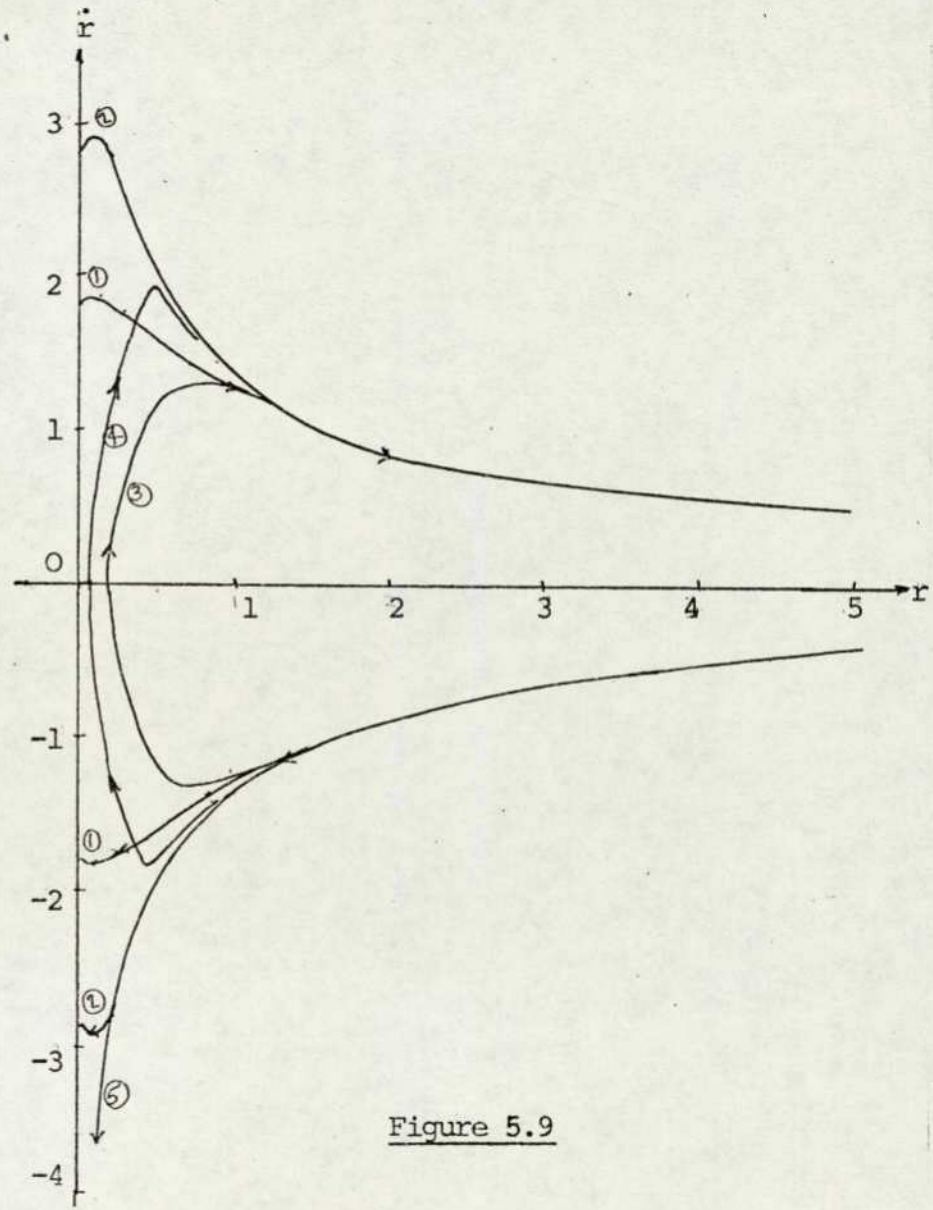


Figure 5.9

5.8) Concluding Remarks

In this chapter, various amendments of the one - and two-dimensional Newtonian equations have been considered. General properties of certain general classes of amendments have been investigated before specific examples have been more closely studied. Of course, not all types of amendment have been considered - the purpose of the chapter was not to classify all possible amendments but rather to suggest several plausible alternatives. A lot of seemingly different alternatives can be seen to produce the same qualitative results.

Having introduced the possibility of Newtonian models, it would be a useful exercise to classify possible amendments according to their qualitative behaviour.

Chapter Six:-

Models of Gravity Approximating to the Schwarzchild Equations

So far, we have been concerned with the non-relativistic Newtonian equations of gravity and with feasible amendments of them. As the Newtonian equations are non-relativistic, the underlying manifold on which the Newtonian C^o vector field is defined is simply Euclidean space $\mathbb{R}^n$ (n=1,2,3). In this chapter, the relativistic Schwarzchild equations and possible amendments are studied. In contrast to the Newtonian case, the Schwarzchild manifold is not flat and needs more than one chart to define it. (Given the Schwarzchild metric in terms of one chart, the equations denied from the metric must contain at least one apparent singularity). Thus, the task of proving any amendment to the equations has global existence of solutions becomes more difficult.

The amendments considered in this chapter reduce to the amended Newtonian equations of the previous chapter in the weak field limit.

6.1) The Schwarzchild Equations of Motion

The Schwarzchild metric is a special case of the general spherically symmetric metric of the form,

$$ds^2 = -A(r)\, dt^2 + B(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \qquad (6.1\text{-}1)$$

The co-ordinate system and metric (6.1-1) have been used in most theoretical models for relativistic stars since the pioneering work of Schwarzchild, Tolman and Oppenheimer and Volkoff.[1]

Having assumed a metric of the form (6.1-1), the Schwarzchild metric, describing the geometry outside any spherically symmetric metric centre of attraction of total mass-energy c, by simply noting that the Ricci tensor associated with (6.1-1) must obey Einstein's equations for free space. The metric is,

$$ds^2 = -\left(1-\frac{2c}{r}\right)dt^2 + \frac{dr^2}{\left(1-\frac{2c}{r}\right)} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \qquad (6.1\text{-}2)$$

The geodesic equations of motion are in general given by (2.2-4). If the $\{x_i\}$ are Schwarzchild co-ordinates $\{t,r,\theta,\phi\}$ and Γ^i_{jk} are the connection coefficients of (6.1-2) in terms of these co-ordinates, (2.2-4) have a first integral and so the equations of motion are expressible as the following first order system:-

$$\begin{cases} \dfrac{dt}{d\lambda} = \dfrac{\pm k_1}{(1-2c/r)} & (6.1\text{-}3)\,(a) \\ \dfrac{dr}{d\lambda} = \sqrt{k_1^2 + (1 - \dfrac{2c}{r})(\varepsilon - \dfrac{k_2}{r^2})} & (6.1\text{-}3)\,(b) \\ \dfrac{d\theta}{d\lambda} = \dfrac{1}{r^2}\sqrt{k_2 - k_3^2 \operatorname{cosec}^2\theta} & (6.1\text{-}3)\,(c) \\ \dfrac{d\phi}{d\lambda} = \dfrac{k_3}{r^2\sin^2\theta} & (6.1\text{-}3)\,(d) \end{cases}$$

($k_i, i = 1,2,3$, are constants of integration). In (6.1-3), ε takes on three values, depending on the nature of the geodesics in question. $\varepsilon = 1$ in the spacelike geodesic case, $\varepsilon = 0$ in the null case and $\varepsilon = -1$ in the timelike case. (In the timelike case, λ can be taken to be proper time - the time as measured by a travelling observer's clock).

Of course, it makes no difference to the motion of a test particle or light ray outside the centre of attraction - a star say, what the geometry is inside, because the particle or ray never gets there (before it can, it collides with the surface of the star). Thus, if $r > 2c$, we need not face the issue of a 'singularity' at $r = 2c$. However, in the limit of gravitational collapse the centre of attraction is a 'black hole'. Then (6.1-2) applies arbitrarily close to $r = 2c$ and the apparent singularity in (6.1-2) leads to an apparent singularity in (6.1-3). However, as shown in chapter two, this apparent singularity can be removed from the metric by choosing a different set of co-ordinates in a neighbourhood of $r = 2c$, such as Kruskal-Szekeres co-ordinates given by (2.4-2). If the geodesic

equations (6.1-3) are expressed in terms of these co-ordinates, the only singularity occurring is at $r = 0$.

As shown in chapter two, in contrast to $r = 2c$, $r = 0$ is a singularity of the space-time manifold itself and so represents a true singularity of the geodesic equations (6.1-3). It is easy to show that this singularity is reached in finite time. (As $r = 2c$ is not a singularity, nothing can prevent a test particle or light ray approaching $r = 2c$ from falling on inwards towards $r = 0$.) For example, if $k_2 = k_3 = 0$, (6.1-3)(b) becomes, in the timelike geodesic case,

$$\frac{dr}{d\lambda} = \sqrt{(K_1^2-1)+\frac{2c}{r}} \Rightarrow \int \frac{dr}{\sqrt{(K_1^2-1)+2c/r}} = \lambda + \lambda_o \qquad (6.1\text{-}4)$$

and so if the root in the integral in (6.1-4) is real in a neighbourhood of $r = 0$, $r = 0$ is reached in a finite value of λ. Also, from (6.1-3)(a) and (b),

$$\int \frac{K\,dr}{(1-2c/r)\sqrt{K_1^2-1)+2c/r}} = t + t_o \qquad (6.1\text{-}5)$$

and so if $r = 0$ is reached in finite λ, it is reached in finite co-ordinate time (if the motion is observed inside $r = 2c$).

This chapter is concerned with the Schwarzchild equations in two space dimensions, so that in (6.1-3), $K_2 = K_3$ and $\theta = \frac{\pi}{2}$ say. The behaviour of orbits in (r,ϕ) space can be determined from,

$$\frac{dr}{d\phi} = \frac{r^2}{K_3}\sqrt{K_1^2 + (1-\frac{2c}{r})(\varepsilon - \frac{K^2}{r^2})} \qquad (6.1\text{-}6)$$

using (6.1-3)(b) with $\theta = \frac{\pi}{2}$. The integration of (6.1-6) is not easy, even in the case $K_1 = 0$, but qualitative behaviour of the orbits can be ascertained by calculating turning points of an 'effective potential'.[2] In the case of a test particle this is given by,

$$V(r) = \sqrt{(1 - \frac{2c}{r})\left\{\frac{K_3^2}{r^2} + 1\right\}} \qquad (6.1\text{-}7)$$

Theorem 6.1-1[2]

Let V_{min}, V_{max} denote the minimum and maximum values of V(r) respectively. (v(r) given by (6.1-7).) Then,

(i) $r = r(\phi)$ is a circular orbit when $K_1 = V_{min}$

(ii) there is precession when K_1 is a little more than V_{min}

(iii) $r = r(\phi)$ is an unstable, almost periodic orbit when $K \doteq V_{max}$

(iv) $r = r(\phi)$ reaches the singularity at $r = 0$ when $K > V_{max}$

□

Notice that the turning points of the potential V(r) are given by, $r = \dfrac{K_3^2 \pm K_3\sqrt{K_3^2 - 12c^2}}{2c}$ and so, in certain cases, there may be no turning points and the fourth possibility of the theorem must hold.

In the case of light rays can again by deduced from a potential. The conclusions are contained in the following theorem,

Theorem 6.1-2[2]

Let $b = \dfrac{K_3}{K_1}$. Then,

(i) if $b > 3\sqrt{3}c$, a photon which falls in from infinity is reflected and returns to infinity:- if $b >> 3\sqrt{3}c$ the orbit is almost a straight line and if $0 < b - 3\sqrt{3} << c$, the photon circles the star many times before flying back to infinity. A photon emitted from $r = 2c$ gets pulled back there.

(ii) if $b < 3\sqrt{3}c$, a photon which falls in from infinity, falls into $r = 2c$ and a photon emitted near $r = 2c$ escapes to infinity.

□

6.2) Feasible Amendments of the Schwarzchild Equations

In order for the equations (6.1-3) to have global existence of solutions, the singularity at $r = 0$ must be removed. In this section, amendments of (6.1-3) possessing global existence of solutions are considered, which are derivable from a metric. The amendments considered reduce to the amended Newtonian equations in the weak field limit.

6.2.1) A More General Set of Geodesic Equations

The purpose of this sub-section is to generalize the first order system (6.1-3) in such a way that the resulting equations are still derivable from a metric. To this end, consider the special case of the metric (6.1-1),

$$ds^2 = - A(r)dt^2 + \frac{1}{A(r)}\, dr^2 + F(r)\,(d\theta^2 + \sin^2\theta d\phi^2) \qquad (6.2\text{-}1)$$

(As only the two space dimensions case is considered, the singularities at $\theta = n\pi$, $n \in \mathbb{Z}$, are no problem. In three dimensions these are only apparent singularities anyway).

Connection coefficients associated with (6.2-1) are given by (2.2-5), where $g_{oo} = -g'' = -A(r)$, $g_{22} = F(r)$, $g_{33} = F(r)\sin^2\theta$, all other components of the metric tensor being zero. If the components of the Ricci tensor associated with (6.2-1) are all zero, the functions A,F will be as in the Schwarzchild case. However, if it is assumed that the Ricci tensor is only approximately zero far away from $r = 0$, that is, the field equations (2.3-1) do not hold in a neighbourhood of $r = 0$, the equations of motion associated with (6.2-1) for general functions $A(r)$, $F(r)$ can be found. From (2.2-4), the second order equations of motion are,

$$\begin{cases} \ddot{t} + (A'/A)\dot{r}\dot{t} = 0 & (6.2\text{-}2)\,(a) \\ \ddot{\phi} + (F'/F)\dot{r}\dot{\phi} + 2\cot\theta\dot{\theta}\dot{\phi} = 0 & (6.2\text{-}2)\,(b) \\ \ddot{r} + \frac{1}{2}AA'\,(\dot{t})^2 - \frac{1}{2}\frac{A'}{A}(\dot{r})^2 - \frac{1}{2}AF'(\dot{\theta})^2 - \frac{1}{2}AF'\sin^2\theta\,(\dot{\phi})^2 = 0 & (6.2\text{-}2)\,(c) \\ \ddot{\theta} + (F'/F)\dot{r}\dot{\theta} - \sin\theta\cos\theta\,(\dot{\phi})^2 = 0 & (6.2\text{-}2)\,(d) \end{cases}$$

where a dot denotes differentiation with respect to the affine parameter λ and a dash denotes differentiation with respect to r.

The equations (6.22) have a first integral:-

$$\ddot{t} + (A'/A)\dot{r}\dot{t} = \frac{1}{A}\frac{d}{d\lambda}(A\dot{t}) = 0 \Rightarrow \dot{t} = \frac{\pm K_1}{A(r)} \qquad (6.2\text{-}3)(a)$$

(K_1 a constant of integration).

$$\ddot{\phi} + (F'/F)\dot{r}\dot{\phi} + 2\cos\theta\dot{\theta}\dot{\phi} = \frac{1}{F\sin^2\theta}\frac{d}{d\lambda}(F\sin^2\theta\dot{\phi}) = 0 \Rightarrow$$

$$\dot{\phi} = \frac{K_3}{F(r)\sin^2\theta} \qquad (6.2\text{-}3)(d)$$

(K_3 a second constant of integration).

The general metric (6.2-1) gives,

$$\left(\frac{ds}{d\lambda}\right)^2 = \varepsilon = -A\left(\frac{dt}{d\lambda}\right)^2 + \frac{1}{A}\left(\frac{dr}{d\lambda}\right)^2 + F(r)\left\{\left(\frac{d\theta}{d\lambda}\right)^2 + \sin^2\theta\left(\frac{d\phi}{d\lambda}\right)^2\right\} \qquad (6.2\text{-}4)$$

so that $\varepsilon = 1, 0, -1$ in the spacelike, null, timelike geodesic cases respectively. Substituting (6.2-4) into (6.2-2)(c) and using (6.2-3)(a), we obtain,

$$\frac{d^2r}{d\lambda^2} + \frac{1}{2}\frac{K_1^2A'}{A} - \frac{1}{2}\frac{A'}{A}\left(\frac{dr}{d\lambda}\right)^2 - \frac{1}{2}AF'\left\{\frac{\varepsilon}{F} + \frac{K_1^2}{AF} - \frac{1}{AF}\left(\frac{dr}{d\lambda}\right)^2\right\} = 0 \Rightarrow$$

$$\frac{du}{dr} + \left\{\frac{F'}{F} - \frac{A'}{A}\right\}u + \left\{\frac{K_1^2A'}{A} - \varepsilon\frac{AF'}{F} - \frac{K_1^2F'}{F}\right\} = 0 \qquad (6.2\text{-}5)$$

where $u = \left(\frac{dr}{d\lambda}\right)^2$. This is a first order, linear equation with integrating factor $\frac{F}{A}$. The solution is,

$$u = \left(\frac{dr}{d\lambda}\right)^2 = K_1^2 + A\left(\varepsilon - \frac{K_2}{F}\right) \qquad (6.2.\text{-}3)(b)$$

where K_2 is a further constant of integration. Substituting for $\dot{t}, \dot{\phi}, \dot{r}$ in (6.2-4),

$$\dot{\theta} = \frac{1}{F}\sqrt{K_2 - K_3^2\,\mathrm{cosec}^2\theta} \qquad (6.2\text{-}3)(c)$$

In conclusion, the following lemma may be stated:-

Lemma 6.2-1

The geodesic equations of motion derived from the general spherically symmetric metric (6.2-1) are,

$$\begin{cases} \dfrac{dt}{d\lambda} = \dfrac{\pm K_1}{A(r)} & (6.2\text{-}3)\,(a) \\ \dfrac{dr}{d\lambda} = \sqrt{K_1^2 + A\left(\varepsilon - \dfrac{K_2}{F}\right)} & (6.2\text{-}3)\,(b) \\ \dfrac{d\theta}{d\lambda} = \dfrac{1}{F}\sqrt{K_2 - \dfrac{K_3^2}{\sin^2\theta}} & (6.2\text{-}3)\,(c) \\ \dfrac{d\phi}{d\lambda} = \dfrac{K_3}{F\sin^2\theta} & (6.2\text{-}3)\,(d) \end{cases}$$

K_i, $i = 1,2,3$ constants, and $\varepsilon = -1$, 0, or $+1$.

□

Obviously, the equations (6.2-3) contain the Schwarzchild equations as a special case, where $A(r) = 1 - \frac{2c}{r}$ and $F(r) = r^2$. In this work, only two space dimensions are considered and with $\theta = \pi/2$, $K_2 = K_3^2$, (6.2-3) reduce to,

$$\begin{cases} \dfrac{dt}{d\lambda} = \dfrac{\pm K_1}{A(r)} & (6.2\text{-}4)\,(a) \\ \dfrac{dr}{d\lambda} = \sqrt{K_1^2 + A(r)\left\{\varepsilon - \dfrac{K_3^2}{F(r)}\right\}} & (6.2\text{-}4)\,(b) \\ \dfrac{d\phi}{d\lambda} = \dfrac{K_3}{F(r)} & (6.2\text{-}4)\,(c) \end{cases}$$

Notice that the equations for the space co-ordinates r and ϕ are independent of the co-ordinate time t and that they have the same general form as (5.5-2) so that theorem 5.5-1 is applicable. Thus, r and ϕ can be extended to any value of λ, provided that $K_1^2 + A\left(\varepsilon - \frac{K_3^2}{F}\right) < 0$ in an arbitrarily small neighbourhood of all finite singular points of $K_1^2 + A\left(\varepsilon - \frac{K_3^2}{F}\right)$ and $F(r)$, if each

of these functions are continuous in r and $(\dot{r})'$ is finite at each critical point of (6.2-4)(b),(c) and at infinity if $\dot{r}(r)$ becomes singular there.

6.6.2) Feasible Amendments to the Schwarzchild Geodesic Equations of the Form (6.2-4)

Looking at the apparent force on a test particle instantaneously at rest and assuming that the gravitational field is weak, the first component of the metric (6.1-2) is given by,

$$-g_{oo} = 2\Phi + 1 = A(r) \qquad (6.2\text{-}5)$$

where Φ is the Newtonian potential. Thus the singular character of the Newtonian potential, inevitably leads to a singular character of the Schwarzchild equations.

It is seen that, from (6.2-5), in order for (6.2-4) to reduce to the amended two-dimensional Newtonian equations (5.6-1) in the weak field limit, $A(r) = 1 + 2\Phi^*$, where Φ^* is the amended Newtonian potential, so that,

$$A(r) = 1 - 2\int f(r)dr \qquad (6.2\text{-}6)$$

where $f(r) \to -\frac{c}{r^2}$ as $r \to \infty$.

Comparing (6.2-4) with this choice of A(r) with (5.6-1), F(r) must be chosen so that $F(r) \to r^2$ as $r \to \infty$. Notice that if $K_3 = 0$, (6.2-4) reduces exactly to the one-dimensional Newtonian equation (5.2-1) in this case, (along timelike geodesics).

Notice that if $A(a) = 0 \Rightarrow \int f(r)dr\Big|_{r=a} = \frac{1}{2}$ from (6.2-6),

r = a corresponds to an amended Schwarzchild radius, that is, the right hand side of (6.2-4)(a) becomes infinite there. In contrast to the Schwarzchild case, in any particular amendment, there could be several amedned Schwarzchild radii or none at all. As in the original case, values of r satisfying A(r) = 0 are apparent singularities and can be removed from the general metric (6.2-1) by considering a transformation to generalized Kruskal-Szekeres

co-ordinates, as shown in a later sub-section.

As the function A(r) is determined by an integral in (6.2-6), the arbitrary constants K_1, K_3 in (6.2-4) may differ considerably from the arbitrary constants in (6.1-3) with $\theta = \frac{\pi}{2}$, for any given initial conditions.

At this stage, a couple of examples of amended models are considered.

Example 6.2-1

Consider the system of equations,

$$\begin{cases} \dfrac{dt}{d\lambda} = \dfrac{\pm K_1}{1 - (2c/a)\tan^{-1}(r/a)} \\ \dfrac{dr}{d\lambda} = \sqrt{K_1^2 + \left(\dfrac{2c}{a}\tan^{-1}\dfrac{r}{a} - 1\right)\left\{1 + \dfrac{K_3^2}{r^2 + a^2}\right\}} \\ \dfrac{d\phi}{d\lambda} = \dfrac{K_3}{r^2 + a^2} \end{cases} \qquad (6.2\text{-}7)$$

$\frac{dr}{d\lambda}$, $\frac{d\phi}{d\lambda}$ are non-singular in r and are, in fact, continuous. Notice that there are no finite critical points. Suppose that $\dot{r}(r=0) = 0$, so that $K_1^2 = 1 + \dfrac{K_3^2}{a^2}$. Then,

$$\frac{d}{dr}\left(\frac{dr}{d\lambda}\right)(r=0) = \left\{\lim_{r\to 0} \frac{1}{r^2}\left(K_1^2 + \left(\frac{2c}{a}\tan^{-1}\frac{r}{a} - 1\right)\left(1 + \frac{K_3^2}{r^2+a^2}\right)\right)\right\}^{\frac{1}{2}}$$

and so does not exist. By theorem 5.5-1, (6.2-7) is a viable amendment of (6.1-3) provided that $K_1^2 \neq 1 + \dfrac{K_3^2}{a^2}$ in the timelike case with $\theta = \pi/2$. If $K_3 = 0$, (6.2-7) reduces to the one-dimensional Newtonian amendment, (5.2-2) given in example 5.2-1. Notice that K_1 in (6.2-7) differs considerably from K_1 in (6.1-3) for any given

initial conditions. For example, suppose that in (6.1-3) $K_1 = 0$, $K_3 = 1$ ($\theta = \pi/2$), so that in the timelike goedesic case, $\dot{\phi} = 1$, $\dot{r} = (4c-2)^{\frac{1}{2}}$ when $r = 1$. With these initial conditions, K_1 and K_3 in (6.2-7) are given by,

$$K_1 = \left\{4c + a^2 - \tan^{-1}\left(\frac{1}{a}\right)\left(\frac{4c}{a} + 2ca\right)\right\}^{\frac{1}{2}}, \quad K_3 = 1 + a^2$$

so that for $|a| << 1$, $K_3 \doteqdot 1$, but K_1 will be complex!

The amended Schwarzchild radii of (6.2-7) are given by, $\tan^{-1}\frac{r}{a} = \frac{a}{2c}$ and so in this case, $r = a\tan\left(\frac{a}{2c}\right)$, (the principle branch of $\tan^{-1}$ being taken), so that for $|a| << 1$, the amended Schwarzchild radius approaches the origin!

Example 6.2-2

Consider the system,

$$\begin{cases} \dfrac{dt}{d\lambda} = \dfrac{\pm K_1}{\left(1 - 2c\,(r^2 + a^2)^{-\frac{1}{2}}\right)} \\ \dfrac{dr}{d\lambda} = \sqrt{K_1^2 + \left(-1 + \dfrac{2c}{(r^2+a^2)^{\frac{1}{2}}}\right)\left(1 + \dfrac{K_3^2}{r^2 + \alpha^2}\right)} \\ \dfrac{d\phi}{d\lambda} = \dfrac{K_3}{r^2 + \alpha^2} \end{cases} \qquad (6.2\text{-}8)$$

($a \neq 0$).

With $\alpha = 0$, $\dot{r}$ is real in a neighbourhood of $r = 0$ for most values of a and c ($2c > a$) in any reasonable amendment and so by theorem 5.5-1, (6.2-8) is not a good amendment of (6.1-3) in this case. If $\alpha \neq 0$, $\dot{r}$ and $\dot{\phi}$ are continuous in r and there are no singularities at infinity and no finite critical points. Suppose that $\dot{r}(r = 0) = 0$, so that $K_1^2 + \left(\frac{2c}{a} - 1\right)\left(1 + \frac{K_3^2}{\alpha^2}\right) = 0$.

$$\frac{d}{dr}(\dot{r})_{(r=0)} = \left\{\lim_{r\to 0} \frac{1}{r^2}\left(K_1^2 + \left(\frac{2c}{(r^2+a^2)^{\frac{1}{2}}} - 1\right)\left(1 + \frac{K_3^2}{r^2 + \alpha^2}\right)\right)\right\}^{\frac{1}{2}}$$

$$= \left\{\left(1 + \frac{K_3^2}{\alpha^2}\right)\left(-\frac{c}{a^3}\right) + \left(\frac{2c}{a} - 1\right)\left(-\frac{K_3^2}{\alpha^4}\right)\right\}^{\frac{1}{2}}$$

and so is finite. (Notice that in this example $F'(0) = 0$, $f(0) = 0$.). Thus, by theorem 5.5-1, (6.2-8) is a viable amendment of (6.1-3), for any values of K_1, K_3 in this case. This time, $\lim_{r\to\infty} A(r) \to 1$ as in the Schwarzchild case and so for any given initial conditions, if $|\alpha|$, $|a| \ll 1$, the arbitrary constants in (6.2-8) and (6.1-3) can be made arbitrarily close, unlike in the previous example. In the weak field region, (6.2-8) approaches the amended two-dimensional Newtonian model (5.5-12) and if $K_3 = 0$, it reduces exactly to the one-dimensional amended Newtonian model (5.2-15).

The amended Schwarzchild radii are given by,

$$r^2 + a^2 = 4c^2 \Rightarrow r = \pm\sqrt{4c^2 - a^2}\ .$$

Thus, in (x,y) space, there is one Schwarzchild radius if $4c^2 > a^2$, as in the Schwarzchild case, and no apparent singularity if $4c^2 < a^2$.

As mentioned previously, if $A(a) = 0$, $r = a$ is an apparent singularity of the equations (6.2-4) as far as the travelling observer is concerned and can be reached in a finite value of λ. Thus, although (6.2-4)(b) and (c) may have global existence of solutions and the timelike geodesic is complete, (6.2-4) as a whole has a singularity which is reached in a finite value of the parameter. Therefore, although the geodesic is complete with respect to the travelling observer, it is not apparently complete on the space-time manifold with respect to an external observer, due to the fact that the continuous vector field defining the geodesic has to be expressed in terms of more than one chart. Of course, this problem does not arise if the Schwarzchild equations are amended in such a way so that no amended Schwarzchild radii are present. In this case, the space-time manifold is expressible in terms of one chart.

6.3 Properties of the Amended Schwarzchild Equations

In this section, various properties of the system (6.2-4), are investigated where $K_1 \neq 0$, $\varepsilon = 0$ or -1 and it is also assumed that, for reasons given in sub-section 6.22), $F(r) \to r^2$ as $r\to\infty$ and $A(r) = 1 - 2\int f(r)dr$ where $f(r) \to -\frac{c}{r^2}$ as $r \to \infty$. It is also assumed that the right-hand sides of (6.2-4)(b) and (c) satisfy the conditions of theorem 5.5-1.

Notice that from (6.2-4)(a) and (b) we get,

$$\frac{dr}{dt} = \frac{A}{K_1}\sqrt{K_1^2 + A\left\{\varepsilon - \frac{K_3^2}{F}\right\}} \Rightarrow \int \frac{K_1 dr}{A\sqrt{K_1^2 + A(\varepsilon - K_3^2/F)}} = t + t_o \qquad (6.3\text{-}1)(a)$$

and from (6.2-4)(a) and (c),

$$\frac{d\phi}{dt} = \frac{K_3}{K_1}\,\frac{A}{F} \qquad (6.3\text{-}1)(b)$$

In order to get some idea of the qualitatative behaviour of orbits of (6.2-4) in space, that is, the orbits of test particles which will actually be seen, it is sufficient to consider only those equations which involve the space co-ordinates. The shape of the orbits as far as an observer travelling along the geodesics is concerned is given by the phase portrait of the system (6.2-4)(b) and (c) with λ as the affine parameter. The shape of the orbits as far as the external observer is concerned is given by the phase portrait of the system (6.3-1).

These two cases can be combined into the general system,

$$\begin{cases} \dfrac{dr}{d\lambda} = P(r) \\ \dfrac{d\phi}{d\lambda} = Q(r) \end{cases} \qquad (6.3\text{-}2)$$

where, for a travelling observer, P and Q are as in (6.2-4) and λ is affine, and for an external observer P and Q are given by (6.3-1)(a), (6.3-1)(b) respectively and λ is co-ordinate time.

Let $x = r\cos\phi$, $y = r\sin\phi$ so that (6.3-2) becomes,

$$\begin{cases} \dfrac{dx}{d\lambda} = \dfrac{x}{r}P(r) - yQ(r) = R(x,y) \\ \dfrac{dy}{d\lambda} = \dfrac{y}{r}P(r) + xQ(r) = S(x,y) \end{cases} \qquad (6.3\text{-}3)$$

Equations (6.3-2) and (6.3-3) are the same as (5.5-1) and (5.6-2) in the two-dimensional Newtonian case and so the analysis of chapter

five applies equally well here.

6.3.1) Critical Points

By the analysis of section 5.6-1), (6.3-3) has a critical point at (0,0) if $P(0) = 0$ and the right hand side also become zero at $r = a$ if $P(a) = Q(a) = 0$.

Suppose that $P(a) = Q(a) = 0$, so that, as before, $r = a$ represents a circle of non-isolated critical points. In the external observer case, this can only happen if $A(a) = 0$, using (6.3-1), and so $r = a$ corresponds to an amended Schwarzchild radius, as expected. In the travelling observer case, (6.2-4) (b) and (c), $r = a$ has the same effect as the Schwarzchild radius with respect to both the external and travelling observer. Thus, if every amended Schwarzchild radius is a critical point of (6.2-4) (b) and (c), a collapsing star could not collapse beyond the largest such radius.

Suppose now that (6.3-3) has a critical point at (0,0), then also by the analysis of section 5.6-1) it follows, that the eigenvalues of the matrix of the linearized system about (0,0) are given by,

$$\lambda = P'(0) \pm iQ(0) \qquad (6.3\text{-}4)$$

In the travelling observer case,

$$P(r) = \sqrt{K_1^2 + A\left(\varepsilon - \frac{K_3^2}{F}\right)}$$

$$\therefore \quad P'(0) = \sqrt{\lim_{h\to 0} \frac{1}{h^2}\left\{K_1^2 + A\left(\varepsilon - \frac{K_3^2}{F}\right)\right\}} \quad \Rightarrow$$

$$P'(0) = \sqrt{\lim_{h\to 0} \frac{1}{2h}\left\{A'\left(\varepsilon - \frac{K_3^2}{F}\right) + K_3^2\,\frac{AF'}{F^2}\right\}}$$

(where a dash on the right-hand side represents differentiation with respect to h).

Now $A(r) = 1 - 2\int f(r)dr \Rightarrow A'(r) = -2f(r)$ and so $f(0) = 0 \Rightarrow A'(0) = 0$. Then in order for $P'(0)$ to exist, $F'(0) = 0$. If $f(0) \neq 0$, then

$$\left\{A'(0)\left(\varepsilon - \frac{K_3^2}{F(0)}\right) + \frac{K_3^2 A(0)F'(0)}{F^2(0)}\right\} = 0 \text{ in order for } P'(0) \text{ to}$$

exist. Then $P'(0)$ is given by, using (6.3-5),

$$P'(0) = \left\{\frac{1}{2}\left[A''(0)\left(\varepsilon - \frac{K_3^2}{F(0)}\right) + \frac{2K_3^2A'(0)F'(0)}{F^2(0)} + \frac{K_3^2A(0)}{F^3(0)} \times \right.\right. \tag{6.3-6}$$

$$\left.\left.(F(0)F''(0) - 2(F'(0))^2)\right]\right\}^{\frac{1}{2}}$$

In the external observer case, using (6.3-1), $P'(0) = A'(0)$ if $A(0) = 0$, that is, if $r = 0$ is an amended Schwarzchild radius, or $P'(0)$ is given by the right-hand side of (6.3-6), multiplied by a factor of $A(0)$.

As far as the nature of the critical point (0,0) of (6.3-3) goes, theorem 5.6-1 is appropriate (with F replaced by P, and G replaced by Q).

The analysis of section 5.6.1) shows that (0,0) cannot be a centre unless $P(r) \equiv 0$, so that the possibilities for (0,0) are a focus, a node, a multiple critical point or a completely isolated critical point. Thus, the nature of critical points gives limited information. Notice, also that lemma 5.6-2 also applies to the general system (6.3-3).

Example 6.3-1

Consider the system (6.3-3) with $P(r)$, $Q(r)$ given by,

$$\begin{cases} P(r) = \left\{K_1^2 + \left(\dfrac{2c}{(r^2 + a^2)^{\frac{1}{2}}} - 1\right)\left(1 + \dfrac{K_3^2}{r^2 + a^2}\right)\right\}^{\frac{1}{2}} \\ \\ Q(r) = \dfrac{K_3}{r^2 + a^2} \end{cases} \tag{6.3-7}$$

so that (6.3-2) are the second two equations of the viable model (6.2-7), with $\alpha = a$.

Notice that $Q(r) \neq 0$ for any finite values of r, so that (6.3-3) has no critical circles in this case.

$$P(0) = 0 \Rightarrow K_1^2 = (1 - \frac{2c}{a})\left(1 + \frac{K_3^2}{a^2}\right)$$

and, in this case, using the analysis of example 6.2-2,

$$P'(0) = \left\{ -\left(1 + \frac{K_3^2}{a^2}\right)\frac{c}{a^3} + \left(1 - \frac{2c}{a}\right)\left(\frac{K_3^2}{a^4}\right)\right\}^{\frac{1}{2}} \qquad (6.3\text{-}8)$$

Suppose that $a > 0$. Then $P'(0) = 0$ if $aK_3^2 - 3cK_3^2 - ca^2 = 0$ and $P'(0) \neq 0$ if $aK_3^2 - 3cK_3^2 - ca^2 > 0$. Otherwise, $P'(0)$ is imaginary. Thus, by theorem 5.6-1 and the later analysis of subsection 5.6.1), (0,0) is a focus of (6.3-3) if $aK_3^2 - 3cK_3^2 - ca^2 \geq 0$, in the travelling observer case.

In the weak field region, (6.3-2) with choices of P and Q given by (6.3-7) reduces to the Newtonian model (5.6-10) given in example 5.6-1, with $\alpha = a$ and K_3 corresponds to k, so that if (0,0) is a focus for the amended Schwarzchild model case, it is a focus for the amended Newtonian model, but not necessarily vice versa. Comparing (6.3-7) with (6.2-4), it can be seen that $A(0) = \frac{1-2c}{a}$ and so by previous comments if $a \neq 2c$, the origin has the same character in both the travelling observer and external observer (inside the amended Schwarzchild radius cases).

Example 6.3-2

Consider the system,

$$\begin{cases} \dot{t} = \dfrac{K_1}{(1-2cr^3(r^4+a^4)^{-1})} \\ \dot{r} = \left\{ K_1^2 + \left(-1 + \dfrac{2cr^3}{r^4+a^3}\right)\left(1 + \dfrac{K_3^2}{r^2+a^2}\right)\right\}^{\frac{1}{2}} \\ \dot{\phi} = \dfrac{K_3}{r^2+a^2} \cdot \end{cases} \qquad (6.3\text{-}9)$$

Comparing (6.3-9) with (6.2-4), $A(r) = 1 - \frac{2cr^3}{(r^4+a^4)}$,

$F(r) = r^2 + a^2$. There are no finite critical points and no real singularities. From (6.2-6), $f(0) = 0$, so that $P'(0)$ exists using (6.3-5) and is given by,

$$P'(0) = \left\{\frac{1}{2} \cdot \left(\frac{2K_3^2}{a^4}\right)\right\}^{\frac{1}{2}} = \pm \frac{K_3}{a^2} \qquad (6.3\text{-}10)$$

The equations (6.3-9) are a viable amendment to (6.1-3) by theorem 5.5-1 and reduce to the Newtonian case (5.6-11) with $\alpha = a$, $n = 3$, in the weak field region and reduce to the model (5.2-17) if $K_3 = 0$. From subsection 5.6.1), the origin of (6.3-3) in this case is a focus for any choice of constants if $K_1^2 = 1 + \frac{K_3^2}{a^2}$ as in the case of (5.6-11) with $n = 3$, $\alpha = a$.

The amended Schwarzchild radii are given by the four roots of

$$r^4 - 2cr^3 + a^4 = 0 \quad (a \neq 0)$$

which only has one turning point (a minimum) at $r = \frac{3c}{2}$ and so there are two real roots if $16a^4 - 27c^4 < 0$, one real root if $16a^4 - 27c^4 = 0$ and no real roots if $16a^4 - 27c^4 > 0$. If $|a| \ll 1$, the two real roots are $r \approx 2c$ and $r \approx 0$.

6.3.2) Existence of Periodic and Almost Periodic Orbits

In contrast to the two-dimensional Newtonian equations, the Schwarzchild equations (6.1-3) do not possess a set of non-isolated periodic orbits in (r,ϕ) space for a range of values of K_1, K_3. If the amended equations (6.3-3) with $P(r)$, $Q(r)$ given in (6.2-4) (b) and (c) contain no critical points, (6.3-3) can only have a finite number of isolated closed orbits, which may or may not be limit-cycles. Some information regarding the non-existence of periodic orbits may be obtained from Dulac's test. Applying theorem 3.5-7 to (6.3-3), with $B = rP(r)$ as in the proof of lemma 5.6-4, the following lemma is obtained:-

Lemma 6.3-1

Let G be a simply connected subregion of $\mathbb{R}^2 - \{0\}$. Then there are no closed curves of (6.3-3) in G if the function

$$\Phi = P^2(r) + rP(r)P'(r)$$

does not change sign in G. □

Notice that G cannot contain points at which P'(r) does not exist. In the case of (6.2-4) (b) and (c) or (6.3-1), the lemma is difficult to apply in practice due to the complicated nature of the function Φ, and is of limited use.

Consider again the general system (6.3-2), with P(r) and Q(r) given either by (6.3-1) or (6.2-4)(b),(c). As in the Newtonian case, if P(a) = 0, r = a is an isolated closed orbit of (6.3-2) if a is real. Suppose that $Q(a) \neq 0$, so that r = a is not a critical circle. As in the Newtonian case, if P'(a) does not exist, (6.3-2) does not have uniqueness of solutions on r = a and so any other orbit may cross r = a or reach r = a in finite time, so that a necessary condition for r = a to be a limit-cycle is that P'(a) exists. Then lemma 5.6-5 is appropriate (with F(r) replaced by P(r)). Notice as (6.3-1) is just (6.2-4)(b),(c) multiplied by a function of r, if r = a is a limit-cycle as far as the travelling observer is concerned, it is also one as far as the external observer is concerned (assuming that if r = a lies within an amended Schwarzchild radius, the external observer is also within the Schwarzchild radius).

Example 6.3-3

Consider the system,

$$\begin{cases} \dot{t} = \dfrac{K_1}{\left(1 - 2c(r^2 + a^2)^{-\frac{1}{2}}\right)} \\ \dot{r} = \left\{ K_1^2 + \left(\dfrac{2c}{(r^2+a^2)^{\frac{1}{2}}} - 1 \right) \left(1 + \dfrac{K_3(K_3 + \gamma(r^2+a^2))}{(r^2 + a^2)} \right) \right\}^{\frac{1}{2}} \\ \dot{\theta} = \dfrac{K_3}{r^2+a^2} + \gamma \end{cases} \qquad (6.3\text{-}11)$$

so that, comparing with (6.2-4), $A(r) = 1 - \frac{2c}{(r^2+a^2)^{\frac{1}{2}}}$,

$$F(r) = \left(\frac{\gamma(r^2+a^2) + K_3}{K_3(r^2+a^2)} \right)^{-1} \quad (\varepsilon = -1).$$

Now $P(b) = 0 \Rightarrow K_1^2 + \left(\frac{2c}{(b^2+a^2)^{\frac{1}{2}}} - 1 \right) \left(1 + \frac{K_3(K_3+\gamma(b^2+a^2))}{(b^2+a^2)} \right) = 0.$

Notice that $A'(0) = 0$ and $F'(0) = 0$, so that from (6.3-5), $P'(0)$ exists. Suppose that $\gamma \neq \frac{-K_3}{b^2+a^2}$, so that $r = b$ is not a critical circle. Then (6.3-11) is a viable amendment to (6.1-3) by theorem 5.5-1. From example 6.2-2, the amended Schwarzchild radius is given by $r = \sqrt{4c^2-a^2}$. Let $4c^2 > a^2$ so that the apparent singularity exists. If $b = \sqrt{4c^2-a^2}$, the amended Schwarzchild radius coincides with an isolated closed path with respect to the travelling or external observer. It acts as a limit-cycle if $P'(b)$ exists. It still erases the difference between what the observers see even if $P'(b)$ does not exist - the closed orbit is just reached in finite, as opposed to infinite parameter time. If $\gamma = \frac{-K_3}{b^2+a^2}$, $r = b$ is a critical circle and in order for the geodesics to be complete, $P'(b)$ must exist, by theorem 5.5-1. In this case, the amended Schwarzchild radius is never reached. ($P'(b)$ is given by 6.3-6 with 0 replaced by b and exists if

$$-A'(b)\left(1 + \frac{K_3^2}{F(b)}\right) + \frac{K_3^2 A(b) F'(b)}{F^2(b)} = 0 \quad \text{from (6.3-5).}$$

Notice that the amended Schwarzchild radius $r = b$ only coincides with a critical circle, limit-cycle or a closed path reached in finite time with respect to the travelling observer, for particular values of γ, K_1, K_3.

Recall that the nature of orbits of the Schwarzchild equations can be determined by finding the turning points of an 'effective potential' (given by (6.1-7) in the material particle case). The conclusions are given by theorems 6.1-1, 6.1-2. The amended equations (6.2-4), also possess an effective potential, given by, in the material particle case,

$$V^2(r) = +A(r)\ (1 + \frac{K_2}{F(r)}) \qquad (6.3\text{-}12)$$

and this clearly reduces to the potential given by (6.1-7) in the weak field limit by the hypotheses on $A(r)$, $F(r)$. The turning points of this amended potential are given by,

$$A'(r)\left(1 + \frac{K_2}{F}\right) - AF'\left(\frac{K_2}{F^2}\right) = 0 \qquad (6.3\text{-}13)$$

and, so provided that departures from the original Schwarzchild functions, $A(r) = 1 - \frac{2c}{r}$, , $F(r) = r^2$, are not too great, departures from the original turning points will not be too great, and orbits will behave qualatively the same, except perhaps in a given small neighbourhood of $r = 0$. In particular, if the original Schwarzchild equations possess almost periodic solutions, any feasible amendment of them should also possess almost periodic solutions. This is, in fact, the case with the numerical examples considered in section 6.5). Having obtained there exist almost periodic solutions in a particular case, it is possible to talk about peri-helion shifts of these orbits. Now, from (6.2-4),

$$\frac{dr}{d\phi} = \frac{F}{K_3}\sqrt{K_1^2 - A(r)\left(\frac{K_3^2}{F} + 1\right)} \qquad \Rightarrow$$

if $\alpha = \frac{1}{K_3^2}$, $\beta = \frac{K_1}{K_3}$, $\frac{dr}{d\phi} = 0$ at $r = r_+$ and r_-, then,

$$\phi = \int_{r_-}^{r_+} \frac{dr}{F\sqrt{\beta^2 - A/F + \alpha A}} \qquad (6.3\text{-}14)$$

gives the angle between opposite perihelions. Of course, in general, the integral in (6.3-14) has to be evaluated numerically.

From the general system (6.3-3), the following integral is obtained:-

$$\int \frac{Q(r)}{P(r)}\,dr = \phi + \phi_0 \qquad (6.3\text{-}15)$$

as in the Newtonian case, so that by (6.2-4)(b) and (c) and (6.3-1), in either the travelling observer or external observer case,

$$I = \int \frac{K_3\, dr}{F(r)\sqrt{K_1^2 + A(r)\left(\varepsilon - \frac{K_3^2}{F(r)}\right)}} = \phi + \phi_o \qquad (6.3\text{-}16)$$

Once again, evaluation of the integral in (6.3-16) is, in general, impossible by analytic means but in certain cases may be evaluated in terms of elliptic integrals and the nature of orbits may be ascertained.

Example 6.3-4

Consider the general model (6.3-3)

with $F(r) = r^2 + a^2$ and $A(r) = 1 - \frac{2cr}{r^2+a^2}$ where $|a| << 1$.

In the timelike case, (6.3-16) becomes,

$$I = K_3 \int \frac{dr}{\sqrt{r^4(K_1^2-1)+2cr^3+r^2(2a^2K_1^2-2a^2-K_3^2)+r(2ca^2+2cK_3^2)+(K_1^2a^4-a^4-a^2K_3^2}} \qquad (6.3\text{-}17)$$

Thus, if the constants are chosen so that $K_1 = 1$, $c = \frac{1}{2}$, $a = 1$, $K_3^2 = 2$, (and so $K_1^2a^4 - a^2K_3^2 \neq 0$ and the model is a viable amendment to the Schwarzchild equations),

$$I = \sqrt{2}\int \frac{dr}{\sqrt{r^3-2r^2+3r-2}} = \sqrt{2}\int \frac{dr}{\sqrt{(r-1)\left((r-\frac{1}{2})^2+\frac{7}{4}\right)}} \Rightarrow$$

$$\phi + \phi_o = 2^{1/4}\, cn^{-1}\left(\frac{\sqrt{2}+1-r}{\sqrt{2}-1+r},\ (\tfrac{1}{2} - \tfrac{\sqrt{2}}{8})^{\frac{1}{2}}\right)_{(r>1)}. \qquad (6.3\text{-}18)^3$$

Now $cn\phi$ is a periodic function (for real values of ϕ) and $-1 \leq cn\phi \leq 1$, independent of the modulus. However, r becomes infinite when $cn\phi = -1$, and so the orbits in this case, cannot be closed. In general, if the quartic in (6.3-17) is factorized as far as possible, r can be expressed in terms of Jacobi elliptic functions, using the

results of[3].

6.3.3) Structural Stability

As stated previously, the space co-ordinates in the two-dimensional amended Schwarzchild case are solutions of the system of equations (6.3-2), which have the same general form as (5.6-2). Thus, the analysis of subsection 5.6.3) is appropriate in this case. In particular, lemma 5.6-5 will hold. Once again, (6.3-3) will be structurally stable for almost all values of the arbitrary constants contained in P(r), Q(r). The set of general equations (6.2-4), which include the equation for the co-ordinate time, have no critical points in (t,x,y) space by hypothesis and so are naturally structurally stable.

Example 6.3-5

Consider the feasible amendments to the Schwarzchild equations (6.2-8), given in example 6.2-2. ($\alpha \neq a$). By lemma 5.6-5, the system is structurally stable as a whole. The system has a critical point in (x,y) space at (0,0) if $K_1^2 = \left(1 - \frac{2c}{a}\right)\left(1 + \frac{K_3^2}{a^2}\right)$ (see example 6.2-3). This will be a focus if $aK_3^2 - 3c\,K_3^2 - ca^2 \geq 0$ and so the system will only be structurally unstable if K_1 and K_3 satisfy the first condition, but K_3 does not satisfy the second.

6.4) Properties of the Amended Metric

In this section, various properties of the general metric from which the amended Schwarzchild equations are derived, namely,

$$ds^2 = -A(r)dt^2 + \frac{1}{A(r)}\,dr^2 + F(r)\,(d\theta^2 + \sin^2\theta d\phi^2) \qquad (6.4\text{-}1)$$

where A(r), F(r) have the properties given in section 6.2), are studied.

6.4.1) Generalized Kruskal-Szekeres Co-ordinates

As mentioned in section 6.2), the equations of motion (6.2-7) contain an apparent singularity at r = a where A(a) = 0. This value of r corresponds to the amended Schwarzchild radius. There may be several

such values of r. In this sub-section, it is shown how these apparent singularities may be removed from the metric (6.4-1).

Consider the general co-ordinates given by,

$$\begin{cases} u = \alpha^{\frac{1}{2}}(r)\ e^{\beta(r)}\ \cosh\ (tK) \\ v = \alpha^{\frac{1}{2}}(r)\ e^{\beta(r)}\ \sinh\ (tK) \end{cases} \qquad (6.4\text{-}2)$$

where $\alpha^{\frac{1}{2}}(r)$ is assumed to be real, $K \in \mathbb{R}$.

$$\text{Now } du = \left\{ \tfrac{1}{2}\alpha'\alpha^{-\frac{1}{2}}e^{\beta}\cosh(tK) + \alpha^{\frac{1}{2}}\beta' e^{\beta(r)}\ \cosh(tK) \right\} dr$$

$$+ K\alpha^{\frac{1}{2}}e^{\beta}\ \sinh\ (tK)dt \qquad (6.4\text{-}3)\,(a)$$

$$dv = \left\{ \tfrac{1}{2}\alpha'\alpha^{-\frac{1}{2}}e\ \sinh(tK) + \alpha^{\frac{1}{2}}\beta' e^{\beta(r)}\ \sinh(tK) \right\} dr$$

$$+ K\alpha^{\frac{1}{2}}e^{\beta}\ \cosh(tK)dt \qquad (6.4\text{-}3)\,(b)$$

Squaring equations (6.4-3)(a) and (b) and subtracting, we obtain, after some simplification,

$$du^2 - dv^2 = e^{2\beta(r)} \left\{ \tfrac{1}{2}\alpha'\alpha^{-\frac{1}{2}} + \alpha^{\frac{1}{2}}\beta' \right\}^2 dr^2 - K^2\alpha e^{2\beta}dt^2 \implies$$

$$\frac{A(r)}{\alpha K^2}\, e^{-2\beta}(du^2 - dv^2) = \frac{A}{\alpha K^2}\left(\frac{1}{4}\frac{(\alpha')^2}{\alpha} + \alpha(\beta')^2 + \alpha'\beta' \right) dr^2 - A dt^2 \qquad (6.4\text{-}4)$$

where A(r) is the known function in the amended Schwarzchild case.

Comparing (6.4-4) with terms in the general metric (6.4-1), in order to get $-Adt^2 + \frac{1}{A}dr^2$ easily in terms of $du^2 - dv^2$, it is clear that,

$$\frac{1}{A} = \frac{A}{\alpha^2}\left(\frac{1}{4}\frac{(\alpha')^2}{\alpha} + \alpha(\beta')^2 + \alpha'\beta' \right) \qquad (6.4\text{-}5)$$

It is now supposed that $\frac{1}{2}\alpha' = \beta'$ (in keeping with the original derivation of Kruskal-Szekers co-ordinates with $r > 2c$), and so (6.4-5) represents a differential equation for the unknown function $\alpha(r)$, in terms of the known function A(r). From (6.4-5) we get,

$$\frac{1}{A^2} = \frac{(\alpha')^2}{\alpha K^2}\left\{\frac{1}{2}\alpha^{-\frac{1}{2}} + \frac{1}{2}\alpha^{\frac{1}{2}}\right\}^2 \quad \Longrightarrow$$

$$\frac{2K}{A} = \pm\alpha'(r)\ \left(\frac{1+\alpha}{\alpha}\right) \quad \Longrightarrow$$

$$\alpha e^{\alpha} = \exp\left\{\pm\int\frac{2K}{A}\,dr + \gamma\right\} \qquad (6.4\text{-}6)$$

where γ is a constant of integration.

Now in the Schwarzchild limit, $A = 1 - \frac{2c}{r}$ and comparing (6.4-6) with the Kruskal-Szekers co-ordinates case for $r > 2c$, it is seen that,

$$\left(\frac{r}{2c} - 1\right)e^{\frac{r}{2c}-1} = \exp\left\{\pm 2K(r + 2c\,\ln|r-2c|) + \gamma\right\}$$

and so $K = \frac{1}{4c}$, $\gamma = -(1+\ln 2c)$ (positive signs being taken).

If, instead, it is supposed that $\frac{1}{2}\alpha' = -\beta'$ (in keeping with the original derivation of Kruskal-Szekers co-ordinates with $(r < 2c)$, it is easily seen that the solution of (6.4-5) is,

$$\alpha e^{-\alpha} = \exp\left\{\pm\int\frac{2K}{A}\,dr + \gamma\right\} \qquad (6.4\text{-}7)$$

and once again, in the Schwarzchild case, positive signs are taken and $K = \frac{1}{4c}$, $\gamma = -(1+\ln 2c)$. Notice, if $r < 2c$ in the original case, the roles of u and v are interchanged, so that when $\alpha(r)$ changes sign (as a function of r), the transformed metric does not. Thus, in this generalized case, the roles of u and v need to be interchanged at $r = a$. Thus, u and v are given by (6.4-2) for $r > a$ and,

$$\begin{cases} u = \alpha^{\frac{1}{2}}(r)e^{\beta(r)}\sinh(tK) \\ v = \alpha^{\frac{1}{2}}(r)e^{\beta(r)}\cosh(tK) \end{cases} \qquad (6.4\text{-}8)$$

for $r < a$.

From (6.4-2), (6.4-8),

$$u^2 - v^2 = \pm\,\alpha e^{2\beta} \qquad (6.4\text{-}9)$$

and from (6.4-4), (6.4-5),

$$\frac{1}{A}dr^2 - Adt^2 = \frac{\pm A}{K^2\alpha} e^{-2\beta}(du^2 - dv^2) \qquad (6.4\text{-}10)$$

and as r passes through a, the transformed metric remains the same. The metric contains no apparent singularity at $r = a$ if, $\lim_{r\to a} \frac{A(r)}{\alpha(r)} = \lim_{r\to a} \frac{A'(r)}{\alpha'(r)}$ is finite, so that $\alpha'(a)$ must be finite. From (6.4-6), (6.4-7),

$$\alpha e^{\pm\alpha} = F(r) \text{ say, } \Rightarrow \alpha' e^{\pm\alpha}(1 \pm \alpha) = F'(r) \qquad (6.4\text{-}11)$$

If constants in F(r) are chosen so that $F(a) = 0 \Rightarrow \alpha(a) = 0$ (a condition required for continuity of co-ordinates), then from (6.4-11), $\alpha'(a) = F'(a)$. Now α always remains positive and so $r = a$ must be a turning point of F(r) and so $F'(a) \neq 0$ as required. That is, the transformed metric will not contain an apparent singularity at $r = a$ if $A'(a)$ is finite. (This will be the case for all viable amendments).

The results of this analysis, can be summarized in the following theorem:-

Theorem 6.4-1

An apparent singularity in the metric (6.4-1) at $r = a$, can be transformed away by use of the co-ordinates,

$$\begin{cases} u = \alpha^{\frac{1}{2}}(r)e^{(\frac{1}{\alpha}+K_1)}\cosh(tK_2) & r > a \\ v = \alpha^{\frac{1}{2}}(r)e^{(\frac{1}{\alpha}+K_1)}\sinh(tK_2) \end{cases}$$

$$\begin{cases} u = \alpha^{\frac{1}{2}}(r)e^{(-\frac{1}{\alpha}+K_1)}\sinh(tK_2) & r < a \\ v = \alpha^{\frac{1}{2}}(r)e^{(-\frac{1}{\alpha}+K_1)}\cosh(tK_2) \end{cases}$$

and $\alpha e^{\alpha} = \exp\left\{\pm\int \frac{2K_2}{A}dr + \gamma\right\}$ $\quad r > a$

$\alpha e^{-\alpha} = \exp\left\{\pm\int \frac{2K_2}{A}dr + \gamma\right\}$ $\quad r < a$

K_1, K_2, γ constants. (Constants being chosen so that $\alpha(a) = 0$).

□

Example 6.4-1

In general, the co-ordinates given by the theorem may be impossible to obtain explicitly and even in simple cases can be quite complicated:- consider the amended Schwarzchild model given in example 6.2-2. In this case,

$$A(r) = 1 - \frac{2c}{\sqrt{r^2 + \delta^2}} \quad (\delta \text{ a small positive constant})$$

$\therefore A(r) = 0 \Rightarrow r = a = \sqrt{4c^2 - \delta^2}$ is the only amended Schwarzchild radius. (δ taken small enough so that $4c^2 - \delta^2 > 0$).

$$\text{Now } \int \frac{dr}{A} = \int \frac{\sqrt{r^2 + \delta^2}}{\sqrt{r^2 + \delta^2} - 2c}\, dr = r + \int \frac{2c}{\sqrt{r^2+\delta^2}-2c}\, dr$$

Let $r = \delta \sinh\theta$ so that,

$$\int \frac{dr}{A} = r + 2c \left\{ \theta + \int \frac{2c\, d\theta}{\delta\cosh\theta - 2c} \right\} \qquad (6.4\text{-}12)$$

$$\text{Now } I = \int \frac{d\theta}{\delta\cosh\theta - 2c} = \begin{cases} \dfrac{2}{\sqrt{\delta^2-4c^2}} \tan^{-1}\left(\dfrac{\delta e^{\theta} - 2c}{\sqrt{\delta^2-4c^2}} \right) & \delta > 2c \\ \dfrac{1}{\sqrt{4c^2-\delta^2}} \ln \left| \dfrac{\delta e^{\theta} - 2c - \sqrt{4c^2-\delta^2}}{\delta e^{\theta} - 2c + \sqrt{4c^2-\delta^2}} \right| & \delta < 2c \\ -\dfrac{1}{2c} \coth\dfrac{\theta}{2} & \delta = 2c. \end{cases}$$

In any viable amendment to the Schwarzchild case, $\delta < 2c$ and the second case is appropriate. (As stated earlier, an amended Schwarzchild radius will not exist if $\delta > 2c$).

Substituting for I in (6.4-12), and then for θ,

$$\int \frac{dr}{A} = r + 2c \ln(r + \sqrt{r^2+\delta^2}) - 2c \ln\delta + \frac{4c^2}{\sqrt{4c^2-\delta^2}} \ln \left| \frac{r + \sqrt{r^2+\delta^2} - 2c - \sqrt{4c^2-\delta^2}}{r + \sqrt{r^2+\delta^2} - 2c + \sqrt{4c^2-\delta^2}} \right| \Rightarrow \qquad (6.4\text{-}13)$$

$$\alpha e^{\pm\alpha} = (r + \sqrt{r^2+\delta^2})^{4K_2c} \exp(2K_2 r + \gamma) \left\{ \frac{r + \sqrt{r^2+\delta^2} - 2c - \sqrt{4c^2-\delta^2}}{r + \sqrt{r^2+\delta^2} - 2c + \sqrt{4c^2-\delta^2}} \right\}^{\frac{8K_2c^2}{\sqrt{4c^2-\delta^2}}}$$

Notice, at $r = a$, $\alpha = 0$, as required (independent of the choice of γ and K_2). Also, differentiating the right-hand-side of (6.4-13), gives $\alpha'(a) \neq 0$ as required.

It may happen that the function $\alpha(r)$ given for $r>a$ never changes sign. In this case, only values of u and v given by (6.4-2) need be used.

It has been assumed that there is only one apparent singularity at $r = a$ present in the metric (6.4-1). Clearly, theorem 6.4-1 can be generalized to cater for any number of apparent singularities by taking successive pairs of co-ordinates and v and swopping these around and changing signs of α everytime one of these 'singularities' are crossed.

6.4-2 Corrected Weak and Strong Energy Conditions

It was shown in section 4.3, how certain metrics which give rise to geodesic equations having global existence of solutions may not satisfy the weak and strong energy conditions. It was shown that the null and timelike convergence conditions only hold if the corrected weak and strong energy conditions hold in the space-time defined by the given metric. Here, a specific example is given of a metric where the null and timelike convergence conditions do not hold.

Consider the general metric (6.4-1) where $A(r)$, $F(r)$ are assumed to have the properties given in section 6.3), so that the derived geodesics are complete. The connection coefficients associated with (6.4-1) are given by (2.2-5) and the components of the Ricci tensor by

$$\begin{cases} R_{00} = \frac{A}{2}\left\{\frac{A'F'}{F} + A''\right\} \\ R_{11} = -\frac{1}{F}\left(F'' - \frac{1}{2}\frac{(F')^2}{F}\right) - \frac{1}{A^2}R_{00} \\ R_{22} = 1 - \frac{1}{2}(A'F' + AF'') \\ R_{33} = \sin^2\theta R_{22} \end{cases} \qquad (6.4\text{-}14)$$

all other components being zero. It is easily seen that, in the case

of the Schwarzchild metric (6.1-2), $R_{\alpha\alpha} = 0$ for all α and so the Einstein field equations in free space (2.3-1), (with $\Lambda = 0$, $T_{ab} = 0$), are satisfied.

Consider the special case where

$$A(r) = 1 - \frac{2}{\sqrt{r^2+a^2}}, \quad F(r) = r^2 + a^2, \quad |a| << 1. \qquad (6.4\text{-}15)$$

so that (6.4-1) gives rise to complete geodesics, by example 6.2-2 (c has been taken to be 1 for simplicity).

Substituting (6.4-15) into (6.4-14),

$$\begin{cases} R_{00} = \left(1 - \dfrac{2}{\sqrt{r^2+a^2}}\right) \cdot \dfrac{a^2}{(r^2+a^2)^{5/2}} \\ R_{11} = \dfrac{-2a^2}{(r^2+a^2)^2} - \left(1 - \dfrac{2}{\sqrt{r^2+a^2}}\right)^{-1} \cdot \dfrac{a^2}{(r^2+a^2)^{5/2}} \\ R_{22} = \dfrac{2a^2}{(r^2+a^2)^{3/2}} \\ R_{33} = \sin^2\theta R_{22} \end{cases} \qquad (6.4\text{-}16)$$

Notice as $r \to \infty$, all components of the Ricci tensor tend to zero as required, but the Einstein equations are not satisfied exactly, especially at $r = 0$, if $a \neq 0$. Also, in the limit as $a \to 0$, all the components tend to zero (for constant r).

Consider a null or timelike vector with components w^i at any point m of the space-time belonging to a co-ordinate domain of the given chart, so that,

$$g_m(w,w) = -c^2 \Rightarrow g_{ij}\, w^i w^j = -c^2 \Rightarrow$$

$$-\left(1 - \frac{2}{\sqrt{r^2+a^2}}\right)(w^0)^2 + \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right)^{-1} (w^1)^2$$

$$+ (r^2+a^2)\left\{(w^2)^2 + \sin^2\theta\,(w^3)^2\right\} = -c^2 \qquad (6.4\text{-}17)$$

where m has co-ordinates (r,θ), and $c \in \mathbb{R} - \{0\}$ if w is a timelike

vector and c = 0 if w is a null vector.

Now $R_{ij}w^iw^j = R_{00}(w^0)^2 + R_{11}(w^1)^2 + R_{22}(w^2)^2 + R_{33}(w^3)^2 \Rightarrow$

$$R_{ij}w^iw^j = \left\{ \frac{a^2}{(r^2+a^2)^{5/2}} \cdot \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right) \right\} (w^0)^2$$

$$- \left\{ \frac{2a^2}{(r^2+a^2)^2} + \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right)^{-1} \cdot \frac{a^2}{(r^2+a^2)^{5/2}} \right\} \times (w^1)^2$$

$$+ \frac{2a^2}{(r^2+a^2)^{3/2}} \left\{ (w^2)^2 + \sin^2\theta (w^3)^2 \right\} \qquad (6.4\text{-}18)$$

from (6.4-16) and so that in the Schwarzchild (a=0) case, the null and timelike convergence conditions are trivially satisfied at all points.

Substituting (6.4-17) into (6.4-18) we get,

$$R_{ij}w^iw^j = \left\{ \frac{a^2}{(r^2+a^2)^{5/2}} \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right) \right\} \times (w^0)^2$$

$$- \left\{ \frac{2a^2}{(r^2+a^2)^2} + \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right)^{-1} \cdot \frac{a^2}{(r^2+a^2)^{5/2}} \right\} \times (w^1)^2$$

$$+ \frac{2a^2}{(r^2+a^2)^{5/2}} \left\{ \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right)(w^0)^2 - \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right)^{-1} (w^1)^2 - c^2 \right\}$$

$$\therefore R_{ij}w^iw^j = \left\{ \frac{3a^2}{(r^2+a^2)^{5/2}} \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right) \right\} (w^0)^2 - \left\{ \frac{3a^2}{(r^2+a^2)^{5/2}} \times \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right)^{-1} + \frac{2a^2}{(r^2+a^2)^2} \right\} (w^1)^2$$

$$- \frac{2c^2a^2}{(r^2+a^2)^{5/2}} \qquad (6.4\text{-}19)$$

Clearly, from (6.4-17) and (6.4-19), $R_{ij}w^iw^j < 0$ when

$$\left(1 - \frac{2}{\sqrt{r^2+a^2}}\right)(w^0)^2 < \left(1 - \frac{2}{\sqrt{r^2+a^2}}\right)^{-1} (w^1)^2 + \frac{2}{3}(r^2+a^2)^{\frac{1}{2}}(w^1)^2 + \frac{2c^2}{3} \Rightarrow$$

$$\frac{c^2}{3} + (r^2+a^2)\left\{(w^2)^2 + \sin^2\theta (w^3)^2\right\} < \frac{2}{3}(r^2+a^2)^{\frac{1}{2}}(w^1)^2 \qquad (6.4\text{-}20)$$

This inequality is clearly-satisfied at an infinite number of points, by choosing the w^i suitably. For instance, if $w^2 = w^3 = 0$, $w^1 = 1$, $R_{ij}w^iw^j<0$ for points where $c^2< 2(r^2+a^2)^{\frac{1}{2}}$ and W^0 is obtained from (6.4-17).

6.5) Numerical Solutions of Specific Models

In this section, various numerical solutions of specific chosen amendments of the Schwarzchild equations are considered, as in the Newtonian cases. Only amendments in two space dimensions and of the specific type (6.2.-4) are considered. Various solutions of the equations are plotted for the material test particle case, so that $\varepsilon = -1$ and λ, the affine parameter, in (6.2-4), represents the proper time. Solutions are only plotted for amendments with $F(r)$ given by $F(r) = r^2+a^2$, $|a| \ll 1$ as it can be shown that there is no significant difference in solutions of amendments with the same choice of $A(r)$, one having $F(r) = r^2+a^2$ and one having an $F(r)$ possessing the same qualitative properties, for example, $F(r) = \dfrac{\pi r^2}{2\tan^{-1}(br^3)}$, $|b| \gg 1$. As in the two-dimensional Newtonian cases, K_3 is taken to be unity, without loss of generality and K_1 is chosen so that orbits of a chosen qualitative character are obtained. Thus, the general amended equations (including the original Schwarzchild equations), are considered in the form,

$$\begin{cases} \dfrac{dt}{d\tau} = \dfrac{K_1}{A(r)} \\ \dfrac{dr}{d\tau} = \left\{ K_1^2 - A(r)\left(1 + \dfrac{1}{r^2+a^2}\right)\right\}^{\frac{1}{2}} \\ \dfrac{d\phi}{d\tau} = \dfrac{1}{r^2+a^2} \end{cases} \qquad (6.5\text{-}1)$$

$(|a| \ll 1)$

where τ denotes proper time, so that distinct amendments are defined by a particular choice of $A(r)$, (any particular choice satisfying the conditions given in subsection 6.22)).

Recall that in the Newtonian case, the constants K,A in the first order two-dimensional equations (5.4-3) can be chosen so that the orbit is hyperbolic, parabolic, or elliptic. Then, with the considered amendments of section 5.7) of the general form (5.6-1), with amended constants, so that initial conditions remained constant for any amendment

of each separate case, the amended orbits were almost hyperbolic, parabolic, or almost closed with a peri-helion shift. In the Schwarzchild case, (equation (6.5-1) with $A(r) = 1 - \frac{2c}{r}$ and $a = 0$), the qualitative behaviour of orbits is dictated by the conditions of theorem 6.1-1, that is, the turning points of the effective potential. For instance, if $K_1 = 0.94$, $K_3 = 4.0$, the solutions are almost periodic orbits (non-eccentric orbits), which never pass through the Schwarzchild radius, and so never approach the origin. An orbit of this type is shown in figure 6.1. Any orbits of amended equations, with the same initial conditions, agree almost exactly with the Schwarzchild orbit, as do the hyperbolic-like orbits (given in the Schwarzchild case, for instance, by $K_1 = 1.998$, $K_3 = 1$).

Unlike in the Newtonian case, the constant K_1 can be chosen in the Schwarzchild case, so that orbits pass through the Schwarzchild radius and must approach the singularity at $r = 0$. Curve (1) in figure 6.2 is such an orbit, and cannot be extended past the origin. (In this case, $c = 1$, without loss of generality and $K_1 = 0.884$). Thus the Schwarzchild case is fundamentally different from the Newtonian case in the sense that Newtonian geodesics are only incomplete if they approach the origin radially, whereas any Schwarzchild geodesic passing through the Schwarzchild radius is incomplete. The amendments of Schwarzchild orbits which pass through the Schwarzchild radius are the most interesting, both from a mathematical viewpoint and from a philosophical viewpoint, and it is these solutions which are now examined.

Solutions of amended equations are compared with the Schwarzchild orbit shown by curve (1) of figure 6.2. The orbit starts from rest at $x = 5$ (at time $\tau=0$), and corresponds to the eccentric elliptic orbit given by curve (1), figure 5.3 in the Newtonian case. K_1 is chosen in the amended equations in such a way that the solutions under consideration have the same initial conditions as this Schwarzchild orbit. The amended solutions agree almost exactly with the Schwarzchild orbit until they pass through the Schwarzchild radius at $r = 2$. It has been stated earlier that the equations can be amended in such a way that a Schwarzchild radius is no longer present. However, whether an apparent singularity

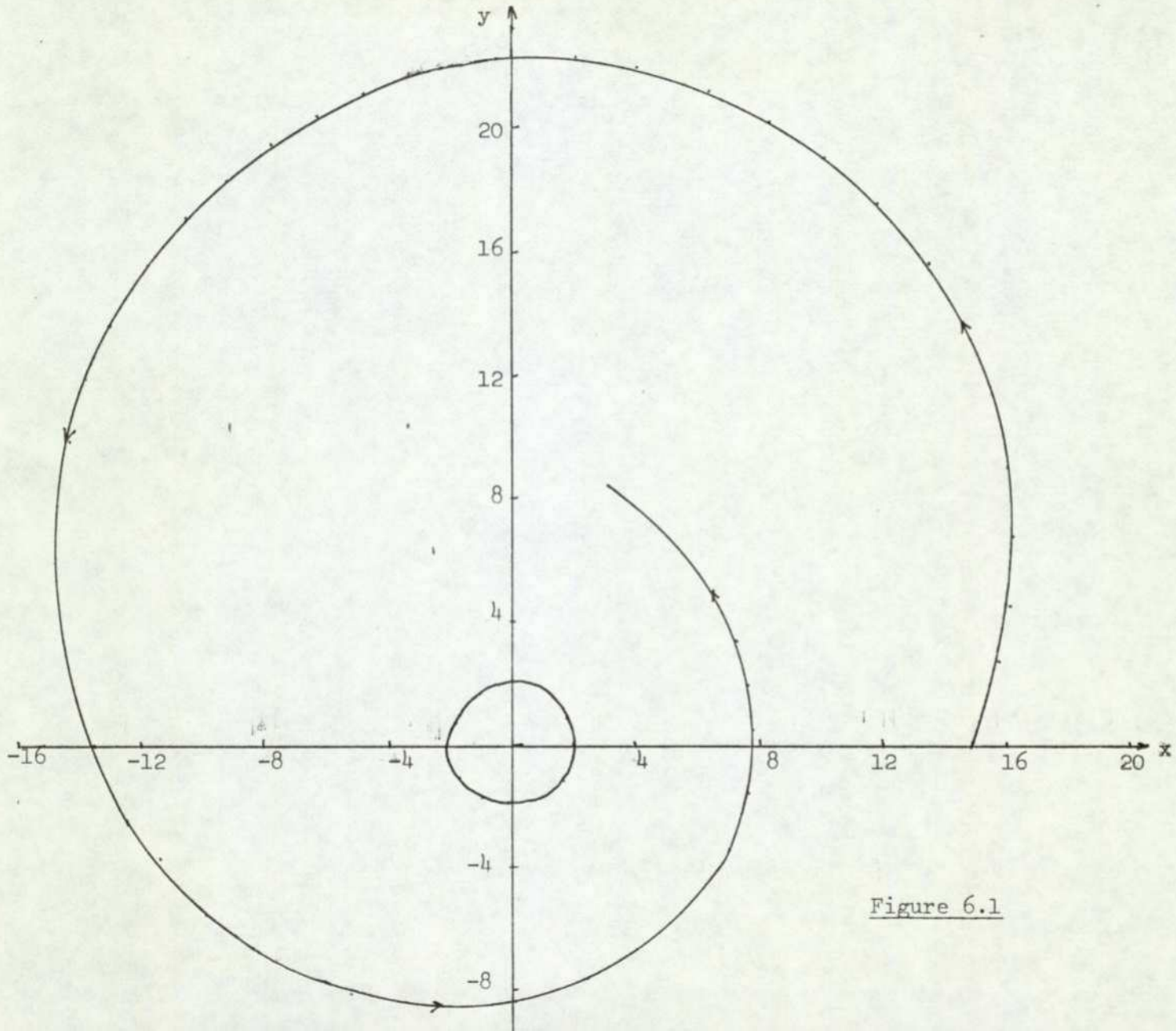


Figure 6.1

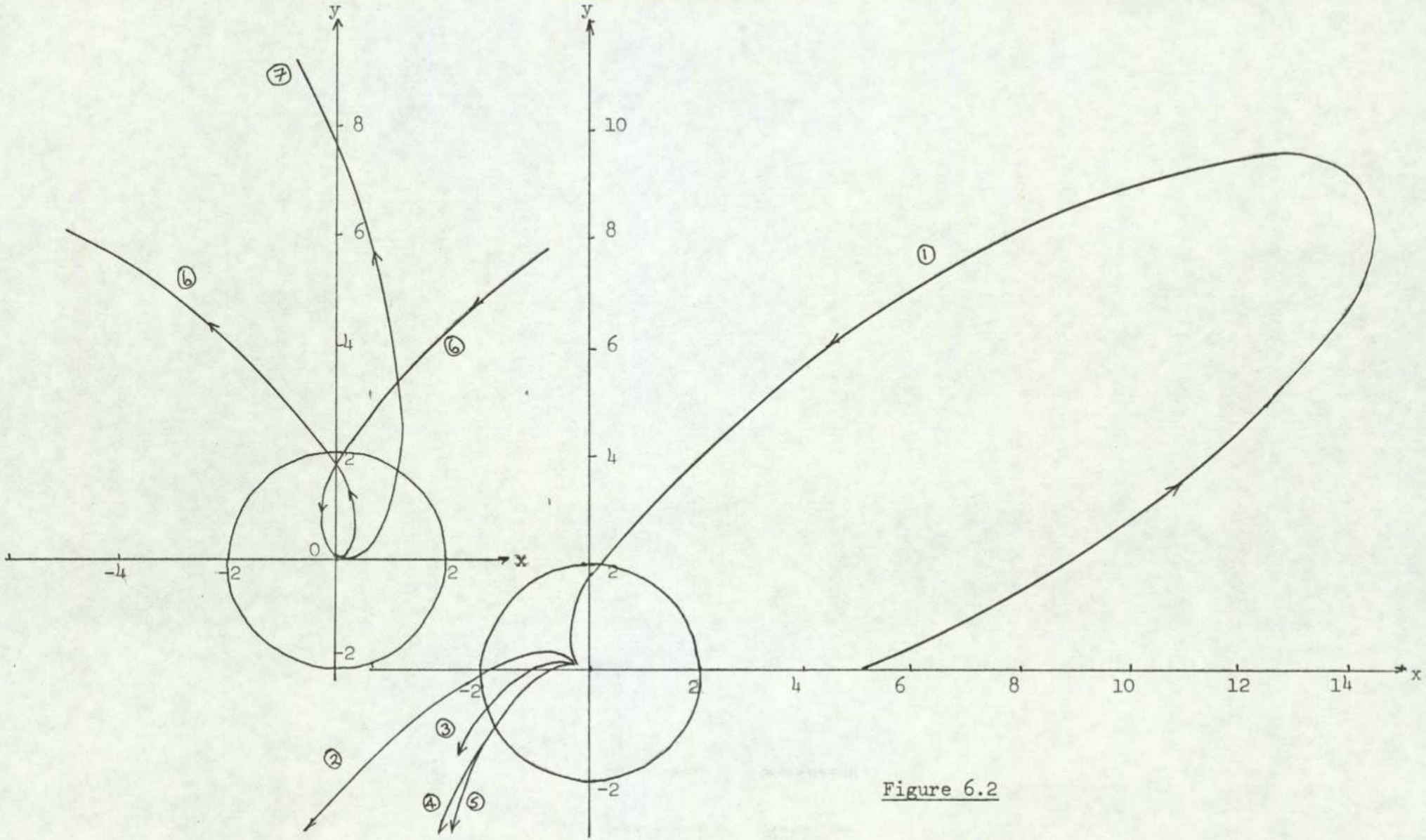


Figure 6.2

is present or not, the qualitative behaviour of the orbit, as far as the travelling observer is concerned, will remain the same. All the amendments considered now have an amended Schwarzchild radius. Clearly, as the amended orbits approach the origin, they will behave qualitatively the same as the one-dimensional counter parts, with respect to the proper time, as given in section 5.7. That is, an orbit will pass through the origin and come back, get thrown outwards again before it reaches the origin, or pass through the origin, so that the radial co-ordinate changes sign, depending on the behaviour of a radial orbit of the same set of equations. For instance, curves (2) and (5) of figure 6.2 (together with curve (1)) are solutions of (6.5-1), with $A(r) = 1-2\int f(r)dr$, $f(r) = \frac{2}{\pi}\tan^{-1}\frac{200r^2}{r^2}$, with $a^2 = 0.1, 0.001$, respectively, (compare with the numerical solutions of (5.7-1) - curves (1), (2) of figure 5.2). Curve (4) of the figure, together with curve (1) is the solution of (6.5-1) with $A(r) = 1 - 2\int f(r)dr$, $f(r) = \frac{-r^2}{r^4 + 0.001}$, (compare with the numerical solutions of (6.7-4)). If the constant denominator of $f(r)$ is 0.1, the amended solution agrees approximately with curve (2). If the function $f(r)$ is perturbed slightly to $f(r) = \frac{0.1 - r^2}{(r^2 + 0.1)^2}$, the corresponding solution is curve (3).

Curves (6) and (7) of figure 6.2 are typical solutions of an amendment in which any orbit approaching the origin radially, passes through and the r co-ordinate becomes negative. They are solutions of (6.5-1) with $A(r) = 1 - \frac{2}{(r^2+a^2)^{\frac{1}{2}}}$ and $a = 0.001, 0.1$ respectively. The curves (5) and (6) are extended in figures 6.3 and 6.4 respectively. The apparent contradictory nature of both the orbits is explained by the fact that the motion of the test particle is described by two sets of differential equations, as opposed to just (6.5-1), depending on the sign of the radial coordinate r, as the motion along any radial path must be symmetrical about the origin (see section 5.7). Thus, although the solutions to (6.5-1) in the given cases have orbits as described in the diagrams, the actual path of a test particle will be an almost periodic orbit about the origin, that is, two of the loops of the solution are not actually traversed. The particle will have

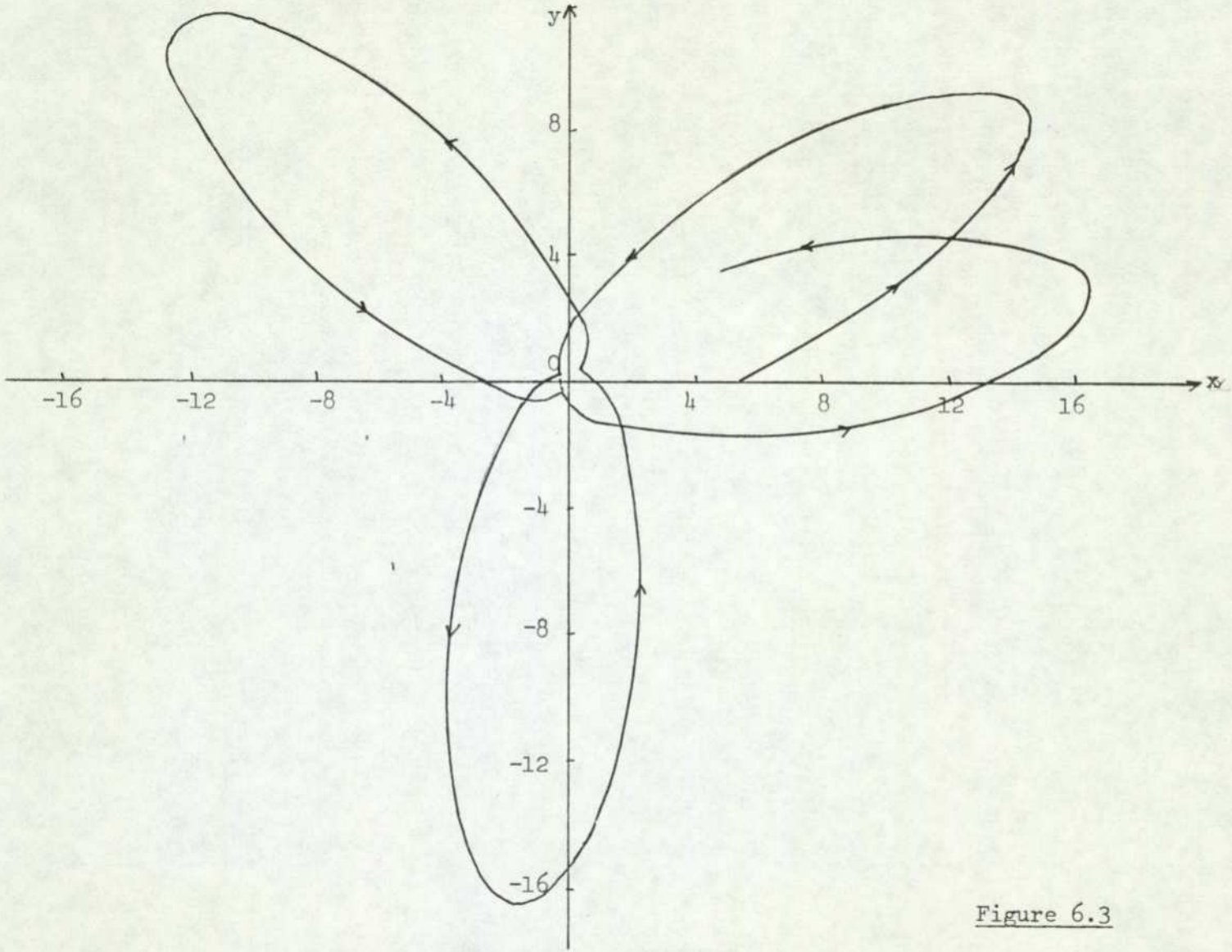


Figure 6.3

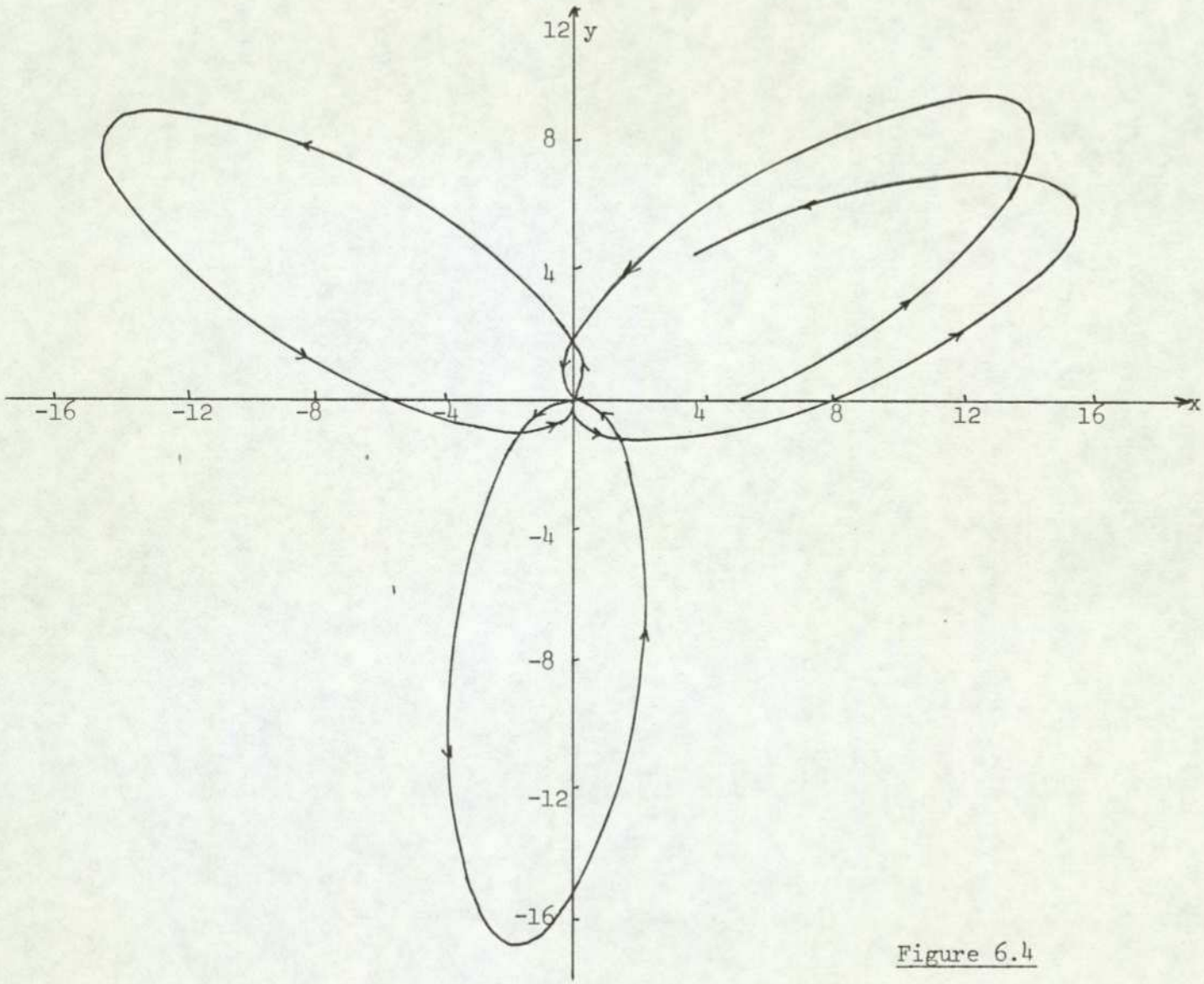


Figure 6.4

almost the same qualitative motion as in the amended Newtonian cases, and the almost-periodic Schwarzchild orbits which do not pass through the Schwarzchild radius. As far as the travelling observer is concerned, the orbit will be almost periodic (with peri-helion shift). Of course, as far as any external observer is concerned, the orbit never reaches the Schwarzchild radius and so any motion inside the radius is not observed. Thus, although the travelling observer passes through the radius, past the origin and out again, it appears to the external observer that he has always been on the outside! The orbit is elongated as far as the external observer is concerned, (with respect to the true orbit).

6.6) Concluding Remarks

This chapter has been concerned with various properties of amendments to the Schwarzchild equations of a specified general nature. As stated in Chapter Five, the work is far from complete - it should be possible to classify general classes of amendments according to their qualitative properties.

The amended Schwarzchild equations and the amended Newtonian equations presented in this and the previous chapter seem to possess similar qualitative properties - the solution bear a closer resemblance than those of their traditional counterparts, (for example, peri-helion shifts etc.). However, the properties of the amended Schwarzchild equations introduce new philosophical problems, such as those associated with orbits passing through an apparent singularity an infinite number of times, as mentioned in the last section. These points are raised in the following, final, chapter.

Chapter 7

Conclusions and Speculations

In this thesis we have studied, from a mathematical point of view, the feasibility of "regularizing" the geodesic equations of certain singular, but otherwise experimentally viable, gravitational spacetimes. That is, spacetimes which include the gravitational effects of matter and which lead to predictions in accord with some of the experimental observations, but in which the effective gravitational force becomes infinite at some point or set of points in the spacetime. The aim of this study is to build a data-base of regularized spacetimes which can then be used in an attempt to develop a nonsingular theory of gravity, much in the same way that Newton used Kepler's data to synthesize his model.

Since this has been a preliminary investigation, we have, naturally enough, concentrated on the two most successful models of gravitation to date. That is, Newtonian gravitation, where gravity is represented by a scalar field acting in a Newtonian spacetime, and Einstein gravitation, where matter gives rise to gravitational effects by deforming Minkowski spacetime into a "curved" psuedo-Riemmanian manifold. Furthermore, we omitted cosmological spacetimes and restricted our attention to the simplest non-trivial solutions of these models. That is, the point-mass field in the Newtonian case and the Schwarzschild spacetime in the Einstein case. The fact that these are unique solutions of their respective models, for the symmetry considered, is an added advantage. Note, however, that our results for the Schwarzschild solution may be transferred to the cylindrically symmetric Kerr metric by means of the complex transformation mentioned in the introduction. Thus, we have effectively covered both the conjectured black-hole states of the Einstein model.

The criterion we used for regularizing the geodesic equation is that the amended equations should form a dynamical system. Since the limit sets of dynamical systems are only approached asymptotically, the resulting geodesic flows will be complete. Our results have confirmed that this is a sufficient criterion for the data-gathering process and, more importantly, have shown that it certainly is possible to eliminate or avoid

singular spacetimes by replacing the singular points by regular, critical or isolated points of a dynamical system. Thus, using both central and noncentral amendments, we have found limit set behaviour corresponding to critical points, limit cycles and isolating limit sets where the critical set disconnects the singularity from the observers' spacetime by forming a natural unimpenetrable boundary. We have, in Chapters 5 and 6, catalogued and analysed these amendments in some considerable detail and in some cases have verified, by numerical computation, that the weak-field properties can be adjusted to fit the experimental data. Thus, the first phase of our programme has been successfully completed.

Our next step, from the mathematical point of view, is to partition the amended geodesic equations into equivalence classes with each class representing a particular type of qualitative behaviour. This would enable us to carry out a detailed study of the differentiability and stability properties of equivalent amendments and, in this way, arrive at a set of criteria to determine the "best" amendment in each class. One could then, in principle, use these best amendments to set up field equations for the corresponding spacetimes. Once this stage has been reached then the conventional approach can be reintroduced.

From the physical point of view, the next step is fairly clear. Each amended spacetime has to be examined carefully and the behaviour of its geodesics must be shown to be "physically reasonable". Furthermore, the final states of dying or collapsing stars must be reassessed now that the black-hole state has been eliminated. However, the broad picture that emerges from our analysis so far can be expressed as follows. If we assume that the gravitational "constant" is a function of the radial coordinate, i.e., $G = G(r)$, then our amendments correspond to situations in which (a) $G(r) \to 0$ as $r \to 0$ in such a way that the effective gravitational force is either finite or zero at the origin, but always remains attractive and (b) $G(r)$ becomes negative for small r so that the effective gravitational potential develops a repulsive core. In the case of noncentral amendments G will, of course, depend on both r and $\dot{r}$, implying that there is a damping term which slows down the collapse to such an extent that it takes an infinite proper time to become singular. Damping usually means dissipation and consequent energy radiation, and this could

possibly occur in the form of gravitational waves.

Now the requirement of a variable G in our interpretation, raises the question as to whether there is any experimental evidence to support this view and, surprisingly enough, there is. Recent work by Long[1,2,3] indicates that there may be a departure from the inverse square law at laboratory dimensions. In fact, his results show that for $4<r<100$, where r is measured in centimetres, the gravitational constant should be replaced by the function

$$G(r) = G_o\left[1 + (0.002)\ln r\right].$$

If this can be extrapolated down to $r=0$, then it indicates that the gravitational force vanishes very close to the origin and becomes infinitely repulsive at $r=0$. We must, of course, wait until these results are confirmed by others before they can be taken seriously. However, their existence is reassuring.

Another interesting point to note, in the case in which the effective gravitational force always remains attractive, but goes to zero as we approach the origin, is that it leads to an "asymptotically free theory". For it is precisely theories of this type which are now being considered in the investigations of the strong interactions between the quarks of elementary particle theory.

Finally, we come to the question of arbitrary parameters. Our modifications must, of necessity, contain arbitrary parameters to give us the flexibility to adjust the weak-field end of the models. At this early stage of our programme we are inclined to think that they are classical indications of an underlying quantum structure; related, perhaps to Planck's quantum of action or the so-called Planck length of quantum geometrodynamics. In other words, we recover the theories of Einstein and Newton when the parameters go to zero.

References

Chapter One

1. A. Einstein, Ann. Phys. Ser. 4 35, 898 (1911).
2. ———— Ann. Phys. Ser. 4 49, 796 (1916).
3. ———— "The Meaning of Relativity", (Princeton Univ. Press, Princeton, N.J. 1921).
4. C.W. Kilmister and G. Stephenson, "An axiomatic criticism of Unified Field Theories", (I and II) Supp. al Nuovo Cimento Nl 11, 91 and 118 (1954).
5. D.R. Brill and R.H. Gowdy, "Quantization of General Relativity", Rep. Prog. Phys., 33, 413 (1970).
6. B. Carter, "Global Structure of the Kerr Family of Gravitational Fields", Phys. Rev. 174, 1559 (1968).
7. S.W. Hawking, "Breakdown of Predictability in Gravitational Collapse", Phys. Rev. D 14, 2460 (1976).
8. R. Geroch, J. Math. Phys. 13, 956 (1972).
9. E.T. Newman, "The Bondi-Metzner-Sacho Group", in "General Relativity and Gravitation", edited by G. Shariv and J. Rosen (Halsted, New York, 1975).
10. J.N. Goldberg, "Canonical Quantization", in "General Relativity and Gravitation", edited by G. Shariv and J. Rosen (Halsted, New York, 1975).
11. C.W. Misner, K.S. Thorne and J.A. Wheeler, "Gravitation" (Freeman, San Francisco, 1973) p. 1202.
12. R. Adler, M. Bazin and M. Schiffer, "Introduction to General Relativity" (McGraw-Hill, New York, 1965) Chap. 13.
13. J.A. Wheeler, "Geometrodynamics" (Academic Press, New York, 1962).
14. H. Poincaré, "Science and Hypothesis" (Dover, New York, 1952) pp 70 and 71.
15. M.J. Rees, "Observational Effects of Black Holes", in "General Relativity and Gravitation", edited by G. Shariv and J. Rosen (Halsted, New York, 1975).
16. J.N. Bahcall, F.J. Dyson, J.I. Katz and B. Paczynski, Astrophys. J. Lett. 189, L. 17 (1974).
17. B. Schmidt, Gen. Rel. Grav. 1, 269 (1971).
18. R.K. Sachs, Comm. Math. Phys. 33, 215 (1973).
19. C.W. Misner and A.H. Taub, Sov. Phys. - JETP 28, 122 (1969).

20. C.W. Misner, in "Relativity Theory and Astrophysics: Relativity and Cosmology", edited by J. Ehlers (American Math. Soc., Providence, 1967), Vol. 8, p. 160.
21. G.F.R. Ellis and A.R. King, "Was the Big Bang a Whimper", Comm. Math. Phys. 38, 119 (1974).
22. C. Brans. and R.H. Dicke, Phys. Rev., 124, 925 (1961).
23. H. Yilmaz, Phys. Rev., 111, 1417 (1958).
24. B.O.J. Tupper, Il Nuovo Cimento, 17B, 135 (1973).
25. C. Isham, R. Penrose and D. Sciama, eds., "Quantum Gravity" (Clarendon Press, Oxford, 1975).
26. S. Deser and P. van Nienwenhuizen, "Uniqueness and Nonrenormalizability of Quantum Gravitation", in "General Relativity and Gravitation", edited by G. Shariv and J. Rosen (Halsted, New York, 1975).
27. G. Stephenson, Il Nuovo Cimento, 9, 263 (1958).
28. C.W. Kilmister, "Les Theories Relativistes de la Gravitation" (Paris: Centre National de la Recherche Scientifique, 1962).
29. C.N. Yang, Phys. Rev. Lett., 33, 445 (1974).
30. G.W. Barrett, L.J. Rose and A.E.G. Stuart, Phys. Lett. 60A, 278 (1977).
31. L.J. Rose, A.E.G. Stuart and G.W. Barrett, J. Phys. A: Math. Gen., 11 709 (1978).
32. A. Muller, U. Heinz, B. Muller and W. Greiner, J. Phys. A: Math. Gen., 11, 1781 (1978).
33. A. Trautman, Nature Phys. Sci., 242, 7 (1973).
34. A.R. Prasanna, Phys. Rev. D 11, 2076 (1975).

Chapter Two

1. H. Minkowski, "Space and Time" in 'The Principle of Relativity', (Dover, New York, 1923).
2. C. Kaceser, "Introduction to the Special Theory of Relativity", (Prentice-Hall, New Jersey).
3. A. Schild, "Time", Texas Quarterly 3,3 pp. 42-62.
4. A. Schild, "Gravitational Theories of the Whitehead Types and the Principle of Equivalence", in C. Møller, ed., "Evidence for Gravitational Theories", (Academic Press, New York).
5. A. Schild, "Lectures on General Relativity Theory" in J. Ehlers, ed., 'Relativity Theory and Astrophysics...", pp. 1-105 (American Mathematical Society, Providence R.I.).

6. F. Brickell and R.S. Clarke, "Differentiable Manifolds" (Van Nostrand, London) pp 106, 107, 122.
7. ———————————— "Differentiable Manifolds" (Van Nostrand, London) pp 54-57.
8. ———————————— "Differentiable Manifolds" (Van Nostrand, London) pp 113, 114.
9. R. Bishop and S.I. Goldberg, "Tensor Analysis on Manifolds" (Macmillan) p. 85-86.
10. F. Brickell and R.S. Clarke, "Differentiable Manifolds" (Van Nostrand, London) pp 183-189.
11. S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Space-time" (Cambridge University Press, 1973). p. 39.
12. F. Brickell and R.S. Clarke, "Differentiable Manifolds" (Von Nostrand, London) p. 73.
13. M. Hausner and J.T. Schwartz, "Lie Groups. Lie Algebras" (Nelson, 1968) Part One.
14. F. Brickell and R.S. Clarke, "Differentiable Manifolds" (Van Nostrand, London) pp 155-158.
15. S.W. Hawking and G.F.R. Ellis, "The Large-Scale structure of Space-time", (Cambridge Univ. Press, 1973), p. 32.
16. H.V. Craig, "Vector and Tensor Analysis" (McGraw-Hill, 1943), pp. 313-317.
17. F. Brickell and R.S. Clarke, "Differentiable Manifolds" (Van Nostrand, London), pp 167, 168.
18. M. Postnikov, "The Variational Theory of Geodesics", (W.B. Saunders Co, 1967). pp 107-108.
19. S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Space-time". (Cambridge University Press, 1973). pp 107-116.
20. M. Postnikov, "The Variational Theory of Geodesics" (W.B. Saunders Co., 1967) pp 114-115.
21. S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Space-time". (Cambridge University Press, 1973) pp 97-102.
22. ———————————— "The Large Scale Structure of Space-time". (Cambridge University Press, 1973). pp 181-182.
23. K. Schwarzchild, "Vber das Gravitationsfield eines Massen punktes nach der Einsteinschen Theorie" (Sitzber. Deut. Akod Wiss. Berlin. Kl Math-Physics, Tech.) pp 189-196.
24. C.W. Misner, K.S. Thorne and J.A. Wheeler, "Gravitation" (W.H. Freeman and Co. San Francisco), p.826.

25. S. Kobayashi and K. Nomizu, "Foundations of Differential Geometry: Vol. I" (Interscience, New York).
26. E.M. Lifschitz and I.M. Khalatnikov, "Investigations in Relativistic Cosmology", Adv. in Physics (Phil. Mag suppl) 12, pp 185-249.
27. R. Penrose, "Gravitational Collapse and Space-time Singularities", Physics Review Letters 14, pp 57-59.
28. S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Space-time" (Cambridge University Press, 1973), pp 88-91, 94-96.
29. ———————————— "The Large Scale Structure of Space-time" (Cambridge University Press, 1973). Chapter 7, p 301.
30. ———————————— "The Large Scale Structure of Space-time" (Cambridge University Press, 1973).pp194,189
31. ———————————— "The Large Scale Structure of Space-time" (Cambridge University Press, 1973). pp 202-204.
32. S.W. Hawking and R. Penrose, "The Singularities of Gravitational Collapse and Cosmology". Proc. of the Royal Society. London A314, pp 529-548.
33. S.W. Hawking, "The Occurrence of Singularities in Cosmology. III. Causality and Singularities". Proc. of the Royal Society. London. A300, pp 187-201.
34. S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Space-time" (Cambridge University Press, 1973). pp 91-94.
35. R.P. Geroch, "What is a Singularity in General Relativity?", Annals of Physics (New York), 48, pp 526-540.
36. B.G. Schmidt, "A New Definition of Singular Points in General Relativity". Journal of General Relativity and Gravitation, 1, pp 269-280.
37. C.W. Misner and A.H. Taub, "A Singularity-free Empty Universe", Sov. Physics J.E.T.P. 28, pp 122-133.

Chapter Three

1. N.P. Bhata, O. Hajek, "Local Semi-Dynamical Systems", (Lecture Notes in Maths. No. 90) pp6-8.
2. E.A. Coddington and N. Levinson, "Theory of Ordinary Differential Equations", (McGraw-Hill, 1955). p.21.
3. V.V. Nemytskii and V.V. Stepanov, "Qualitative Theory of Differential Equations", (Princeton University Press) p.9.

4. A. Strauss, "The Use of Liapunov Functions for Global Existence" in 'Seminar on Differential Equations and Dynamical Systems' (Lecture Notes in Maths, No. 60). pp76-81.
5. S. Bernfeld, "Perturbations of Systems with Global Existence", in 'Seminar on Differential Equations and Dynamical Systems, II' (Lecture Notes in Maths. No 144) pp 16-21.
6. E.A. Coddington and N. Levinson, "Theory of Ordinary Differential Equations", (McGraw-Hill, 1955). Chapter One.
7. F. Brickell and R.S. Clarke, "Differentiable Manifolds" (Van Nostrand, London) pp26,27.
8. ———————————— "Differentiable Manifolds" (Van Nostrand, London) p.140.
9. W. Kaplan, "Regular Curve Families Filling the Plane I" Duke Math J7, pp 154-185.
10. R. Abraham, "Foundations of Mechanics" (W. Benjamin Inc. New York, 1967), pp156-158.
11. R. Abraham, "Foundations of Mechanics" (W. Benjamin Inc. New York, 1967). pp158-159, 165.
12. E.A. Coddington and N. Levinson, "Theory of Ordinary Differential Equations", (McGraw-Hill, 1955) Chapter fifteen.
13. D.W. Jordan and P. Smith, "Nonlinear Ordinary Differential Equations" (Clarendon Press, Oxford).
14. V.V. Nemytskii and V.V. Stepanov "Qualitative Theory of Differential Equations" (Princeton University Press) pp 135-136.
15. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.G. Maier, "Qualitative Theory of Dynamic Systems on the Plane" (Israel Prog. for Sci. Trans. Jerusalem 1971). pp 205-208.
16. V.V. Nemytskii and V.V. Stepanov, "Qualitative Theory of Differential Equations" (Princeton Univ. Press). pp.136-137, 139-149.
17. V.V. Nemytskii and V.V. Stepanov, "Qualitative Theory of Differential Equations", (Princeton Univ. Press) p.138.
18. R. Reissig, G. Sansone and R. Conti, "Nonlinear Differential Equations of Higher Order" (Noordhoff Int. Publ. Leyden) Chapter Three.
19. V.V. Nemytskii and V.V. Stepanov, "Qualitative Theory of Differential Equations", (Princeton Univ. Press) pp 183-193.
20. V.V. Nemytskii, "Lyapunov's method of Rotating Functions for Investigating Oscillations", Dokl. Akad. Nawk SSSR 97 (1954), pp33-36.

21. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.G. Maier, "Qualitative Theory of Dynamic Systems on the Plane" (Israel Prog. for Sci. Trans. Jerusalem 1971). Chapter Nine.
22. A. Schwartz, "A Generalization of a Poincaré-Bendixson Theorem To Closed Two-dimensional Manifolds". Amer. Journ. Math. 85. pp453-458.
23. R. Abraham, "Foundations of Mechanics". (W. Benjamin Inc. N.Y.1967) pp 158-161, 164.
24. ———— "Foundations of Mechanics". (W. Benjamin Inc. N.Y.1967) pp 170-171.
25. Jack K. Hale, "Ordinary Differential Equations" (Wiley, New York), Chapter 10.
26. A. Andronov and L. Pontriagin, "Systems Grossiers", Dokl. Akad. Nauk SSSR 14. pp 247-251.
27. R. Abraham, "Foundations of Mechanics" (W. Benjamin Inc. N.Y.1967) pp 168-169.
28. M.M. Peixoto, "On Structural Stability", Annals of Maths, 69, 1 (1959)
29. ———— "Structural Stability on 2-dimensional Manifolds", Topology 2, pp101-121.
30. S. Smale, "Structurally Stable Systems are Not Dense". Amer. J. Math, 88, pp 491-496.
31. L. Markus, "Structurally stable Differential Systems" Ann of Math. 73 pp 1-19.
32. M. Peixoto, "Qualatative Theory of Differential Equations and Structural Stability in 'Inter. Sym. on Nonlinear Diff. Equations' (New York, 1967).
33. E.A. Coddington and N. Levinson, "Theory of Ordinary Differential Equations", (McGraw-Hill, 1955) Chapter Fourteen.
34. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.G. Maier, "Theory of Bifurcations of Dynamic Systems on a Plane". (Israel Prog. for Sci. Trans. Jerusalem 1971). p 243.
35. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.G. Maier, "Theory of Bifurcations of Dynamic Systems on a Plane". (Israel Prog. for Sci. Trans. Jerusalem 1971). pp 254-261.
36. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.G. Maier, "Theory of Bifurcations of Dynamic Systems on a Plane" (Israel Prog. for Sci. Trans. Jerusalem 1971). Chapter ten.
37. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.G. Maier, "Theory of Bifurcations of Dynamic Systems on a Plane" (Israel Prog. for Sci. Trans. Jerusalem 1971) pp 402-408.

38. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.G. Maier, "Theory of Bifurcations of Dynamic Systems on a Plane". (Israel Prog. for Sci. Trans. Jerusalem 1971) pp 414-421.
39. D.R.J. Chillingworth, "Differential Topology with a view to Applications" (Pitman Publishing London-San Francisco-Melbourne).

Chapter Four

1. F. Brickell and R.S. Clarke, "Differentiable Manifolds" (Van Nostrand, London) pp 169-173.
2. S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Space-time", (Cambridge University Press, 1973). p. 195.
3. E. Calabi and L. Markus, Annals of Maths. 75, 63 (1962).
4. S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Space-time", (Cambridge University Press, 1973). p. 263-265.
5. ———————————— "The Large Scale Structure of Space-time", (Cambridge University Press, 1973), p. 192-194.
6. ———————————— "The Large Scale Structure of Space-time", (Cambridge University Press, 1973), p. 267.

Chapter Five

1. M.W. Hirsch and S. Smale, "Differential Equations, Dynamical Systems, and Linear Algebra" (Academic Press, London 1974). Chapter Two.
2. M.R. Spiegel, "Advanced Calculus" (Schaum's Outline Series McGraw-Hill). pp 260, 262.
3. ———————— "Advanced Calculus" (Schaum's Outline Series McGraw-Hill) pp 260, 263-264.
4. ———————— "Mathematical Handbook of Formulas and Tables" (Schaum's Outline Series McGraw-Hill) p.75.
5. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.G. Maier, "Theory of Bifurcations of Dynamic Systems on a Plane". (Israel Program for Scientific Translations. Jerusalem. 1971). pp 24, 25.
6. Jack K. Hale, "Ordinary Differential Equations" (Wiley, New York), Chapter ten.
7. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.G. Maier, "Qualatative Theory of Dynamic Systems on the Plane" (Israel Program for Scientific Translations. Jerusalem. 1971). p.195.

8. M.R. Spiegel, "Mathematical Handbook of Formulas and Tables". (Schaum's Outline Series. McGraw-Hill). p.96.
9. A.A. Andronov, E.A. Leontovich, I.I. Gordon and A.E. Maier, "Theory of Bifurcations of Dynamic Systems on a Plane" (Israel Program for Scientific Translations. Jerusalem, 1971). pp 261-265
10. P.F. Byrd and M.D. Friedman, "Handbook of Elliptic Integrals for Engineers and Scientists" (Springer-Verlag, Berlin Heidelberg, New York 1971) pp 42-147.

Chapter Six

1. C.W. Misner, K.S. Thorne and J.A. Wheeler, "Gravitation" (W.H. Freeman and Co. San Francisco). pp 593-609, 819-823, 842-844.
2. ———————————————— "Gravitation" (W.H. Freeman and Co. San Francisco). pp 660-663, 674-675.
3. P.F. Byrd and M.D. Friedman, "Handbook of Elliptic Integrals for Engineers and Scientists". (Springer-Verlag, Berlin Heidelberg, New York, 1971). pp 42-147.

Chapter Seven

1. D.R. Long, Phys. Rev., D,9, 850 (1974)
2. D.R. Long, Phys. Rev., D, 10 1677 (1974)
3. D.R. Long, Nature, 260, 417 (1976).